



Universidad de Valladolid



PROGRAMA DE DOCTORADO EN INVESTIGACIÓN BIOMÉDICA

TESIS DOCTORAL:

**ANÁLISIS TRANSCRIPTÓMICO DEL
REMODELADO DE LA HOMEOSTASIS DEL
CALCIO INTRACELULAR EN CÁNCER DE COLON**

Presentada por Enrique Pérez Riesgo para optar al
grado de Doctor por la Universidad de Valladolid

Dirigida por:
Dr. Carlos Villalobos Jorge



AUTORIZACIÓN DEL DIRECTOR/A DE TESIS

(Art. 7.2 de la Normativa para la presentación y defensa de la Tesis Doctoral en la UVa)

D. Carlos Villalobos Jorge, con D.N.I./Pasaporte 09305670P
Investigador Científico (CSIC) del departamento de Fisiología Celular y Molecular
Centro Instituto de Biología y Genética Molecular (IBGM).....
Dirección a efecto de notificaciones IBGM, C/ Sanz y Forés 3, 47003 Valladolid
e-mail carlosv@ibgm.uva.es.....
como Director de la Tesis Doctoral titulada *Análisis transcriptómico del remodelado del calcio intracelular en cáncer de colon*.....
realizada por D. Enrique Pérez Riesgo
alumno/a del Programa de Doctorado en *Investivación Biomédica*
autoriza su presentación, considerando que la tesis reúne todas las condiciones de novedad, excelencia científica y originalidad.

D. Carlos Villalobos Jorge..... (Director/a de la tesis doctoral)¹:

- Declara que cumple los requisitos para poder ejercer la dirección de la tesis doctoral que establecen el RD 99/2011 (modificado por el RD 195/2016) y el Acuerdo del Comité de Dirección de la Escuela de Doctorado de la Universidad de Valladolid del 17 de febrero de 2014, es decir, que posee “al menos un período de actividad investigadora reconocido de acuerdo con las previsiones del RD 1086/1989 de 28 de agosto”².

Valladolid, 8 de Diciembre de 2021

El Director de la Tesis,

Fdo.: Carlos Villalobos Jorge

SR/SRA. PRESIDENTE/A DE LA COMISIÓN DE DOCTORADO

¹ Esta declaración deben realizarla todos los codirectores de la tesis, en el caso de que haya más de un director.

² En el caso de que la figura contractual no sea evaluable por tramos o se trate de un profesor de una universidad extranjera, deberá haber aportado a la Comisión académica del Programa de doctorado un CV u otro documento acreditativo de poseer méritos equivalentes a un tramo de actividad investigadora.

Gracias por no soltarme la mano durante todo este camino, por haberme ayudado a quitar las piedras que había en el camino y por haberme ayudado a levantarme cada vez que me he caído.

Agradecimientos

Si hay una frase o dicho capaz de resumir mi paso por la tesis, ese es el proverbio africano “Si quieres ir rápido camina solo, si quieres llegar lejos ve acompañado”. Y es que no podría ser más acertado, pues nunca habría podido llegar hasta este punto sin el apoyo que, por suerte, he recibido: la ayuda y oportunidades que la vida me ha presentado; la educación que he tenido la suerte de recibir; las buenas amistades que, afortunadamente, he ido descubriendo a lo largo de los años, algunas de las cuales se han convertido en los hermanos que nunca tuve, y, por supuesto, el apoyo incondicional de mi familia. De hecho, sería muy soberbio por mi parte pretender, en unas pocas páginas, nombrar y agradecer a todas las personas que me han apoyado, ayudado y empujado hacia adelante; que han confiado en mí, y que me han hecho valorarme y cambiar la visión que tenía de mí... De hecho, gracias a todos y cada uno de ellos, poco a poco mi camino ha ido tomando la dirección que me ha llevado a realizar esta tesis. Por ello, en homenaje a todos vosotros, os nombraré solo a unos pocos, pero, aparezcáis o no en estas páginas ¡MUCHAS GRACIAS!

Cuando uno echa la vista atrás y piensa en cómo ha llegado hasta ese punto, es inevitable pensar en aquellos profesores de la infancia que nos cogieron como si de un pedazo de arcilla nos tratásemos y nos moldearon acorde a ciertos valores. No obstante, me acuerdo aún más de aquellos que, aparte de inculcarnos ciertos valores, también nos dieron las herramientas necesarias para poder moldear y crear nuestros propios valores y pensamientos, para pensar por nosotros mismos y para saber que para brillar no hay que apagar al resto, sino apoyarse en ellos. Por ello, me gustaría dar las gracias a un gran profesor que me marcó desde que yo no era más que un renacuajo y que siempre nos enseñó a pensar por nosotros mismos, a ser libres y a respetar a todas las personas. Le conocí en el bus que realizaba la ruta escolar, pues él era el profesor asignado a dicha ruta. Curiosamente, fue un día en el que me bajé en una parada equivocada (culpa mía), cuando comencé a tener más relación con él. Unos años después, en quinto y sexto de primaria, tuve el placer de que él fuera mi tutor. También recuerdo que cuando mi papi se fue de misión a Bosnia yo estaba muy triste y él se volcó totalmente conmigo para animarme. Desafortunadamente, a raíz de irme a vivir a Tenerife, perdimos el contacto, lo que no fue un impedimento para que cuando volvimos a vernos, tras cuatro años, me trató como si el tiempo nunca hubiera pasado. De hecho, si no recuerdo mal, me invitó a un lazo de crema en la cafetería del colegio que nos comimos fuera en el patio al lado de las fuentes, lazo que para mí tuvo un significado importante dada la situación que, entonces, estaba atravesando. Por ello, puesto que estoy seguro que gran parte de mi forma de ser te la debo a ti, Don Andrés Camilo Bermejo Y González, querido profesor de mi infancia, te doy las gracias.

También me gustaría dar las gracias a aquella familia que uno elige, aquellos amigos que, más que amigos, son los hermanos que nunca tuve y que, a pesar de la distancia que nos separa y de las obligaciones que el día a día nos impone, el vínculo tan fuerte que nos une hace que el tiempo no sea una piedra en el camino de nuestra amistad, sino una prueba de lo mucho que nos hacemos falta. Tanto es así que, aunque nos veamos cada vez con más arrugas, en nuestros corazones el tiempo apenas ha transcurrido.

Gracias James, por los momentos compartidos desde el primer día de carrera, en clase, en el gym como en novena -y en mi primera donación de sangre previa a la primera novena aguantando mis mareos-... y por haber estado ahí desde entonces sabiendo que siempre podré contar contigo. Gracias Santodo, porque desde que te conocí has sido un importante apoyo, tanto a nivel personal como académico, y por seguir ahí a pesar de que te robé un *tupper* de tortilla (que te devolveré cuando me presentes a esa cosita llamada Ánder).

Gracias a Manu por esas broncas telefónicas cada vez que me veía decaído, desmotivado o que me obstaculizaba a mí mismo y no avanzaba, pero también por ese primer piti el día de mis novatadas que, a modo de flechas cruzadas, supuso el inicio de una gran amistad que perdurará por siempre.

Gracias a Mario Sexy, pues, aunque nuestro primer cruce de palabras no fue más que un “Hola Emo” de ti hacia mí en el jardín trasero de la fácul vieja, el resto de conversaciones que hemos tenido y batallas que hemos librado han sido sumamente valiosas para mí; también por esas broncas en Suiza que, junto con James y jelen, me echasteis por no haber acabado todavía la tesis.

Gracias Miralles, pues, aunque siempre hemos tensado la cuerda y “jugado” a picarnos -desde aquel día en el que estabas en el “Paris” con una máquina de hacer *piercings*- has sido un gran apoyo para mí; también me has ayudado de una forma indescriptible en momentos verdaderamente malos -incluso durante horas al otro lado del teléfono- y, afortunadamente, también hemos podido disfrutar juntos de otros tantos buenos momentos.

Gracias al Señor Glez, pues desde aquel paquete a pachas en el otro trastero te has convertido en una persona importante en mi vida y me has ayudado a valorarme y a confiar en mí mismo; todavía recuerdo aquella conversación cuando os dije en Villa SAM a ti y a Ramón que quería estudiar el QUIR y me respondisteis que ni se me ocurriera, que yo tenía que dedicarme a la investigación.

Gracias a Ramona, pues quién iba a pensar que de aquel cruce de miradas marcando territorio en los aparcamientos de la facultad vieja, y a pesar de tus rastas y de mi flequillo rojo, te convertiste en una persona muy importante en mi vida. No obstante, fue un día en mi buhardilla cuando realmente nos dimos cuenta de que podríamos confiar absolutamente el uno en el otro. Te estoy y estaré eternamente agradecido, Ramona, no por tus placajes, sino por haberme valorado tanto y por transmitirme una visión de mí que, a día de hoy, me sigue emocionando y que, al mismo tiempo, me asusta, pues pienso que todavía estoy lejos de alcanzarla. Sé de un afortunado mini Ramón, que no se salta del día de pierna en el gym, al que no le va a faltar cariño y que va a tener un gran ejemplo de padre a quien seguir.

Gracias a Clarita por haberme apoyado y aguantado tantos años y, a pesar de ello, seguir quedando conmigo, crispando a la gente y haciendo nuestros bailes; personas como tú (y hay pocas) me hacen pensar que es totalmente cierto que los ojos son el espejo del alma, pues eres una bellísima persona.

Por último, y no por ello menos importante, gracias a todos aquellos sumos con los que he tenido la suerte de compartir mesa en Novena, a la stirpe de los jinchos, a los VillaSAMianos y, en general, a todo el equipo de la élite por todos estos años, apoyo y momentos que, a pesar del Sanzoles, se han convertido en inolvidables.

Por supuesto, gracias a Carlos y a Lucía, pues, desde el minuto uno, habéis confiado en mí, valorado mis ideas y me habéis guiado en este maravilloso viaje que ha sido para mí la tesis doctoral y que, aunque este viaje esté llegando a su fin, espero -y estoy seguro de ello- que nuestra relación perdure en el tiempo. Carlos, todavía recuerdo aquel mensaje del día 10 de mayo de 2016 que, aunque suene a cliché, me cambió la vida, pues me ha permitido encontrar y desarrollar mi vocación, pues con él me diste la oportunidad de comenzar la tesis con vosotros debido, y cito textualmente, a mi “querencia por la estadística y la bioinformática”. También, muchas gracias por la confianza que habéis depositado en mí, pues desde el primer momento me habéis permitido compaginar la tesis con mi formación académica en el maravilloso mundo de la ciencia de datos, la bioestadística y la bioinformática. Por ello, doy gracias porque hayáis sido mis mentores todo este tiempo y, sinceramente, espero que esta relación que la tesis ha creado perdure a lo largo del tiempo. También, y aunque fuera anterior a comenzar la tesis, gracias a Carmen, pues con ella comencé mi paseo aleatorio por el increíble mundo de la ciencia de datos y por las buenas palabras que siempre ha tenido sobre mí. A Laura, y aunque nos conocemos hace relativamente poco, me gustaría agradecerle su ayuda y sus consejos, su apoyo y la confianza que ha depositado en mí desde el primer día.

Gracias a David, pues su trabajo ha sido clave para la obtención de gran parte de los resultados de esta tesis. Pero no solo cuenta el trabajo realizado, sino también tu dedicación, pues al igual que yo con los datos, se nota cuando a una persona le apasiona lo que hace por cómo lo hace, y no hay más que verte trabajar para darte cuenta del mimo con el que realizas cada uno de los experimentos y, por ello, de la gran calidad de los resultados que de ellos se obtienen. También me gustaría dar las gracias al resto de mis compis de laboratorio, tanto aquellos que ya no están (Mónica y Lucía) como aquellos que todavía permanecen en él (Vero, Sara, Sendoa y Jelen). Gracias por aguantar mis “chapas” de estadística y por los consejos que me habéis dado. Por supuesto, gracias también por no enfadaros conmigo todas aquellas veces que, sin previo aviso, he acaparado el cuarto de cultivos y el equipo de imagen; intentaré centrarme más en los números para que eso no vuelva a ocurrir.

Gracias a mi familia, pues su apoyo y respaldo ha sido clave para convertirme en el hombre que soy hoy en día. Gracias por la educación que me habéis dado, por los valores que me habéis transmitido y por el cariño que siempre he sentido. También, gracias por haberme permitido realizar y alcanzar todos mis sueños (aunque todavía tengo pendiente aprender a tocar el violín), y por no haberme puesto nunca ningún impedimento a ninguna de mis ideas, por muy raras que pudieran sonar, como cuando os dije: “Papá, mamá ¿Puedo aprender a tocar el piano?”; “Papa, mamá ¿Puedo dejar de tocar

el piano para aprender a patinar?"; "Papa, mama, Elena, después de una licenciatura, tres máster y mientras hago el doctorado ¿Qué os parece que empiece otra carrera?"... No obstante, también gracias por haber evitado que cometiera otros errores, como cuando os dije: "Papá, mamá ¿Puedo dejar los estudios para irme a la mina para conducir un Dumper?". También, gracias por haberme hecho ver la luz cuando mi ceguera me hacía incapaz de salir del pozo de autodestrucción en el que caí ya casi, se dice pronto, hace casi 20 años, cuando durante más de cuatro años la anorexia, la bulimia y la depresión tomaron las riendas de "mi vida" y tiraban de aquellos escasos 42 kilos; pues, de no ser por todos vosotros, en especial por mis papis (y de aquella napolitana de chocolate que me visteis tirar al barranco yendo hacia el instituto y que, mágicamente, apareció en casa cuando volví de este junto con una nota totalmente desgarradora y absolutamente cierta), estoy seguro de que las casi 700 hojas de esta tesis nunca se hubieran podido escribir. Por ello, gracias mamuchi y papuchi, pues siempre estáis ahí para cuidarme y apoyarme en todo; gracias a mis abuelitos, que desde el cielo me cuidan; gracias a mis abuelitas, a mis tíos y a mis primos, pues siempre me hacéis saber que me queréis y que estáis orgullosos de mí. Por supuesto, gracias a mis princesitas Fabiana y Elia; a mis principitos Toñín y Pablín, y mi querido ahijado Danielín, que por desgracia no os veo tanto como me gustaría. Solo con ver lo orgullosos que estáis cuando decís "Kikín es investigador" me da una fuerza hercúlea que me ayuda a seguir siempre adelante sin importar lo dura que pueda resultar en ocasiones esta vida investigadora que, por otro lado, tantas alegrías me han dado.

Evidentemente, me gustaría también dar las gracias a mi otra nueva familia a la que he tenido la gran suerte de conocer gracias a Elena, mi compañera de viaje desde, aproximadamente, mi tercer año de carrera (hace ya 9 años y una semana ahora mismo), pues desde el primer momento me abrieron sus puertas, me aceptaron como uno más y me han animado a seguir mis y también me han apoyado en todos los pasos que he dado desde entonces. También a ti Daddy, que tanta suerte has tenido de la familia que te ha tocado, gracias por aquella oscura tarde en la que el agobio y estrés de la tesis se desvanecieron al observarte y pensar en lo feliz que estabas jugando contigo mismo frente al espejo, lo concentrado que estabas en mi rodilla mirando las luces del ordenador mientras escribía la tesis, recorriendo la casa entera en tacatá, e, incluso, viendo Pepa Pig para que no lloraras más después del susto que te di jugando al "cucu tras".

Gracias a mi compañera de vida que, como ya dije anteriormente, me ha acompañado durante más de 9 años y que espero que sigas a mi lado, y yo al tuyo, para seguir creando y compartiendo momentos, compartiendo pasiones, apoyándonos y complementándonos. Desde el día en que te conocí, en plena Novena en el bar "el desconcierto", fui incapaz de quitarte de mi cabeza. De hecho, un año después, cuando nuestros caminos volvieron a cruzarse, tuve la suerte de que tú también te fijaras en mí y de que me aceptaras en tu vida. Gracias por todos estos años, por todo lo que me has aportado, ayudado, confiado, animado, e incluso por esos momentos en el gym que tengo la suerte de compartir contigo... Sabes que sin tu apoyo y sin tu ayuda no hubiera llegado hasta este punto, pues fuiste una de las personas que confió en mí y me animaste a dar el paso definitivo para comenzar la tesis y para seguir formándome en lo que, a día de hoy, sigue siendo mi pasión, la ciencia de datos y la biomedicina. Además, tampoco hubiera podido alcanzar otras tantas metas que me he ido proponiendo y consiguiendo con tu ayuda. Afortunadamente, he tenido la suerte de

que no solo personal sino también laboral y académicamente nos apasionen las mismas cosas y que, por ello, nuestra carrera investigadora continúe en el mismo laboratorio. De hecho, y me siento orgulloso de ello, gran parte del trabajo que presento en esta tesis no existiría sin tu gran ayuda y, por supuesto, por tu pasión y dedicación a la ciencia. Por todo ello, y más, ¡Gracias Aceitunita!

Finalmente, y no por ello menos importante, gracias a la Asociación Española Contra el Cáncer, pues sin su apoyo y ayuda no habría sido posible realizar esta tesis. Gracias por haber confiado en mí y por haberme dado la posibilidad de aportar mi granito de arena en la lucha contra el cáncer. También, me gustaría dar las gracias a todas aquellas personas que apoyan la lucha contra el cáncer y que, grano a grano, hacen posible que la AECC tenga cada día más fuerza para luchar contra el cáncer. Por supuesto, a todas las personas que lidiáis día a día con el cáncer, ya sea porque lo padecéis o porque alguien cercano a vosotros lo sufre ¡GRACIAS POR VUESTRA FUERZA, ENTEREZA Y ESPERANZA!

Aportaciones científicas

Durante el desarrollo de la tesis doctoral, se han realizado las siguientes publicaciones científicas:

- **Pérez-Riesgo, E.**, Gutiérrez, L. G., Ubierna, D., Acedo, A., Moyer, M. P., Núñez, L., & Villalobos, C. (2017). Transcriptomic Analysis of Calcium Remodeling in Colorectal Cancer. *International journal of molecular sciences*, 18(5), 922. <https://doi.org/10.3390/ijms18050922>
- Núñez, L., Bird, G. S., Hernando-Pérez, E., **Pérez-Riesgo, E.**, Putney, J. W., Jr, & Villalobos, C. (2019). Store-operated Ca²⁺ entry and Ca²⁺ responses to hypothalamic releasing hormones in anterior pituitary cells from *Orai1*^{-/-} and heptaTRPC knockout mice. *Biochimica et biophysica acta. Molecular cell research*, 1866(7), 1124–1136. <https://doi.org/10.1016/j.bbamcr.2018.11.006>
- Hernando-Pérez, E., **Pérez-Riesgo, E.**, Cepeda, S., Arrese, I., Sarabia, R., Villalobos, C., & Núñez, L. (2021). Differential Ca²⁺ responses and store operated Ca²⁺ entry in primary cells from human brain tumors. *Biochimica et biophysica acta. Molecular cell research*, 1868(8), 119060. <https://doi.org/10.1016/j.bbamcr.2021.119060>
- Parra-Izquierdo, I., Sánchez-Bayuela, T., López, J., Gómez, C., **Pérez-Riesgo, E.**, San Román, J. A., Sánchez Crespo, M., Yacoub, M., Chester, A. H., & García-Rodríguez, C. (2021). Interferons Are Pro-Inflammatory Cytokines in Sheared-Stressed Human Aortic Valve Endothelial Cells. *International journal of molecular sciences*, 22(19), 10605. <https://doi.org/10.3390/ijms221910605>

Además, han sido presentados en forma de comunicación en los siguientes congresos científicos:

- Congreso /Evento: XV International Meeting of the European Calcium Society (ECS2018). Núñez L, Bird GS, Hernando-Pérez E, **Pérez-Riesgo E**, Birnbaumer L, Villalobos C and Putney JW. Store-operated Ca²⁺ entry and Ca²⁺ responses to hypothalamic releasing hormones in anterior pituitary cells from *Orai1* and hepta TRPC knockout mice.
Comunicación en forma de Póster. Publicación: Abstract publicado en libro de abstracts y Artículo completo en *BBA Mol Cel Res*. Lugar de celebración: Hamburgo, Alemania Fecha: 9-13 septiembre de 2018.

-
- Congreso /Evento: XV International Meeting of the European Calcium Society (ECS2018).
Pérez-Riesgo E and Villalobos. Illuminating transcriptomics of calcium remodeling in colorectal cancer.
Comunicación en forma de Póster. Publicación: Abstract publicado en libro de abstracts.
Lugar de celebración: Hamburgo, Alemania Fecha: 9-13 septiembre de 2018.
 - Congreso /Evento: XXXIX Congreso de la Sociedad Española de Ciencias Fisiológicas (SECF).
Pérez-Riesgo E and Villalobos. Illuminating transcriptomics of calcium remodeling in colorectal cancer.
Comunicación en forma de Póster. Publicación: Abstract publicado en libro de abstracts.
Lugar de celebración: Cádiz, España Fecha: 18-21 Septiembre de 2018.
 - Congreso /Evento: VII RECI CONGRESS
Pérez-Riesgo E and Villalobos. Illuminating transcriptomic analysis of calcium remodeling in colorectal cancer cells.
Comunicación en forma de Póster. Publicación: Abstract publicado en libro de abstracts.
Lugar de celebración: Cáceres, España Fecha: 15-17 mayo de 2019.

“Nada tiene tanto poder para ampliar la mente como la capacidad de investigar de forma sistemática y real todo lo que es susceptible de observación en la vida”

Marco Aurelio

Índice

| | |
|--|--------------|
| Lista de Figuras | VIII |
| Lista de Tablas | XIX |
| Lista de Abreviaturas | XXI |
| Resumen | XXVII |
| | |
| I Introducción | 1 |
| | |
| 1. Cáncer de colon | 1 |
| 1.1. El cáncer | 1 |
| 1.2. Sellos o <i>hallmarks</i> del cáncer | 2 |
| 1.2.1. Mantenimiento de la señal proliferativa | 3 |
| 1.2.2. Evasión de las señales inhibitorias del crecimiento | 6 |
| 1.2.3. Resistencia a la muerte | 8 |
| 1.2.4. Potencial ilimitado de replicación | 17 |
| 1.2.5. Inducción de la angiogénesis | 18 |
| 1.2.6. Activación de la invasión y la metástasis | 20 |
| 1.2.7. Inestabilidad genómica y mutación | 26 |
| 1.2.8. Reprogramación metabólica | 27 |
| 1.2.9. Evasión del sistema inmune | 31 |
| 1.2.10. Estado inflamatorio inducido por el tumor | 32 |
| 1.3. Cáncer de colon | 34 |

| | | |
|-----------|---|-----------|
| 1.3.1. | Incidencia del cáncer de colon y sus factores de riesgo | 34 |
| 1.3.2. | Vías de desarrollo del CCR | 35 |
| 1.3.2.1. | Inestabilidad cromosómica | 39 |
| 1.3.2.2. | Intestabilidad de microsátélites y mecanismos de reparación de apareamientos erróneos | 39 |
| 1.3.2.3. | Hiperactivación de la ruta de señalización Wnt/ β – catenina | 40 |
| 1.3.2.4. | Activación de Ras/MAPK | 42 |
| 1.3.2.5. | Receptores nucleares (NR) | 43 |
| 1.3.3. | Metástasis de CCR | 44 |
| 1.3.4. | Prevención y tratamientos | 45 |
| 1.3.5. | Inhibición de la síntesis de poliaminas | 47 |
| 2. | Homeostasis del Calcio Intracelular | 53 |
| 2.1. | El calcio intracelular | 53 |
| 2.2. | Señalización por Ca^{2+} intracelular | 55 |
| 2.3. | La membrana plasmática | 58 |
| 2.3.1. | Mecanismos <i>ON</i> de la MP | 58 |
| 2.3.1.1. | Canales de calcio operados por voltaje (VOCC) | 59 |
| 2.3.1.2. | Canales de Ca^{2+} operados por receptor (ROCC) | 64 |
| 2.3.1.3. | Canales de Ca^{2+} operados por segundo mensajero (SMOCC) | 69 |
| 2.3.1.4. | Canales de Ca^{2+} operados por depósitos (SOCC) | 70 |
| 2.3.2. | Mecanismos <i>OFF</i> de la MP | 72 |
| 2.3.2.1. | Bombas Ca^{2+} /ATPasa de la MP | 72 |
| 2.3.2.2. | Intercambiadores Na^{+}/Ca^{2+} | 74 |
| 2.4. | El retículo endoplasmático (RE) | 75 |
| 2.4.1. | Mecanismos <i>ON</i> del RE | 76 |
| 2.4.1.1. | Receptores de inositol trifosfato (IP3R) | 76 |
| 2.4.1.2. | Receptores de rianodina (RyR) | 79 |
| 2.4.2. | Mecanismos <i>OFF</i> del RE | 80 |
| 2.4.2.1. | ATPasas de Ca^{2+} del retículo sarco/endoplasmático (SERCA) | 80 |
| 2.4.2.2. | ATPasas de Ca^{2+} de la vía secretora (SPCA) | 81 |

| | |
|---|------------|
| 2.5. La mitocondria | 82 |
| 2.5.1. El complejo uniportador de Ca ²⁺ mitocondrial (MCU) | 86 |
| 2.5.2. Canales aniónicos dependientes de voltaje (VDAC) | 89 |
| 2.5.3. Intercambiadores de Ca ²⁺ mitocondriales | 89 |
| 2.6. Entrada capacitativa de Ca ²⁺ | 91 |
| 2.6.1. Descubrimiento de SOCE | 91 |
| 2.6.2. Bases moleculares de la SOCE | 93 |
| 2.6.3. ORAI | 95 |
| 2.6.4. STIM | 102 |
| 2.6.5. MS4A12 | 109 |
| 2.6.6. Moduladores de SOCE | 110 |
| 2.6.6.1. CRACR2A | 111 |
| 2.6.6.2. Septinas | 113 |
| 2.6.6.3. ORMDL3 | 114 |
| 2.6.6.4. SARAF | 115 |
| 2.6.6.5. SIMTATE | 116 |
| 2.6.6.6. MBP | 116 |
| 2.7. Superfamilia de los TRP | 117 |
| 2.7.1. TRPC | 118 |
| 2.7.2. TRPV | 122 |
| 2.7.3. TRPM | 123 |
| 2.7.4. TRPA | 125 |
| 2.7.5. TRPP | 126 |
| 2.7.6. TRPML | 127 |
| 3. Remodelado del Ca²⁺ intracelular en el cáncer | 129 |
| 3.1. Remodelado de los VOCC | 129 |
| 3.2. Remodelado de las PMCA | 132 |
| 3.3. Remodelado de los NCX | 132 |
| 3.4. Remodelado de las SERCA | 133 |

| | |
|---|------------|
| 3.5. Remodelado de la liberación de Ca ²⁺ desde el RE | 133 |
| 3.6. Remodelado de la liberación de Ca ²⁺ de la vía secretora | 134 |
| 3.7. Remodelado del transporte de Ca ²⁺ a través de los TRP | 135 |
| 3.7.1. TRPC | 135 |
| 3.7.2. TRPV | 138 |
| 3.7.3. TRPM | 140 |
| 3.7.4. TRPP | 143 |
| 3.7.5. TRPML | 143 |
| 3.8. Remodelado de SOCE | 144 |
| 3.9. Remodelado del transporte de Ca ²⁺ mitocondrial | 153 |
| | |
| II Hipótesis y Objetivos | 157 |
| | |
| 4. Hipótesis de trabajo y objetivos | 159 |
| | |
| III Material y Métodos | 161 |
| | |
| 5. Material | 163 |
| 5.1. Modelos de estudio | 163 |
| 5.1.1. Modelos celulares | 163 |
| 5.1.2. Modelos de tejido humano | 164 |
| 5.2. Medios, reactivos y materiales de cultivo celular | 165 |
| 5.3. Medios de imagen de Ca ²⁺ | 165 |
| 5.4. Reactivos y materiales para la extracción de proteínas y <i>Western Blot</i> | 166 |
| 5.5. Reactivos y materiales para la extracción de ARN | 167 |
| 5.6. Sondas, reactivos de las soluciones y otros compuestos | 167 |
| 5.7. Bases de datos de conjuntos de genes | 167 |
| | |
| 6. Métodos | 169 |
| 6.1. Cultivos Celulares | 169 |
| 6.2. Tratamiento de células NCM460 y HT29 con DFMO | 169 |

| | | |
|----------|---|-----|
| 6.3. | Imagen de Ca^{2+} intracelular en célula única | 171 |
| 6.3.1. | Fundamentos teóricos | 171 |
| 6.3.2. | Procedimiento y diseño experimental para el análisis funcional del efecto del DFMO en los modelos celulares HT29 y NCM460 | 175 |
| 6.3.3. | Análisis de los datos de experimentos de imagen de Ca^{2+} | 176 |
| 6.4. | Extracción de ARN | 177 |
| 6.4.1. | Extracción de ARN a partir de líneas celulares | 177 |
| 6.4.2. | Extracción de ARN a partir de tejido | 178 |
| 6.5. | Técnicas transcriptómicas de alto rendimiento | 180 |
| 6.5.1. | Fundamentos teóricos | 180 |
| 6.5.1.1. | Microarrays de ARN | 181 |
| 6.5.1.2. | RNA-seq | 189 |
| 6.5.2. | Procedimientos y diseños experimentales | 197 |
| 6.5.2.1. | Análisis transcriptómico de las líneas celulares HT29 y NCM460 | 198 |
| 6.5.2.2. | Análisis transcriptómico del efecto del DFMO en los modelos celulares HT29 y NCM460 | 198 |
| 6.5.2.3. | Análisis transcriptómico de muestras de tejido humano de cáncer de colon | 199 |
| 6.5.3. | Análisis de datos transcriptómicos de alto rendimiento | 200 |
| 6.5.3.1. | Análisis de datos de microarrays | 200 |
| 6.5.3.2. | Análisis de datos de RNA-seq | 213 |
| 6.6. | Extracción de proteínas | 214 |
| 6.7. | Análisis de la expresión de proteínas mediante <i>Western Blot</i> | 215 |
| 6.7.1. | Fundamentos teóricos y protocolo | 215 |
| 6.7.2. | Procedimiento y diseño experimental | 217 |
| 6.7.3. | Análisis de datos de <i>Western Blot</i> | 218 |
| 6.8. | Metaanálisis | 221 |
| 6.8.1. | Fundamentos teóricos | 221 |
| 6.8.1.1. | Combinación de los efectos | 226 |
| 6.8.1.2. | Combinación de p valores | 227 |
| 6.8.2. | Procedimiento y diseño experimental | 228 |

| | |
|---|------------|
| <i>ÍNDICE</i> | 22 |
| IV Resultados | 229 |
| 7. Resultados | 231 |
| 7.1. Remodelado de la homeostasis del Ca^{2+} intracelular en líneas celulares de cáncer de colon | 231 |
| 7.1.1. Estudio funcional del remodelado del Ca^{2+} intracelular mediante imagen de Ca^{2+} en líneas celulares | 232 |
| 7.1.2. Análisis transcriptómico de las líneas celulares de cáncer de colon HT29 y de colon normal NCM460 | 236 |
| 7.1.2.1. Análisis de expresión diferencial entre las líneas celulares HT29 y NCM460 | 238 |
| 7.1.2.2. Análisis de significación biológica | 248 |
| 7.1.3. Efecto del tratamiento con DFMO sobre la homeostasis del Ca^{2+} intracelular en colonocitos normales y en células de cáncer de colon | 252 |
| 7.1.3.1. Estudio del efecto del DFMO sobre el Ca^{2+} citosólico, el tamaño de los depósitos de Ca^{2+} intracelular y SOCE | 252 |
| 7.1.3.2. Análisis transcriptómico de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO | 257 |
| 7.2. Análisis del remodelado de la homeostasis del Ca^{2+} intracelular en tejido humano de cáncer de colon | 272 |
| 7.2.1. Análisis transcriptómico de muestras pareadas de tejido humano de cáncer de colon | 272 |
| 7.2.1.1. Análisis de expresión diferencial | 273 |
| 7.2.1.2. Análisis de significación biológica | 279 |
| 7.2.2. Análisis de expresión de proteínas implicadas en SOCE | 290 |
| 7.2.3. Reanálisis transcriptómico de muestras de tejido humano de cáncer de colon extraídas de repositorios públicos y metaanálisis | 295 |
| 7.2.3.1. Reanálisis y metaanálisis de expresión diferencial | 296 |
| 7.2.3.2. Metaanálisis de enriquecimiento | 307 |
| V Discusión | 313 |
| 8. Discusión | 315 |

| | |
|--|------------|
| 8.1. Remodelado del calcio intracelular en cáncer colorrectal en células modelo de colon humano normal (NCM460) y tumoral (HT29). | 316 |
| 8.1.1. Remodelado de los canales de calcio dependientes de voltaje | 319 |
| 8.1.2. Remodelado de la entrada capacitativa de calcio (SOCE) y los canales TRPs | 320 |
| 8.1.3. Remodelado de los Sistemas de Extrusión de Calcio | 322 |
| 8.1.4. Remodelado de la liberación de Ca^{2+} de los depósitos | 323 |
| 8.1.5. Remodelado del transporte de Ca^{2+} en la mitocondria | 323 |
| 8.2. Efecto de la depleción de poliaminas sobre el remodelado del Calcio Intracelular en Cáncer Colorrectal observado en células modelo de colon humano normal y tumoral | 324 |
| 8.3. Remodelado del Calcio Intracelular en muestras pareadas de tejido de colon humano normal y tumoral | 330 |
| 8.4. Metaanálisis del Remodelado del Calcio Intracelular en Cáncer Colorrectal | 335 |
| | |
| VI Conclusiones | 349 |
| | |
| 9. Conclusiones | 351 |
| | |
| VII Apéndices | 355 |
| | |
| A. Fundamentos básicos para el análisis de datos | 357 |
| A.1. Estadística descriptiva | 357 |
| A.1.1. Estadística descriptiva univariante | 359 |
| A.1.2. Análisis descriptivo multivariante | 361 |
| A.2. Variables compuestas o redundantes | 365 |
| A.3. Análisis de componentes principales | 366 |
| A.4. Detección multivariante de outliers | 368 |
| A.5. Coeficiente de correlación de Pearson | 370 |
| A.6. Estudios, diseño de experimentos y modelos | 371 |
| A.6.1. Modelos lineales de efectos fijos | 375 |
| A.6.1.1. Mínimos cuadrados ordinarios (MCO) | 376 |
| A.6.1.2. Mínimos cuadrados generalizados | 384 |

| | | |
|-----------|---|------------|
| A.6.1.3. | Modelos Lineales Generalizados (MLG) | 386 |
| A.6.1.4. | Suposiciones del modelo de regresión lineal multivariante | 387 |
| A.6.1.5. | Selección de Variables | 398 |
| A.6.1.6. | Detección y tratamiento de Outliers | 401 |
| A.6.2. | Modelos Lineales Mixtos | 402 |
| A.6.2.1. | Inferencias | 406 |
| A.6.2.2. | Selección del Modelo | 408 |
| A.7. | Modelos Predictivos: Machine Learning | 409 |
| A.7.1. | Algoritmos de Árboles de Clasificación y Bosques Aleatorios (Random Forest) | 409 |
| B. | Análisis de datos transcriptómicos | 413 |
| B.1. | Análisis de datos transcriptómicos basados en RNA-seq | 413 |
| B.1.1. | Control de calidad y preprocesado de los datos transcriptómicos de los experimentos basados en RNA-seq para la comparación entre las líneas celulares HT29 y NCM460 | 413 |
| B.1.2. | Análisis de enriquecimiento de genes de la expresión diferencial entre las líneas celulares HT29 y NCM460 de los experimentos basados en RNA-seq | 416 |
| B.2. | Caso práctico: análisis de microarrays de las muestras pareadas de tejido humano de cáncer de colon | 432 |
| B.2.0.1. | Análisis exploratorio y control de calidad | 433 |
| B.2.0.2. | Preprocesado de los datos | 439 |
| B.2.0.3. | Filtrado y análisis de expresión diferencial | 441 |
| B.2.0.4. | Análisis de significación biológica | 451 |
| B.3. | Análisis transcriptómico de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las líneas HT29 y NCM460 | 472 |
| B.3.1. | Control de calidad y preprocesado de los datos transcriptómicos de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO | 472 |
| B.3.2. | Análisis de expresión diferencial de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO | 474 |
| B.3.3. | Análisis de enriquecimiento de genes de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO | 478 |

| | | |
|-----------|---|------------|
| B.3.3.1. | Análisis de enriquecimiento de genes de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las células NCM460478 | |
| B.3.3.2. | Análisis de enriquecimiento de genes de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las células HT29 | 482 |
| B.3.3.3. | Análisis de enriquecimiento de genes de la expresión diferencial entre las células HT29 y NCM460 | 491 |
| B.4. | Reanálisis transcriptómico de muestras de tejido humano de cáncer de colon extraídas de repositorios públicos y metaanálisis | 500 |
| B.4.1. | Reanálisis y metaanálisis de expresión diferencial | 500 |
| B.4.2. | Metaanálisis de enriquecimiento de genes | 505 |
| C. | Análisis funcional del efecto del DFMO sobre la homeostasis del Ca^{2+} intracelular | 543 |
| C.1. | Caso Práctico: modelado de la variable respuesta <i>Área Bajo la Curva</i> (AUC) de la señal correspondiente al tamaño de los depósitos | 543 |
| C.1.0.1. | evaluación del modelo sobreóptimo | 544 |
| C.1.0.2. | Estimaciones y comparaciones dos a dos del modelo lineal mixto Clásico | 555 |
| C.1.0.3. | Estimación y comparaciones dos a dos del modelo lineal mixto mediante <i>Boostrapping</i> paramétrico | 558 |
| C.2. | Modelado de la variable respuesta <i>Incremento Máximo</i> de la señal correspondiente a los depósitos | 562 |
| C.3. | Modelado de la variable respuesta <i>Área Bajo la Curva</i> (AUC) de la señal correspondiente a SOCE | 572 |
| C.4. | Modelado de la variable respuesta <i>Incremento Máximo</i> de la señal correspondiente a SOCE | 582 |
| C.5. | Modelado de la variable respuesta $[\text{Ca}^{2+}]_{\text{cit}}$ basal | 593 |
| D. | Análisis de datos de Western Blot | 599 |
| D.1. | Modelado de la expresión de la proteína ORAI1 en muestras de tejido humano de cáncer de colon y normal | 599 |
| D.2. | Modelado de la expresión de la proteína ORAI2 en muestras de tejido humano de cáncer de colon y normal | 603 |
| D.3. | Modelado de la expresión de la proteína TRPC1 en muestras de tejido humano de cáncer de colon y normal | 610 |

| | |
|--|-----|
| D.4. Modelado de la expresión de la proteína STIM1 en muestras de tejido humano de cáncer de colon y normal | 614 |
| D.5. Modelado de la expresión de la proteína STIM2 en muestras de tejido humano de cáncer de colon y normal | 621 |

| | |
|--------------------------|------------|
| VIII Bibliografía | 629 |
|--------------------------|------------|

| | |
|---------------------|------------|
| Bibliografía | 631 |
|---------------------|------------|

Lista de Figuras

| | |
|---|----|
| 1.1. <i>Hallmarks</i> del cáncer. | 2 |
| 1.2. Ruta de la quinasa de fosfatidilinositol-3 (PI3K). | 4 |
| 1.3. Mecanismos de la apoptosis. | 10 |
| 1.4. Mecanismos de la autofagia. | 13 |
| 1.5. Mecanismos de la necrosis. | 15 |
| 1.6. Relación entre la apoptosis, autofagia y necrosis. | 16 |
| 1.7. Cariotipo de leucemia mielógena crónica. | 27 |
| 1.8. Reprogramaciones metabólicas implicadas en diferentes <i>hallmarks</i> | 31 |
| 1.9. Incidencia y mortalidad de los cánceres más comunes. | 34 |
| 1.10. Vías implicadas en el desarrollo de cáncer colorrectal. | 37 |
| 1.11. Rutas clave en el cáncer de colon. | 38 |
| 1.12. Expresión de diferentes receptores nucleares a lo largo de la progresión del CCR. | 43 |
| 1.13. Mecanismos de acción de las poliaminas. | 48 |
| 1.14. Síntesis de poliaminas. | 49 |
| | |
| 2.1. Señalización mediada por Ca^{2+} | 57 |
| 2.2. Estructura de los VOCC. | 59 |
| 2.3. Estructura de la subunidad α_1 y dominios reguladores para las diferentes familias de VOCC. | 60 |
| 2.4. Receptores Operados por Receptor. | 65 |
| 2.5. Activación de los canales de Ca^{2+} operados por depósitos. | 71 |
| 2.6. Estructura de las PMCA y regulación. | 73 |
| 2.7. Activación de los canales IP3R. | 76 |

| | |
|--|-----|
| 2.8. Regulación de los canales IP3R. | 78 |
| 2.9. Regulación por la mitocondria de la liberación de Ca ²⁺ desde el RE. | 84 |
| 2.10. Complejo uniportador mitocondrial (MCU). | 87 |
| 2.11. Activación de SOCE a través de SIMT1 y ORA1. | 94 |
| 2.12. Representación de las tres isoformas de ORAI humano. | 96 |
| 2.13. Modelo del reclutamiento aniónico de la apertura de ORAI y la permeabilidad de Ca ²⁺ | 100 |
| 2.14. Interacción de STIM1 con ORAI1. | 103 |
| 2.15. Estructura y dominios de las proteínas STIM. | 104 |
| 2.16. Moduladores de SOCE. | 110 |
| 2.17. Superfamilia de los TRP. | 117 |
| 2.18. Sobreexpresión de componentes de complejos señaladores. | 121 |
| | |
| 3.1. Remodelado de SOCE y de los depósitos en el cáncer de colon. | 145 |
| 3.2. Reversión parcial del remodelado de SOCE en el cáncer de colon por DFMO. | 146 |
| 3.3. Implicación de ORAI y STIM en los <i>hallmarks</i> de cáncer. | 148 |
| 3.4. Control mitocondrial de SOCE en células normales vs. células de cáncer de colon. | 155 |
| | |
| 6.1. Diseño experimental para el análisis transcriptómico y funcional del efecto del DFMO en las líneas celulares HT29 y NCM460. | 170 |
| 6.2. Sonda ratiométrica Fura2 y el proceso de entrada a la célula de su forma Fura2/AM. | 172 |
| 6.3. Esquema del equipo de imagen de fluorescencia. | 174 |
| 6.4. Control de calidad basado en el parámetro RIN. | 178 |
| 6.5. Proceso de fabricación de microarrays mediante la técnica de fotolitografía. | 182 |
| 6.6. Proceso de fabricación de microarrays mediante la técnica <i>microspotting</i> | 183 |
| 6.7. Proceso de fabricación de microarrays mediante <i>inkjetting</i> | 184 |
| 6.8. Microarrays de dos colores. | 185 |
| 6.9. Microarrays de expresión de un color. | 186 |
| 6.10. Microarrays de un color de Affymetrics. | 187 |
| 6.11. Microarrays de expresión de Affymetrix. | 188 |

| | |
|---|-----|
| 6.12. Coste por secuenciación del genoma humano. | 190 |
| 6.13. Estrategias de inmovilización y amplificación de los fragmentos. | 193 |
| 6.14. Técnicas de secuenciación. | 196 |
| 6.15. Ejemplo de sondas redundantes. | 204 |
| 6.16. Obtención de los valores de densidad óptica. | 219 |
| 6.17. Modelo de efecto fijos. | 224 |
| 7.1. Registros en célula única del tamaño de los depósitos intracelulares de Ca^{2+} y SOCE en las líneas celulares HT29 y NCM460. | 233 |
| 7.2. Estimaciones del efecto del DFMO sobre la $[\text{Ca}^{2+}]_{\text{cit}}$ basal. | 234 |
| 7.3. Estimación del tamaño de los depósitos de Ca^{2+} y SOCE para las líneas celulares NCM460 y HT29. | 235 |
| 7.4. Análisis de la expresión diferencial entre las células HT29 (fenotipo tumoral) y NCM460 (fenotipo normal) para los genes de la familia VOCC. | 239 |
| 7.5. Análisis de la expresión diferencial entre las células HT29 (fenotipo tumoral) y NCM460 (fenotipo normal) para los genes de la familia SOCE. | 240 |
| 7.6. Análisis de la expresión diferencial entre las células HT29 (fenotipo tumoral) y NCM460 (fenotipo normal) para los genes que codifican para proteínas reguladoras de SOCE. | 241 |
| 7.7. Análisis de la expresión diferencial entre las células HT29 (fenotipo tumoral) y NCM460 (fenotipo normal) para los genes de la familia TRP. | 243 |
| 7.8. Análisis de la expresión diferencial entre las células HT29 y NCM460 para los genes de la familia CPT. | 244 |
| 7.9. Análisis de la expresión diferencial entre las células HT29 y NCM460 para los genes de la familia ERC. | 245 |
| 7.10. Análisis de la expresión diferencial entre las células HT29 y NCM460 para los genes de la familia MCT. | 246 |
| 7.11. Resumen del análisis de expresión diferencial entre las células HT29 vs NCM460 de los genes implicados en el transporte de Ca^{2+} intracelular. | 247 |
| 7.12. Perfiles de expresión. | 249 |
| 7.13. Efecto del DFMO en el tamaño de los depósitos y SOCE. | 253 |
| 7.14. Estimaciones del efecto del DFMO sobre la $[\text{Ca}^{2+}]_{\text{cit}}$ basal. | 255 |
| 7.15. Estimaciones del efecto del DFMO en el tamaño de los depósitos. | 256 |
| 7.16. Estimaciones del efecto del DFMO en SOCE. | 257 |

| | |
|---|-----|
| 7.17. Porcentaje de DEG tras el tratamiento con DFMO. | 258 |
| 7.18. Diagramas de Venn para el análisis de expresión diferencial | 259 |
| 7.19. DEG tras el tratamiento con DFMO en el mismo sentido en ambas líneas celulares | 260 |
| 7.20. DEG tras el tratamiento en NCM460 con un patrón de comportamiento diferente en HT29. | 261 |
| 7.21. Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia VOCC. | 262 |
| 7.22. Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia SOCE. | 262 |
| 7.23. Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia TRP. | 263 |
| 7.24. Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia CPT. | 263 |
| 7.25. Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia ERC. | 264 |
| 7.26. Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia MCT. | 264 |
| 7.27. Resumen del análisis de expresión diferencial de los genes implicados en el transporte de Ca^{2+} tras la depleción de poliaminas en cada una de las líneas celulares HT29 y NCM460. | 265 |
| 7.28. Resumen del análisis de expresión diferencial de los genes implicados en el transporte de Ca^{2+} intracelular entre las células HT29 vs NCM460 y entre las células tratadas con DFMO y sin tratar. | 266 |
| 7.29. Perfiles de expresión. | 268 |
| 7.30. Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes VOCC. | 274 |
| 7.31. Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes SOCE. | 275 |
| 7.32. Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes TRP. | 276 |
| 7.33. Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes CPT. | 277 |
| 7.34. Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes ERC. | 278 |

| | |
|--|-----|
| 7.35. Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes MCT. | 279 |
| 7.36. Análisis de componentes principales y algoritmo de clasificación supervisada (árboles de clasificación). | 280 |
| 7.37. Análisis de la varianza de la variable respuesta PC1 en función del tipo de tejido y algoritmo de clasificación supervisada (análisis cuadrático discriminante). | 281 |
| 7.38. Correlación de los genes de la familia VOCC con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. | 283 |
| 7.39. Correlación de los genes de la familia SOCE con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. | 283 |
| 7.40. Correlación de los genes de la familia TRP con la PC1 capaz de explicar las diferencias entre los tipos de tejido, PC1. | 284 |
| 7.41. Correlación de los genes de la familia CPT (bombas e intercambiadores de calcio) con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. | 285 |
| 7.42. Correlación de los genes de la familia ERC con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. | 285 |
| 7.43. Correlación de los genes de la familia MCT con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. | 286 |
| 7.44. Resumen del análisis de expresión diferencial de los genes relacionados con el transporte de Ca ²⁺ intracelular y de su influencia en la clasificación entre tejido tumoral y sano. | 286 |
| 7.45. Perfiles de expresión. | 287 |
| 7.46. Análisis de expresión de las proteínas ORAI1, ORAI2 y TRPC1 en muestras pareadas de tejido humano de cáncer de colon y normal mediante <i>Western Blot</i> | 293 |
| 7.47. Análisis de expresión de las proteínas STIM1 y STIM2 en muestras pareadas de tejido humano de cáncer de colon y normal mediante <i>Western Blot</i> | 294 |
| 7.48. Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes VOCC. | 299 |
| 7.49. Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes SOCE. | 300 |
| 7.50. Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes TRP. | 302 |
| 7.51. Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes CPT. | 304 |

| | |
|---|-----|
| 7.52. Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes ERC. | 305 |
| 7.53. Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes MCT. | 306 |
| 7.54. Resumen del metaanálisis de expresión diferencial entre el tejido de cáncer de colon y normal de los genes relacionados con el transporte de Ca^{2+} intracelular. | 307 |
| 8.1. Expresión diferencial de genes relacionados con la homeostasis del Ca^{2+} intracelular entre las células HT29 y NCM460. | 319 |
| 8.2. Diferencias en la homeostasis del Ca^{2+} intracelular entre las células HT29 y NCM460 tratadas y sin tratar con DFMO. | 329 |
| 8.3. Expresión diferencial de genes implicados en la homeostasis del Ca^{2+} intracelular en muestras pareadas de tejido de cáncer de colon humano vs normal. | 335 |
| 8.4. Metaanálisis de expresión diferencial entre el tejido de cáncer de colon y normal de los genes relacionados con el transporte de Ca^{2+} intracelular. | 347 |
| A.1. Ejemplo de la hipótesis nula generada mediante <i>Boostrapping</i> paramétrico partir de los $t_j^{(i)}$ para el estadístico t_j asociado al parámetro β_0 | 393 |
| A.2. Residuos vs. valores predichos por el modelo. | 394 |
| B.1. Control de calidad y detección de <i>outliers</i> con métodos generales. | 415 |
| B.2. Normalización y transformación de los datos de conteo de RNA-seq. | 416 |
| B.3. Análisis exploratorio de los datos <i>outliers</i> | 434 |
| B.4. Análisis de Componentes Principales y detección de <i>outliers</i> mediante la distancia de Mahalanobis. | 436 |
| B.5. Control de calidad y detección de <i>outliers</i> con métodos generales. | 437 |
| B.6. Control de calidad y detección de <i>outliers</i> con métodos específicos para microarrays. | 438 |
| B.7. Corrección del ruido de fondo, normalización y resumen mediante el método RMA. | 440 |
| B.8. Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes VOCC. | 446 |
| B.9. Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes CP. | 447 |

| | |
|--|-----|
| B.10. Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes CLC. | 448 |
| B.11. Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes SOCE. | 449 |
| B.12. Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes TRP. | 450 |
| B.13. Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes MCT. | 451 |
| B.14. Perfiles de expresión. | 452 |
| B.15. Análisis de componentes principales y algoritmo de clasificación supervisada (árboles de clasificación). | 453 |
| B.16. Análisis de la varianza de la variable respuesta PC1 en función del tipo de tejido y algoritmo de clasificación supervisada (análisis cuadrático discriminante). | 454 |
| B.17. Control de calidad y detección de <i>outliers</i> con métodos generales. | 473 |
| B.18. Control de calidad y detección de <i>outliers</i> con métodos específicos para microarrays. | 473 |
| B.19. Corrección del ruido de fondo, normalización y resumen mediante el método RMA. | 474 |
| C.1. Diagnóstico de los residuos de los modelos lineales mixtos para la variable respuesta AUCDepositos y $\log(\text{AUCDepositos})$ | 545 |
| C.2. Diagnóstico de los efectos aleatorios de los modelos lineales mixtos para la variable respuesta AUCDepositos y $\log(\text{AUCDepositos})$ | 546 |
| C.3. Diagnóstico del modelo lineal mixto final (completo) para la variable respuesta $\log(\text{AUCDepositos})$ tras eliminar los <i>outliers</i> | 553 |
| C.4. Predicciones realizadas con el modelo lineal mixto final (completo) para la variable respuesta $\log(\text{AUCDepositos})$ tras eliminar los <i>outliers</i> | 554 |
| C.5. Contrastes de hipótesis mediante <i>Boostrapping</i> | 560 |
| C.6. Diagnóstico del modelo lineal de efectos fijos para la variable respuesta $\log(\text{MaxDepositos})$ tras eliminar los <i>outliers</i> | 567 |
| C.7. Contrastes de hipótesis mediante <i>Boostrapping</i> | 570 |
| C.8. Diagnóstico del modelo lineal de efectos fijos para la variable respuesta $T(\text{AUCSOCE})$ tras eliminar los <i>outliers</i> | 577 |
| C.9. Contrastes de hipótesis mediante <i>Boostrapping</i> | 580 |

| | |
|---|-----|
| C.10. Diagnóstico del modelo lineal de efectos fijos para la variable respuesta log (MaxSOCE) tras eliminar los <i>outliers</i> | 587 |
| C.11. Contrastes de hipótesis mediante <i>Boostrapping</i> | 591 |
| C.12. Diagnóstico del modelo lineal de efectos fijos para la variable respuesta T (CaBasal) _{ijcr} tras eliminar los <i>outliers</i> | 596 |
| D.1. Diagnóstico del modelo lineal de efectos fijos para la variable respuesta log (DO _{ORAI1} /DO _{β-actina}) tras eliminar los <i>outliers</i> | 600 |
| D.2. Diagnóstico del modelo lineal de efectos fijos para la variable respuesta log (DO _{ORAI2} /DO _{β-actina}) tras eliminar los <i>outliers</i> | 607 |
| D.3. Diagnóstico del modelo lineal de efectos fijos para la variable respuesta T (DO _{TRPC1} /DO _{β-actina}) tras eliminar los <i>outliers</i> | 611 |
| D.4. Diagnóstico del modelo lineal de efectos fijos para la variable respuesta log (DO _{STIM1} /DO _{β-actina}) tras eliminar los <i>outliers</i> | 618 |
| D.5. Diagnóstico del modelo lineal de efectos fijos para la variable respuesta T (STIM2) tras eliminar los <i>outliers</i> | 626 |

Listas de Tablas

| | |
|---|-----|
| 2.1. Clasificación de los VOCC. | 63 |
| 3.1. Implicación de los TRPC en el cáncer. | 137 |
| 3.2. Implicación de los TRPV en el cáncer. | 140 |
| 3.3. Implicación de los TRPM en el cáncer. | 143 |
| 5.1. Series GEO seleccionadas. | 165 |
| 5.2. Anticuerpos primarios utilizados. | 166 |
| 7.1. Resumen del filtrado y de los DEG para cada experimento. | 238 |
| 7.2. Análisis de enriquecimiento de genes para la expresión diferencial entre las células HT29 y NCM460 llevada a cabo con el microarray Clariom D Human. | 250 |
| 7.3. Análisis de enriquecimiento de genes para la expresión diferencial entre las células HT29 y NCM460 llevada a cabo con Illumina (RNA-seq). | 251 |
| 7.4. Resumen de los DEG para diferentes contrastes de interés tras el tratamiento con DFMO. | 258 |
| 7.5. Análisis de enriquecimiento de genes para el tratamiento con DFMO en NCM460. | 269 |
| 7.6. Análisis de conjuntos de genes para el tratamiento con DFMO en HT29. | 270 |
| 7.7. Análisis de conjuntos de genes relacionados con el tejido humano de CCR. | 288 |
| 7.8. Series GEO seleccionadas. | 296 |
| 7.9. Resumen del pre-tratamiento de los conjuntos de datos transcriptómicos seleccionados. | 296 |
| 7.10. Resumen de la selección de los modelos para cada conjunto de datos transcriptómicos. | 297 |
| 7.11. Resumen del reanálisis y metaanálisis de expresión diferencial. | 298 |

| | |
|---|-----|
| 7.12. Comparación entre modelos del análisis de expresión diferencial para VOCC. | 299 |
| 7.13. Comparación entre modelos del análisis de expresión diferencial para SOCE. | 301 |
| 7.14. Comparación entre diferentes modelos del análisis de expresión diferencial para TRP. | 303 |
| 7.15. Comparación entre modelos del análisis de expresión diferencial para CPT. | 304 |
| 7.16. Comparación entre diferentes análisis de expresión diferencial para CLC. | 305 |
| 7.17. Comparación entre diferentes análisis de expresión diferencial para MCT. | 306 |
| 7.18. Conjuntos de genes que rechazan alguna hipótesis nula tras el reanálisis y metaanálisis de enriquecimiento. | 308 |
| 7.19. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>calcium</i> | 309 |
| 7.20. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>cancer</i> | 309 |
| 7.21. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>WNT</i> | 310 |
| 7.22. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>mitochondria</i> | 310 |
| 7.23. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>reticulum</i> | 311 |
| 7.24. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>polyamine</i> | 311 |
| 8.1. Comparación entre modelos del análisis de expresión diferencial para VOCC. | 336 |
| 8.2. Comparación entre modelos del análisis de expresión diferencial para SOCE. | 337 |
| 8.3. Comparación entre modelos del análisis de expresión diferencial para TRP. | 338 |
| 8.4. Comparación entre modelos del análisis de expresión diferencial para CPT. | 339 |
| 8.5. Comparación entre modelos del análisis de expresión diferencial para CLC. | 339 |
| 8.6. Comparación entre modelos del análisis de expresión diferencial para MCT. | 340 |
| B.1. Análisis de enriquecimiento de genes para la expresión diferencial entre las células HT29 y NCM460 llevada a cabo con Illumina (RNA-seq). | 417 |
| B.2. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA). | 418 |

| | |
|---|-----|
| B.3. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOMF (CAMERA). | 418 |
| B.4. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA). | 418 |
| B.5. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA). | 419 |
| B.6. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MH (CAMERA). | 420 |
| B.7. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MH (ORA). | 420 |
| B.8. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (CAMERA). | 420 |
| B.9. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA). | 421 |
| B.10. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MOG (CAMERA). | 432 |
| B.11. Porcentaje del mejor ajuste de todos los modelos evaluados. | 445 |
| B.12. Porcentaje del mejor ajuste de todos los modelos evaluados (genes de interés). | 445 |
| B.13. Análisis de conjuntos de genes para la expresión diferencial entre el tejido humano de CCR y el normal. | 456 |
| B.14. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA). | 457 |
| B.15. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (ORA). | 458 |
| B.16. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOMF (CAMERA). | 458 |
| B.17. Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (CAMERA). | 459 |
| B.18. Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (ORA). | 459 |
| B.19. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA). | 459 |
| B.20. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA). | 466 |

| | |
|--|-----|
| B.21. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MH (CAMERA). | 471 |
| B.22. Porcentaje del mejor ajuste de todos los modelos evaluados. | 472 |
| B.23. DEG de la familia VOCC (NCM460 DFMO vs. Control). | 475 |
| B.24. DEG de la familia VOCC (HT29 DFMO vs. Control). | 475 |
| B.25. DEG de la familia SOCE (NCM460 DFMO vs. Control). | 475 |
| B.26. DEG de la familia SOCE (HT29 DFMO vs. Control). | 475 |
| B.27. DEG de la familia CLC (NCM460 DFMO vs. Control). | 475 |
| B.28. DEG de la familia CLC (HT29 DFMO vs. Control). | 475 |
| B.29. DEG de la familia CP (NCM460 DFMO vs. Control). | 476 |
| B.30. DEG de la familia CP (HT29 DFMO vs. Control). | 476 |
| B.31. DEG de la familia MCT (NCM460 DFMO vs. Control). | 476 |
| B.32. DEG de la familia MCT (HT29 DFMO vs. Control). | 476 |
| B.33. DEG de la familia TRP (NCM460 DFMO vs. Control). | 477 |
| B.34. DEG de la familia TRP (HT29 DFMO vs. Control). | 477 |
| B.35. Análisis de enriquecimiento de genes para el tratamiento con DFMO en NCM460. | 479 |
| B.36. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA). | 480 |
| B.37. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA). | 480 |
| B.38. Análisis de conjuntos de genes para el tratamiento con DFMO en HT29. | 483 |
| B.39. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA). | 484 |
| B.40. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (ORA). | 484 |
| B.41. Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (CAMERA). | 484 |
| B.42. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA). | 484 |
| B.43. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA). | 486 |
| B.44. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA). | 490 |

| | |
|--|-----|
| B.45. Análisis de enriquecimiento de genes para la expresión diferencial entre las células HT29 y NCM460 llevada a cabo con el microarray Clariom D Human. | 492 |
| B.46. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA). | 493 |
| B.47. Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (CAMERA). | 493 |
| B.48. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA). | 493 |
| B.49. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA). | 495 |
| B.50. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA). | 498 |
| B.51. Metaanálisis del conjunto de genes VOCC. | 501 |
| B.52. Metaanálisis del conjunto de genes SOCE. | 501 |
| B.53. Metaanálisis del conjunto de genes TRP. | 502 |
| B.54. Metaanálisis del conjunto de genes CP. | 503 |
| B.55. Metaanálisis del conjunto de genes CLC. | 503 |
| B.56. Metaanálisis del conjunto de genes MCT. | 504 |
| B.57. Total de conjuntos de genes que rechazan alguna de las hipótesis tras el reanálisis y metaanálisis de enriquecimiento de genes. | 506 |
| B.58. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>calcium</i> . | 507 |
| B.59. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>cancer</i> . | 508 |
| B.60. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>WNT</i> . | 509 |
| B.61. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>mitochondria</i> . | 510 |
| B.62. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>reticulum</i> . | 511 |
| B.63. Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés <i>polyamine</i> . | 512 |
| B.64. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA). | 513 |

| | |
|---|-----|
| B.65. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (ORA). | 516 |
| B.66. Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOMF (CAMERA). | 516 |
| B.67. Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (CAMERA). | 516 |
| B.68. Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (ORA). | 517 |
| B.69. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA). | 517 |
| B.70. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA). | 533 |
| B.71. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MH (CAMERA). | 541 |
| B.72. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (CAMERA). | 542 |
| B.73. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA). | 542 |
| B.74. Análisis de enriquecimiento de los conjuntos de genes de la base de datos MOG (CAMERA). | 542 |
| C.1. Selección de modelos para la variable respuesta $\log(\text{AUC Depositos})$ con cAIC. | 547 |
| C.2. Parte aleatoria de los modelos problemáticos para $\log(\text{AUC Depositos})$ | 548 |
| C.3. Parte aleatoria de los modelos problemáticos para $\log(\text{AUC Depositos})$ | 549 |
| C.4. Parte aleatoria de los modelos problemáticos para $\log(\text{AUC Depositos})$ | 550 |
| C.5. Parte aleatoria del modelo seleccionado para AUC Depositos | 551 |
| C.6. Selección de modelos para la variable respuesta $\log(\text{AUC Depositos})$ con AIC. | 551 |
| C.7. Supuestos del modelo seleccionado. | 552 |
| C.8. Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto seleccionado para $\log(\text{AUC Depositos})$ | 555 |
| C.9. Estimaciones de la varianza de la parte aleatoria del modelo mixto seleccionado para $\log(\text{AUC Depositos})$ | 555 |
| C.10. Covarianzas entre los coeficientes del modelo mixto para $\log(\text{AUC Depositos})$ | 556 |

| | |
|--|-----|
| C.11. Estimaciones de la variable respuesta $\log(\text{AUC Depósitos})$ con el modelo lineal mixto seleccionado. | 556 |
| C.12. Diferencias dos a dos entre las estimaciones de las diferentes condiciones experimentales, $\log(\text{AUC Depósitos})$, con el modelo lineal mixto seleccionado. | 558 |
| C.13. Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto <i>Bootstrap</i> para $\log(\text{AUC Depósitos})$ | 561 |
| C.14. Diferencias dos a dos entre las estimaciones obtenidas para las diferentes condiciones experimentales mediante el del modelo lineal mixto <i>Bootstrap</i> para $\log(\text{AUC Depósitos})$ | 561 |
| C.15. Selección de modelos para la variable respuesta $\log(\text{MaxDepositos})$ con cAIC. | 562 |
| C.16. Parte aleatoria de los modelos problemáticos para $\log(\text{MaxDepositos})$ | 563 |
| C.17. Parte aleatoria de los modelos problemáticos para $\log(\text{MaxSOCE})$ | 564 |
| C.18. Parte aleatoria de los modelos problemáticos para $\log(\text{MaxSOCE})$ | 565 |
| C.19. Parte aleatoria del modelo seleccionado para MaxDepositos | 565 |
| C.20. Selección de modelos para la variable respuesta $\log(\text{MaxDepositos})$ con AIC. | 566 |
| C.21. Supuestos del modelo seleccionado (p valores). | 568 |
| C.22. Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para $\log(\text{MaxDepositos})$ | 568 |
| C.23. Covarianzas entre los coeficientes del modelo final para $\log(\text{MaxDepositos})$ | 568 |
| C.24. Estimaciones de la varianza de la parte aleatoria del modelo mixto final para $\log(\text{MaxDepositos})$ | 568 |
| C.25. Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para $\log(\text{MaxDepositos})$ | 569 |
| C.26. Estimaciones de MaxDepositos con el modelo seleccionado. | 569 |
| C.27. Diferencias dos a dos entre las condiciones experimentales de sus valores $\log(\text{MaxDepositos})$ estimados con el modelo seleccionado. | 569 |
| C.28. Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto <i>Bootstrap</i> | 571 |
| C.29. Covarianzas entre los coeficientes del modelo mixto <i>Bootstrap</i> | 571 |
| C.30. Diferencias dos a dos entre las condiciones experimentales del modelo lineal mixto <i>Bootstrap</i> | 571 |
| C.31. Selección de modelos para la variable respuesta $\log(\text{AUCSOCE})$ con cAIC. | 572 |
| C.32. Parte aleatoria de los modelos problemáticos para $\log(\text{AUCSOCE})$ | 573 |

| | |
|---|-----|
| C.33. Parte aleatoria de los modelos problemáticos para $\log(\text{AUCSOCE})$ | 574 |
| C.34. Parte aleatoria del modelo seleccionado para AUCSOCE | 575 |
| C.35. Selección de modelos para la variable respuesta $\log(\text{AUCSOCE})$ con AIC | 575 |
| C.36. Supuestos del modelo seleccionado (p valores). | 576 |
| C.37. Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para $T(\text{AUCSOCE})$ | 578 |
| C.38. Covarianzas entre los coeficientes del modelo final para $T(\text{AUCSOCE})$ | 578 |
| C.39. Estimaciones de la varianza de la parte aleatoria del modelo mixto final para $T(\text{AUCSOCE})$ | 578 |
| C.40. Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para $T(\text{AUCSOCE})$ | 579 |
| C.41. Estimaciones de AUCSOCE con el modelo seleccionado. | 579 |
| C.42. Diferencias dos a dos entre las condiciones experimentales de sus valores $T(\text{AUCSOCE})$ estimados con el modelo seleccionado. | 579 |
| C.43. Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto <i>Bootstrap</i> | 581 |
| C.44. Covarianzas entre los coeficientes del modelo mixto <i>Bootstrap</i> | 581 |
| C.45. Diferencias dos a dos entre las condiciones experimentales del modelo lineal mixto <i>Bootstrap</i> | 581 |
| C.46. Selección de modelos para la variable respuesta $\log(\text{MaxSOCE})$ con cAIC | 582 |
| C.47. Parte aleatoria de los modelos problemáticos para $\log(\text{MaxSOCE})$ | 583 |
| C.48. Parte aleatoria de los modelos problemáticos para $\log(\text{MaxSOCE})$ | 584 |
| C.49. Parte aleatoria del modelo seleccionado para MaxSOCE | 585 |
| C.50. Selección de modelos para la variable respuesta $\log(\text{MaxSOCE})$ con AIC | 585 |
| C.51. Supuestos del modelo seleccionado (p valores). | 586 |
| C.52. Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para $\log(\text{MaxSOCE})$ | 588 |
| C.53. Covarianzas entre los coeficientes del modelo final para $\log(\text{MaxSOCE})$ | 588 |
| C.54. Estimaciones de la varianza de la parte aleatoria del modelo mixto final para $\log(\text{MaxSOCE})$ | 588 |
| C.55. Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para $\log(\text{MaxSOCE})$ | 589 |

| | |
|--|-----|
| C.56. Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado sin interacción para $\log(\text{MaxSOCE})$. | 589 |
| C.57. Estimaciones de MaxSOCE con el modelo seleccionado. | 589 |
| C.58. Diferencias dos a dos entre las condiciones experimentales de sus valores $\log(\text{MaxSOCE})$ estimados con el modelo seleccionado. | 589 |
| C.59. Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto <i>Bootstrap</i>. | 592 |
| C.60. Covarianzas entre los coeficientes del modelo mixto <i>Bootstrap</i>. | 592 |
| C.61. Diferencias dos a dos entre las condiciones experimentales del modelo lineal mixto <i>Bootstrap</i>. | 592 |
| C.62. Selección de modelos para la variable respuesta $[\text{Ca}^{2+}]_{\text{cit}}$ basal con cAIC. | 593 |
| C.63. Parte aleatoria del modelo seleccionado para $[\text{Ca}^{2+}]_{\text{cit}}$ basal. | 594 |
| C.64. Selección de modelos para la variable respuesta $[\text{Ca}^{2+}]_{\text{cit}}$ basal con AIC. | 594 |
| C.65. Supuestos del modelo seleccionado (p valores). | 595 |
| C.66. Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para T (CaBasal). | 597 |
| C.67. Covarianzas entre los coeficientes del modelo final para T (CaBasal). | 597 |
| C.68. Estimaciones de la varianza de la parte aleatoria del modelo mixto final para T (CaBasal). | 597 |
| C.69. Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para T (CaBasal). | 597 |
| C.70. Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado sin interacción para $\log(\text{CaBasal})$. | 598 |
| C.71. Estimaciones de $[\text{Ca}^{2+}]_{\text{cit}}$ basal con el modelo seleccionado. | 598 |
| C.72. Contrastes de interés (T (CaBasal)) estimados con el modelo seleccionado. | 598 |
| D.1. Supuestos del modelo seleccionado (p valores). | 601 |
| D.2. Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo seleccionado para $\log(\text{DO}_{\text{ORAI1}}/\text{DO}_{\beta\text{-actina}})$. | 601 |
| D.3. Covarianzas entre los coeficientes del modelo seleccionado para $\log(\text{DO}_{\text{ORAI1}}/\text{DO}_{\beta\text{-actina}})$. | 601 |
| D.4. Tabla ANOVA (tipo II) del modelo seleccionado para $\log(\text{DO}_{\text{ORAI1}}/\text{DO}_{\beta\text{-actina}})$. | 602 |
| D.5. Estimaciones de la expresión de ORAI1 ($\text{DO}_{\text{ORAI1}}/\text{DO}_{\beta\text{-actina}}$) con el modelo seleccionado. | 602 |

| | |
|---|-----|
| D.6. Diferencias dos a dos entre las condiciones experimentales de sus valores $\log(DO_{ORAI1}/DO_{\beta-actina})$ estimados con el modelo seleccionado. | 602 |
| D.7. Selección de modelos para la variable respuesta $\log(ORAI2)$ con cAIC. | 603 |
| D.8. Parte aleatoria de los modelos problemáticos para $\log(DO_{ORAI2}/DO_{\beta-actina})$. | 604 |
| D.9. Parte aleatoria de los modelos problemáticos para $\log(DO_{ORAI2}/DO_{\beta-actina})$. | 605 |
| D.10. Parte aleatoria del modelo seleccionado para $\log(DO_{ORAI2}/DO_{\beta-actina})$. | 606 |
| D.11. Selección de modelos para la variable respuesta $\log(ORAI2)$ con AIC, BIC y LR. | 606 |
| D.12. Supuestos del modelo seleccionado (p valores). | 608 |
| D.13. Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para $\log(DO_{ORAI2}/DO_{\beta-actina})$. | 608 |
| D.14. Covarianzas entre los coeficientes del modelo final para $\log(DO_{ORAI2}/DO_{\beta-actina})$. | 608 |
| D.15. Estimaciones de la varianza de la parte aleatoria del modelo mixto final para $\log(DO_{ORAI2}/DO_{\beta-actina})$. | 608 |
| D.16. Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado para $\log(DO_{ORAI2}/DO_{\beta-actina})$. | 609 |
| D.17. Estimaciones de la expresión de ORAI2 ($DO_{ORAI2}/DO_{\beta-actina}$) con el modelo seleccionado. | 609 |
| D.18. Supuestos del modelo seleccionado (p valores). | 610 |
| D.19. Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo seleccionado para T ($DO_{TRPC1}/DO_{\beta-actina}$). | 611 |
| D.20. Covarianzas entre los coeficientes del modelo seleccionado para T ($DO_{TRPC1}/DO_{\beta-actina}$). | 612 |
| D.21. Tabla ANOVA (tipo II) del modelo seleccionado para T ($DO_{TRPC1}/DO_{\beta-actina}$). | 612 |
| D.22. Estimaciones de la expresión de TRPC1 ($DO_{TRPC1}/DO_{\beta-actina}$) con el modelo seleccionado. | 612 |
| D.23. Diferencias dos a dos entre las condiciones experimentales de sus valores T ($DO_{TRPC1}/DO_{\beta-actina}$) estimados con el modelo seleccionado. | 613 |
| D.24. Selección de modelos para la variable respuesta $\log(STIM1)$ con cAIC. | 614 |
| D.25. Parte aleatoria de los modelos problemáticos para $\log(DO_{STIM1}/DO_{\beta-actina})$. | 615 |
| D.26. Parte aleatoria de los modelos problemáticos para $\log(DO_{STIM1}/DO_{\beta-actina})$. | 616 |
| D.27. Parte aleatoria de los modelos problemáticos para $\log(DO_{STIM1}/DO_{\beta-actina})$. | 617 |
| D.28. Selección de modelos para la variable respuesta $\log(STIM1)$ con AIC, BIC y LR. | 617 |
| D.29. Supuestos del modelo seleccionado (p valores). | 619 |

| | |
|--|-----|
| D.30. Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo seleccionado para $\log(DO_{STIM1}/DO_{\beta-actina})$ | 619 |
| D.31. Covarianzas entre los coeficientes del modelo seleccionado para $\log(DO_{STIM1}/DO_{\beta-actina})$ | 619 |
| D.32. Tabla ANOVA (tipo II) del modelo seleccionado para $\log(DO_{STIM1}/DO_{\beta-actina})$ | 620 |
| D.33. Estimaciones de la expresión de STIM1 ($DO_{STIM1}/DO_{\beta-actina}$) con el modelo seleccionado para $\log(DO_{STIM1}/DO_{\beta-actina})$ | 620 |
| D.34. Diferencias dos a dos entre las condiciones experimentales de sus valores $\log(DO_{STIM1}/DO_{\beta-actina})$ estimados con el modelo seleccionado. | 620 |
| D.35. Selección de modelos para la variable respuesta $\log(STIM2)$ con cAIC. | 621 |
| D.36. Parte aleatoria de los modelos problemáticos para $\log(DO_{STIM2}/DO_{\beta-actina})$ | 622 |
| D.37. Parte aleatoria de los modelos problemáticos para $\log(DO_{STIM2}/DO_{\beta-actina})$ | 623 |
| D.38. Parte aleatoria del modelo seleccionado para STIM2. | 624 |
| D.39. Selección de modelos para la variable respuesta T(STIM2) con AIC, BIC y LR. | 624 |
| D.40. Supuestos del modelo seleccionado (p valores). | 625 |
| D.41. Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para T (STIM2). | 627 |
| D.42. Covarianzas entre los coeficientes del modelo seleccionado para T ($DO_{STIM2}/DO_{\beta-actina}$). | 627 |
| D.43. Estimaciones de la varianza de la parte aleatoria del modelo mixto final para T (STIM2). | 627 |
| D.44. Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para T (STIM2). | 627 |
| D.45. Estimaciones de la expresión de STIM2 ($DO_{STIM2}/DO_{\beta-actina}$) con el modelo lineal seleccionado. | 628 |
| D.46. Diferencias dos a dos entre las condiciones experimentales de sus valores T ($DO_{STIM2}/DO_{\beta-actina}$) estimados con el modelo seleccionado. | 628 |

Lista de Abreviaturas

- $[Ca^{2+}]_{cit}$: concentración de Ca^{2+} citosólica
- $[Ca^{2+}]_i$: concentración de Ca^{2+} intracelular
- $[Ca^{2+}]_{MT}$: concentración de Ca^{2+} en la mitocondria
- $[Ca^{2+}]_{RE}$: concentración de Ca^{2+} en el retículo endoplásmico
- $\Delta\Psi$: potencial de membrana mitocondrial
- Ach: acetilcolina
- ACP: análisis de componentes principales
- ADN: ácido desoxirribonucleico
- AIC: criterio de información de Akaike
- APC: poliposis coli adenomatosa (del inglés, *adenomatous polyposis coli*).
- GSK3 β : Quinasa Glucogeno Sintasa β (del inglés, *glycogen synthase kinase 3 β*)
- ARC: canales de ácido araquidónico (del inglés, *arachidonic acid channel*)
- ARN: ácido ribonucleico (del inglés, *ribonucleic acid*)
- ATP: adenosín trifosfato
- AUC: área bajo la curva (del inglés, *area under curve*)
- BLUE: estimadores lineales e insesgados de mínima varianza
- CaM: calmodulina
- CaN: calcineurina
- CBD: dominio de unión a CaM

-
- CCE: entrada capacitativa de Ca^{2+} (del inglés, *capacitative Ca^{2+} entry*)
 - CCR: cáncer de colon
 - CICR: liberación de Ca^{2+} inducida por Ca^{2+} (del inglés, *calcium-induced calcium release*)
 - CK1: quinasa Caseína 1 (del inglés, *casein kinase 1*)
 - CNGC: canales operados por nucleótidos cíclicos (del inglés, *cyclic nucleotide-gated channel*)
 - CP: bomba de Ca^{2+} (del inglés, *calcium pump*)
 - CPT: bombas y transportadores de Ca^{2+}
 - CPA: ácido ciclopiazónico (del inglés, *cyclopiazonic acid*)
 - CRAC: canal activado por la liberación de Ca^{2+} (del inglés, *Ca^{2+} release-activated channel*)
 - CTC: células tumorales circulantes
 - CTL: células T citotóxicas (del inglés, *cytotoxic T cell*)
 - CV: coeficiente de variación
 - CycD: ciclina D
 - DAG: diacilglicerol (del inglés, *diacylglycerol*)
 - DFMO: difluorometilornitina
 - dNTP: desoxinucleótido trifosfato
 - DO: densidad óptica
 - EDTA: ácido etilenglicol-diamino-tetra-acético
 - EGTA: ácido etilenglicol-bis-(b-aminoetil-eter)-N,N,N',N'-tetra-acético
 - EMRE: regulador esencial de MCU (del inglés, *essential MCU regulator*)
 - ER: receptor de estrógenos (del inglés, *estrogen receptor*)
 - ERC: conjunto de genes de canales de liberación de Ca^{2+} desde el retículo endoplasmático
 - ETON: dominio N-terminal transmembrana extendido
 - FBS: suero bovino fetal (del inglés, *fetal bovine serum*)
 - GFP: proteína fluorescente verde (del inglés, *green fluorescent protein*)

-
- GPCR: receptores acoplados a proteínas G (del inglés, *G protein-coupled receptors*)
 - HVA: alto umbral de activación (del inglés, *high voltage activation*)
 - iGluR: receptor ionotrópico de glutamato
 - IM: motivo IQ modificado
 - IP3: inositol 1,4,5-trisfosfato (del inglés, *inositol 1,4,5-triphosphate*)
 - IP3R: receptor de Inositol 1,4,5-trisfosfato (del inglés, *inositol 1,4,5-triphosphate receptor*)
 - LGIC: receptores ionotrópicos u operados por ligando (del inglés, *ligand-gated ion channels*)
 - LTD: depresión a corto plazo (del inglés, *long-term depression*)
 - LTP: potenciación a largo plazo (del inglés, *long-term potentiation*)
 - LVA: bajo umbral de activación (del inglés, *low voltage activation*)
 - mAChR: receptor muscarínico de acetilcolina
 - MAM: membranas asociadas a las mitocondrias (del inglés, *mitochondria-associated membranes*)
 - MCD: determinante de mínima covarianza
 - MCG: mínimos cuadrados generalizados
 - MCO: mínimos cuadrados ordinarios
 - MCP: mínimos cuadrados ponderados
 - MCT: conjunto de genes de transporte de Ca^{2+} mitocondrial
 - MCU: uniportador de Ca^{2+} mitocondrial (del inglés, *mitochondrial Ca^{2+} uniporter*)
 - MDSC: células supresoras derivadas de origen mieloide (del inglés, *myeloid-derived suppressor cells*)
 - MEC: medio externo completo
 - mGluR: receptor muscarínico de glutamato
 - mHCX: intercambiador $\text{H}^+/\text{Ca}^{2+}$ mitocondrial (del inglés, *mitochondrial $\text{H}^+/\text{Ca}^{2+}$ exchanger*)
 - MICU: carga de Ca^{2+} mitocondrial (del inglés, *mitochondrial Ca^{2+} uptake*)
 - MME: membrana mitocondrial externa

-
- MMI: membrana mitocondrial interna
 - MP: membrana plasmática
 - MVE: elipsoide de mínimo volumen
 - nAChR: receptor nicotínico de acetilcolina
 - NCKX: intercambiador $\text{Na}^+/\text{Ca}^{2+} - \text{K}^+$ (del inglés, *$\text{Na}^+/\text{Ca}^{2+} - \text{K}^+$ exchanger*)
 - NCLX: intercambiador $\text{Na}^+/\text{Li}^{2+}/\text{Ca}^{2+}$ (del inglés, *$\text{Na}^+/\text{Li}^{2+}/\text{Ca}^{2+}$ exchanger*)
 - NCX: intercambiador $\text{Na}^+/\text{Ca}^{2+}$ (del inglés, *mitochondrial $\text{Na}^+/\text{Ca}^{2+}$ exchanger*)
 - NFAT: factor nuclear de células T activadas (del inglés, *nuclear factor of activated T-cells*)
 - NGS: secuenciación de nueva generación (del inglés, *next-generation sequencing*)
 - NK: *natural killer*
 - NMDAR: receptor de N-metil-D-aspartato (del inglés, *N-methyl-D-aspartate receptor*)
 - NR: receptores nucleares (del inglés, *nuclear receptor*)
 - OCT: optimal cutting temperature
 - ODC: ornitina descarboxilasa
 - PBS: tampón fosfato salino (del inglés, *phosphate buffered saline*)
 - PC: componente/s principal/es
 - PD-1: muerte celular programada 1 (del inglés, *programmed death 1*)
 - PIP2: fosfatidilinositol 4,4-bifosfato (del inglés, *inositol 1,4-bisphosphate*)
 - PKB: proteína quinasa B
 - PLC: fosfolipasa C (del inglés, *phospholipase C*)
 - PMCA: ATPasa de Ca^{2+} de la membrana plasmática (del inglés, *plasma membrane Ca^{2+} ATPase*)
 - PTP: *PicoTiterPlate*
 - PTPm: poro de permeabilidad transitoria mitocondrial (del inglés, *mitochondrial permeability transition pore*)
 - qPCR: PCR cuantitativa (del inglés, *quantitative PCR*)

-
- RIN: número de integridad del ARN (del inglés, *RNA integrity number*)
 - RMO: *Robust Multi-array Average*
 - ROCC: canales de Ca^{2+} operados por receptor (del inglés, *receptor-operated calcium channel*)
 - ROS: especies reactivas de oxígeno (del inglés, *reactive oxygen species*)
 - RTK: receptor tirosín quinasa (del inglés, *receptor tyrosine kinase*)
 - RyR: receptores de rianodina (del inglés, *ryanodine receptor*)
 - SERCA: ATPasa de Ca^{2+} del retículo endoplásmico (del inglés, *endoplasmic reticulum Ca^{2+} ATPase*)
 - SMOCC: canales de Ca^{2+} operados por segundos mensajeros (del inglés, *second messenger-operated Ca^{2+} channel*)
 - SOCC: canales de Ca^{2+} operados por depósitos (del inglés, *store-operated Ca^{2+} channel*)
 - SOCE: entrada de calcio operada por depósitos (del inglés, *store-operated calcium entry*)
 - SPCA: ATPasa de Ca^{2+} de la vía secretora (del inglés, *secretory pathway Ca^{2+} ATPase*)
 - SSE: suma de cuadrados explicada por el modelo
 - SSR: suma de cuadrados residual
 - STIM: molécula de interacción estromal (del inglés, *stromal interaction molecule*)
 - STT: suma de cuadrados totales
 - TM: dominio transmembrana
 - TRP: receptor de potencial transitorio (del inglés, *transient receptor potential*)
 - TRPA: receptor de potencial transitorio anquirina (del inglés, *transient receptor potential ankyrin*)
 - TRPC: receptor de potencial transitorio canónico (del inglés, *transient receptor potential canonical*)
 - TRPM: receptor de potencial transitorio melastatina (del inglés, *transient receptor potential melastatin*)
 - TRPML: receptor de potencial transitorio mucolipina (del inglés, *transient receptor potential mucolipin*)

-
- TRPP: receptor de potencial transitorio policistina (del inglés, *transient receptor potential polycystin*)
 - TRPV: receptor de potencial transitorio vaniloide (del inglés, *transient receptor potential vanilloid*)
 - VDAC: canales aniónicos dependientes de voltaje (del inglés, *voltage-dependent anion channel*)
 - VOCC: canales de Ca²⁺ operados por voltaje (del inglés, *voltage-operated calcium channel*)

Resumen

El cáncer de colon, en 2020, ocupó, a nivel mundial, el tercer puesto en incidencia y el segundo en mortalidad. Se han descrito 10 características o hallmarks que distinguen a las células normales de las tumorales: autonomía proliferativa, resistencia a la muerte, capacidad migratoria y metastásica. La forma más frecuente de aparición de este es a través de alteraciones en la ruta Wnt/ β -catenina, lo que puede dar lugar a un aumento de la síntesis de poliaminas las cuales, en condiciones normales, participan en diferentes procesos como la restitución epitelial e, incluso, en la regulación de la homeostasis del calcio intracelular. Es más, algunos de los hallmarks del cáncer están estrechamente relacionados con los sistemas de transporte del calcio intracelular. El calcio intracelular es un ion sumamente importante en numerosos procesos celulares, tales como la contracción muscular, proliferación celular, muerte celular, migración celular... Por ello, los sistemas de transporte intracelular de este ion han de estar celosamente regulados. Existen diferentes mecanismos celulares de transporte de calcio, aparte de depósitos intracelulares de este con sus correspondientes sistemas de transporte.

En las células epiteliales, como son los colonocitos, el sistema de transporte más importante de calcio es la “Entrada Capacitativa de Calcio”, desregulada en el cáncer de colon. No obstante, aunque a nivel funcional se ha descrito que los sistemas de transporte de calcio están desregulados, los mecanismos moleculares que subyacen a estos cambios todavía no se conocen con detalle. Por todo ello, la presente tesis trata de aclarar algunas de las bases moleculares implicadas en la transformación tumoral y en las características de estas células, para lo que se ha empleado el análisis transcriptómico en diferentes modelos de colon normal y tumoral humano.

En primer lugar, en modelos celulares se ha estudiado la homeostasis del calcio intracelular tanto a nivel transcriptómico, para lo que se han empleado diferentes técnicas de alto rendimiento, como a nivel funcional mediante la técnica de imagen de calcio. En estos mismos modelos, y de un modos similar, se estudio el efecto sobre la homeostasis del calcio intracelular de la depleción de poliaminas inducida por el tratamiento con DFMO, un inhibidor suicida de la síntesis de poliaminas. Posteriormente, para estudiar el transcriptoma de modelos más cercanos a la realidad del cáncer de colon, se estudió el transcriptoma de muestras pareadas de cáncer de colon y normal donadas por 10 pacientes diferentes junto con la expresión de proteínas seleccionadas por western blotting. Finalmente, a partir del reanálisis de diferentes experimentos transcriptómicos extraídos de repositorios públicos junto con los obtenidos por nuestro grupo se realizó un metaanálisis de

más de 1000 microarrays de muestras de tejido humano de cáncer de colon y normal, pues estos resultados reflejarán más fielmente la realidad de los cambios transcriptómicos que subyacen al cáncer de colon humano.

De los modelos celulares se obtuvo que, en el cáncer de colon, los mecanismos de transporte de calcio están profundamente alterados tanto a nivel transcriptómico (56 genes diferencialmente expresados) como a nivel funcional, pues se ha observado que las células tumorales, respecto de las normales, presentan una mayor concentración de calcio libre intracelular, un menor tamaño de los depósitos y una mayor entrada capacitativa.

El análisis transcriptómico triple realizado ha detectado la sobreexpresión de 24 genes implicados en transporte de calcio en las células tumorales: los canales *CAVI.3*, *ORAI2* (y sus reguladores *STIM1*, *MBP*, y las *SEPTIN2,4,7,9-11*), *TRPC5*, *TRPV6*, *TRPM7* y *TRPP1*, las bombas *PMCA1*, *SERCA2*, *SPCA2*, el intercambiador *NCX2*, los canales intracelulares *IP3R1* e *IP3R3*, el uniportador de calcio mitocondrial *MCU*, su modulador *MICU1*, *MCUR1*, y el canal *VDAC2* de la membrana mitocondrial externa, junto con la infraexpresión de 32 genes incluyendo los canales *CAVI.2*, *CAV2.2* y *CAV2.3*, *ORAI1* (y sus moduladores *ORMDL3*, *SARAF*, *CRACR2A* y las *SEPTINI,3* y *6*), *TRPC7*, *TRPP2,3* y *5*, *TRPM2,3,5* y *8*, *TRPV1*, *TRPA1* y *TRPML1,2* y *3*, las bombas *PMCA4* y *SPCA1*, los canales intracelulares *IP3R2* y *RyR2*, los reguladores negativos de la captación de Ca^{2+} mitocondrial *MICU2* y *MCUb*, los canales de la membrana mitocondrial externa *VDAC1* y *3* y el intercambiador mitocondrial *NCLX*.

En cuanto al tratamiento con DFMO sobre estos mismos modelos, se observó un efecto casi selectivo sobre las células tumorales frente a las normales. Por un lado, a nivel transcriptómico se han identificado más de 4000 genes diferencialmente expresados en las células tumorales frente a los, únicamente, 165 identificados en las normales. De estos, en las células tumorales se han identificado 16 relacionados con la homeostasis del calcio intracelular y sólo 1 en las normales. Cabe destacar que de estos 16 genes diferencialmente expresados en las células tumorales, 10 revierten la expresión diferencial mostrada entre las células tumorales y normales en condiciones control. Específicamente, DFMO disminuye la expresión de *TRPC5*, *TRPV6*, *TRPP1* y *SPCA2* que estaban sobreexpresados en las células tumorales, a la vez que aumenta la expresión de *CRACR2A*, *ORMDL3*, *SEPTIN6*, *TRPP2*, *PMCA4* y *VDAC3* que estaban infraexpresados en las células tumorales. En cambio, en las células normales, DFMO solo modula la expresión de un único gen, el canal intracelular *IP3R1*.

El análisis en muestras pareadas procedentes de 10 pacientes muestra, como era de esperar, un menor número de cambios en el transcriptoma, destacando los cambios consistentes con el aumento de la captación mitocondrial de calcio en el fenotipo tumoral. Finalmente, el metaanálisis, mostró un evidente remodelado de la homeostasis del calcio intracelular que difería en buena medida de los cambios observados en los modelos celulares y las muestras pareadas. Los cambios coinciden en 19 genes incluyendo *CAVI.3*, *CAV3.2*, *ORAI2*, *SARAF*, las *SEPTINI,2,4,6,7,9*, *TRPV1*, *TRPML1,2*, *TRPM5*, *TRPM8*, *TRPA1*, *PMCA4*, *MICU2* y *NCLX*. Respecto a las muestras pareadas, los datos del metaanálisis coinciden en solo 10 genes incluyendo el canal específico de tejido *MS4A12*, los

moduladores de SOCE *STIMI* y las *SEPTINI,2,6* y *7*, los canales *TRPM1*, la bomba *SPCA1*, el canal del retículo *IP3R1* y el canal mitocondrial *VDAC1*. Finalmente, en los tres análisis llevados a cabo solo coinciden 4 genes, las *SETPINI,2,6* y *7*.

En conclusión, tanto en los modelos celulares como en los de tejido humano, el fenotipo tumoral está asociado con una evidente dishomeostasis del calcio intracelular que, curiosamente, difiere notablemente entre los distintos modelos. Además, la depleción de poliaminas afecta casi en exclusiva a las células tumorales pero no a las normales, de tal manera que consigue revertir parcialmente el fenotipo tumoral tanto a nivel transcripómico como a nivel funcional.

Parte I

Introducción

“Desde mi punto de vista, la investigación del cáncer es absolutamente apartidista. El cáncer es muy democrático en el sentido de que ataca a las personas independientemente de su raza, su género, su origen nacional o sus convicciones políticas”.

David H. Koch

“El calcio ni se crea ni se destruye, sólo se transporta”.

Mónica García Durillo

Capítulo 1

Cáncer de colon

En el presente capítulo se va a realizar una visión general sobre el cáncer y sus características para, posteriormente, introducir el cáncer de colon, pues es el centro de la investigación de la presente tesis. Finalmente, se mostrará cómo se relaciona la homeostasis o, mejor dicho, la dishomeostasis del Ca^{2+} intracelular con el cáncer de colon.

1.1. El cáncer

En la literatura puede encontrarse numerosas definiciones para tumor, neoplasia o cáncer. Por ejemplo, la RAE define tumor como “hinchazón y bulto que se forma anormalmente en alguna parte del cuerpo de un animal” o “masa de células transformadas, con crecimiento y multiplicación anormales,” neoplasia como “multiplicación o crecimiento anormal de células en un tejido del organismo,” y al cáncer como “enfermedad que se caracteriza por la transformación de las células, que proliferan de manera anormal e incontrolada.” A grandes rasgos, el cáncer, del latín *karkinos* (cangrejo) es una enfermedad que tiene su origen en la acumulación de mutaciones a nivel genético, pues en la mayoría de los casos es necesario más de una mutación para dar lugar a un cáncer, a lo largo de la vida de un organismo, las cuales son nocivas para el mismo pero que confieren una clara superioridad estratégica en la supervivencia del tumor en detrimento de su huésped. Otro concepto importante en el contexto tumoral es la metástasis, que en griego significa “cambio de lugar,” es, según la RAE, la “propagación de un foco canceroso en un órgano distinto de aquel en que se inició.”

A pesar de las definiciones anteriores, es importante describir, a nivel fisiológico, qué es una célula cancerígena y qué la diferencia del resto de células. Por ello, a continuación, se van a presentar una serie de características, en concreto diez, que diferencian las células tumorales de las sanas, denominadas sellos o *hallmarks* del cáncer. La bibliografía básica que se ha usado en la exposición de los diferentes *hallmarks* es Hanahan and Weinberg (2000), Hanahan and Weinberg (2011) y Weinberg (2013), por lo que esta se da aquí por citada.

1.2. Sellos o *hallmarks* del cáncer

En 2000, Douglas Hanahan y Robert A. Weinberg propusieron seis características o *hallmarks* que distinguen a una célula tumoral de una eucariota sana (Hanahan and Weinberg (2000)). Sin embargo, en años venideros fueron surgiendo nuevos descubrimientos y avances en el campo de la investigación contra el cáncer. En particular, se demostró la clara participación en el proceso tumoral del microambiente tumoral tanto en su patogenicidad como en la tumorigénesis. Esto supuso un cambio de paradigma, pues a partir de entonces se empezó a considerar al cáncer y al proceso tumorigénico como una serie de procesos patológicos en los que no sólo intervienen las células tumorales, sino también diferentes tipos celulares normales del microambiente tumoral, denominado *estroma tumoral*. Así, a lo largo de todos estos años se ha demostrado en numerosas investigaciones que el estroma tumoral participa activamente en la tumorigénesis.

A consecuencia de ello, en 2011, los mismos autores que plantearon los seis *hallmarks* el cáncer en el año 2000 se vieron en la necesidad de revisarlos y ampliarlos hasta un total de 10 (figura 1.1) (Hanahan and Weinberg (2011)). Cabe destacar que las células tumorales incipientes adquieren estos sellos de forma progresiva a lo largo del proceso tumoral, desde que se dan una serie de eventos que propician que las células adquieran la capacidad tumorigénica hasta que adquieren la condición de malignidad. Así, un cáncer se desarrolla gradualmente a partir de un número creciente de células aberrantes en las que la acumulación de mutaciones y cambios epigenéticos producen errores en los controles normales en división, apoptosis y diferenciación celular, lo que contribuye al desarrollo y progresión del cáncer. A continuación, se resumirán los diez *hallmarks* con el fin de dar una visión más global del cáncer para, posteriormente, poder comprender con mayor profundidad las peculiaridades del cáncer de colon.



Figura 1.1: *Hallmarks* del cáncer. Se han descrito hasta diez características que distinguen a una célula eucariota normal de una cancerígena las cuales son adquiridas por estas células de forma progresiva. De los diez mostrados, seis de ellos fueron descritos por Hanahan and Weinberg (2000) y los cuatro restantes son los denominados *hallmarks* emergentes, descritos una década después también por Hanahan and Weinberg (2011). Estos últimos son la reprogramación metabólica, evasión del sistema inmune, estado inflamatorio inducido por el tumor e inestabilidad genómica y mutación (Modificado de Hanahan and Weinberg (2011)).

1.2.1. Mantenimiento de la señal proliferativa

En condiciones normales, las células de los diferentes tejidos mantienen una tasa de proliferación sumamente controlada y regulada a través de diferentes vías de señalización en las que se liberan y producen diferentes factores que promueven la proliferación y la entrada en el ciclo celular. Entre los diferentes mecanismos de control se encuentran: la inhibición por contacto, por la cual un tejido no puede superar ciertos límites con el fin de que mantenga su arquitectura y homeostasis tisular, así como asegurar un equilibrio entre la proliferación y la apoptosis; el acortamiento de los telómeros; señales inhibitorias del crecimiento ante un daño en el ADN que tenga lugar durante la replicación, donde adquieren una gran importancia genes supresores de tumores tales como *TP53* o *RBI* (gen del *retinoblastoma (Rb)*), que regulan negativamente el proceso de proliferación, etc.

A pesar del complejo entramado de mecanismos que velan por el control y regulación de la tasa de proliferación, las células tumorales son capaces de sortearlos y, por ende, presentan su característica autonomía proliferativa. Esto lo hacen mediante dos estrategias diferentes: a través del mantenimiento de la señal proliferativa o de la evasión de las señales inhibitorias del crecimiento. No obstante, hay ocasiones en que en una misma ruta se dan ambas situaciones en función de qué componentes estén alterados. La mayor parte de estos mecanismos están basados en la regulación de la señalización mediada por factores de crecimiento, encargados de controlar tanto el número de células como su localización en un tejido. Estos actúan sobre los *receptores tirosín quinasa (TKR)* de la superficie celular, por lo que estos se activan y acaban por repercutir en otros procesos tales como la supervivencia celular o el metabolismo, tal y como sucede, por ejemplo, en la ruta de señalización de la quinasa de fosfatidilinositol-3 (PI3K), mostrada en la figura 1.2 (Cheng et al. (2008), Lemmon and Schlessinger (2010), Hanahan and Weinberg (2011), Witsch, Sela, and Yarden (2010)).

En las células tumorales, los mecanismos por los cuales la señalización por factores de crecimiento está sobreactivada pueden ser la sobreexpresión de sus receptores o mutaciones de ganancia de función de estos receptores que les permite estar constitutivamente activos en ausencia de la unión de su ligando. En cuanto a la biodisponibilidad de estos factores de crecimiento, está regulada por mecanismos de secuestro en la matriz extracelular, por lo que la acción de enzimas como las proteasas juega un papel clave en cómo y cuándo se liberan. Es más, aunque en condiciones normales los factores de crecimiento actúen, posiblemente, a través de una señalización paracrina, en el contexto tumoral pueden estimular a las células del estroma para que liberen factores de crecimiento e, incluso, que las propias células tumorales expresen y liberen sus propios factores de crecimiento y, en consecuencia, esta señalización se vuelva autocrina (Cheng et al. (2008), Bhowmick, Neilson, and Moses (2004), Hanahan and Weinberg (2011), Witsch, Sela, and Yarden (2010)).

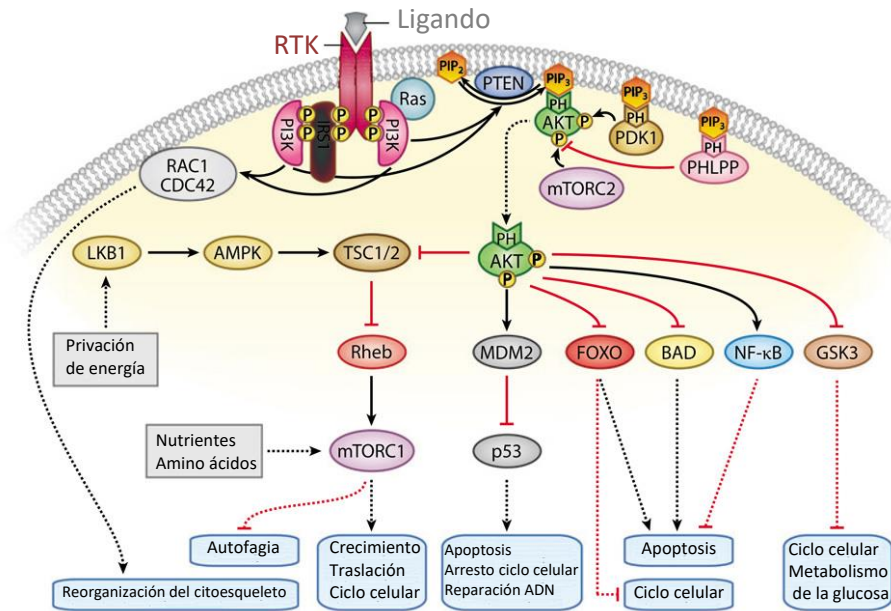


Figura 1.2: Ruta de la quinasa de fosfatidilinositol-3 (PI3K). La activación del TKR por un ligando, como las integrinas, recluta y activa PI3K que, a su vez, fosforila el fosfolípido de membrana fosfatidilinositol-4,5 bisfosfato (PIP₂) para dar PIP₃. Esta molécula es capaz de reclutar proteínas hacia la membrana plasmática (MP) a través de su dominio de unión a pleckstrina (PH), tales como la oncoproteína quinasa de serina/treonina Akt, la proteína quinasa dependiente de 3-fosfoinositida (PDK-1) o la proteína fosfatasa rica en repeticiones de leucina (PHLPP). En este punto, Akt se activa a través de la fosforilación por PDK-1 y del complejo diana de rapamicina de mamíferos insensible a la rapamicina (mTORC2), lo que conlleva, en función del sustrato fosforilado por Akt, la activación de rutas de supervivencia, de crecimiento y proliferación celular, entre otras. Por ejemplo, Akt puede actuar como inhibidor de la apoptosis a través de la inactivación de BAD y procaspasa-9, o activar rutas de supervivencia y proliferativas a través de la inhibición de CHK1 y p21 y activación de CDK2. No obstante, PHLPP puede defosforilar AKT. Cabe destacar que PI3K también es capaz de regular otros mecanismos como, por ejemplo, la reorganización del citoesqueleto a través de RAC/CDC42. Akt también puede inhibir la inactivación del mTORC1 por el complejo de esclerosis tuberosa 1 y 2 (TSC1 y TSC2), si bien es cierto que LKB1 y la proteína quinasa activada por AMP (AMPK) pueden regular de forma positiva TSC1 y TSC2. Abreviaturas: GSK, quinasa glucógeno sintasa; NF- κ B, factor nuclear κ B; RAC1, sustrato 1 de la toxina botulínica C3 relacionado con Ras (modificado de Chalhoub and Baker (2009)).

En cualquier caso, el resultado final es la sobreactivación de la señalización proliferativa desencadenada por TKR. Además, también puede haber mutaciones en diferentes componentes de esta misma ruta de señalización u otras corriente a bajo de estos receptores que les hagan estar constitutivamente activas independientemente de los factores de crecimiento. No obstante, estas son muchas y muy diversas y, por ello, sólo se expondrán brevemente algunas de ellas:

- *Activación de RAS:* una de las consecuencias directas de la activación de los TKR es la activación del transductor de señales RAS. No obstante, en las células tumorales se han descrito diferentes mutaciones que conllevan la sobreactivación de RAS: mutación del dominio de unión a GTP de RAS que favorece su unión a GTP y, por ende, su activación constitutiva; mutación de SOS que, como consecuencia, resulta en la sobreactivación de RAS al favorecer el recambio de GTP, o la mutación de GAP que impide la inactivación de RAS, entre otras (Lemmon and Schlessinger (2010)).

La consecuencia directa de la activación de RAS es el inicio de la activación de la ruta de las *proteínas quinasas activadas por mitógenos* (MAPK, de las siglas en inglés *mitogen-activated protein kinases*) que, como resultado, se activa ERK el cual se dirige al núcleo donde ejerce diferentes acciones que conllevan la activación de la expresión de MYC que, a su vez, activa la expresión de CycD que promueve la entrada en el ciclo celular (Davies and Samuels (2010)). Por lo tanto, las alteraciones de ganancia de función en TKR, en la función de RAS o en las proteínas MAPK, resultan en el incremento del crecimiento y de la proliferación de las células. Así, los requerimientos energéticos de estas aumentan y, por ello, también su metabolismo, lo que favorece aún más la supervivencia y la proliferación celular (Hanahan and Weinberg (2011), Davies and Samuels (2010)).

- *mTOR*: mTOR es un regulador clave en la fisiología celular que favorece el crecimiento y división celular. Esta proteína forma dos complejos proteicos diferentes: mTORC1 y mTORC2. El primero es un complejo soluble que participa en la activación de factores de inicio de la traducción, lo que permite la unión del ARNm a la subunidad pequeña del ribosoma 40S con el fin de iniciar la síntesis proteica y, por ende, favorecer el crecimiento celular. Además, también es el principal regulador del metabolismo y participa en la inhibición de la autofagia.

El segundo, mTORC2, que, a diferencia de mTORC1, se localiza en la cara citoplasmática de la MP gracias a la proteína MSIN1, fosforila a la proteína Akt la cual es un inhibidor de la apoptosis, pues en condiciones normales Akt inactiva BAD y procaspasa-9, y también un activador de rutas de supervivencia y proliferativas ya que inhibe CHK1 y p21 y activa CDK2, entre otras. Curiosamente, Akt también puede inhibir la inactivación del mTORC1 por el complejo de esclerosis tuberosa 1 y 2 (TSC1 y TSC2), si bien es cierto que estos complejos inhibitorios pueden ser regulados positivamente por otras proteínas como LKB1 y la proteína quinasa activada por AMP (AMPK) (figura 1.2). Debido a su papel tanto en el crecimiento como en la proliferación celular y, además, a que se ha encontrado evidencia de su activación por parte de rutas de señalización oncogénicas, mTOR es considerada una diana terapéutica en el cáncer (Chalhoub and Baker (2009), Sudarsanam and Johnson (2010)).

1.2.2. Evasión de las señales inhibitorias del crecimiento

Otra de las estrategias de las células tumorales para adquirir la capacidad de autonomía proliferativa, aparte del mantenimiento de la señal proliferativa, es la evasión de las señales inhibitorias del crecimiento. En este sentido, las células tumorales actúan sobre diferentes mecanismos para evadir estas señales inhibitorias: *p53*, *NF2*, *PTEN*, $TGF\beta$... A continuación, se realizará un breve resumen de los más conocidos:

- *p53*: uno de los mecanismos que emplean las células tumorales es a través de los genes supresores de tumores, como el *TP53*, conocido como el *guardián del genoma*. Este gen es capaz de integrar señales de diferentes vías intracelulares y puede actuar a través de interacciones proteína-proteína o como factor de transcripción. En condiciones normales, *p53* (proteína codificada por *TP53*) es ubiquitinizada por MDM2 para su degradación. Sin embargo, cuando hay daño en el ADN, *p53* se estabiliza: ATM fosforila tanto a *p53* como a *CHK2* la cual, a su vez, también es capaz de fosforilar a *p53*, y, como resultado, MDM2 es incapaz de unirse a *p53* para ubiquitinizarlo. Así, *p53* se acumula en el núcleo y activa diferentes genes encargados de la reparación del ADN y de bloquear el progreso del ciclo celular. Es más, en aquellos casos en los que el daño del material genético es demasiado grave, *p53* es capaz de inducir la apoptosis por la activación de la ruta intrínseca.

En un gran número de tumores se ha encontrado que *TP53* está mutado, silenciado o inactivado, por lo que no puede actuar como supresor tumoral. También se han encontrado mutaciones en ATM que impiden su acción sobre *p53*. Como resultado, se favorece la acumulación de mutaciones en las células tumorales y su supervivencia.

- *NF2*: otro mecanismo basado en los supresores tumorales incluye a *NF2*, también conocido como *Merlín*, que participa en la inhibición por contacto de la proliferación celular. Por lo tanto, la pérdida de función está relacionada con la proliferación y progresión tumoral. En condiciones normales, cuando está presente, junto con α y β -catenina, *NF2* interactúa con las cadherinas para que las células permanezcan unidas y, por ello, es importante en el mantenimiento de la integridad y homeostasis tisular.

Se ha observado que en las células tumorales hay deficiencia de *NF2* la cual se ha relacionado con la liberación de β -catenina. Esta última se acumula en el citoplasma y se estabiliza, por lo que aumenta su concentración citosólica y, en consecuencia, se trasloca al núcleo. Cabe destacar que la β -catenina regula la expresión de genes que promueven la proliferación celular, tales como *MYC* y *CycD*. Además, puesto que se deshacen las uniones intracelulares se favorece la separación de las células tumorales y, por ende, se potencia su capacidad metastásica.

- *Regulación por PTEN*: PTEN es un supresor tumoral cuya función principal es inactivar la vía de señalización de proliferación celular PI3K/Akt. Esta ruta se activa por las proteínas de membrana *integrinas*. Por un lado, el dominio extracelular de estas proteínas reconoce y une proteínas de la matriz extracelular y, por ello, es importante en el mantenimiento de la integridad tisular. Por otro lado, el dominio intracelular actúa como anclaje de elementos del citoesqueleto como FAK, que activa rutas de supervivencia y migración celular. Además, el dominio intracelular también es capaz de transducir señales intracelulares en respuesta a estímulos mecánicos (Chalhoub and Baker (2009), You et al. (2015)).

En condiciones normales, PTEN fosforila a la *quinasa de adhesión focal* (FAK, inglés *focal adhesion kinase*). Posteriormente, FAK recluta a PI3K para fosforilar a PIP2 y generar PIP3 (figura 1.2). Además, PTEN también es capaz de desfosforilar directamente al propio PIP3. A este se le une la oncoproteína Akt para ser activada por PDK1 lo que conlleva la activación de rutas de supervivencia (ver apartado “1.2.3 Resistencia a la muerte”). FAK también es capaz de fosforilar SRC que, a su vez, fosforila FAK, por lo que se forma un sitio de unión con GRB2 al que después se le une SOS y, finalmente, RAS. Así, RAS se activa y se dispara la ruta de las MAPK la cual acaba con la acumulación de ERK en el núcleo y, en consecuencia, se activan genes relacionados con la proliferación como CycD (You et al. (2015)).

En algunos tipos de cáncer se han identificado mutaciones de PTEN lo que conlleva a que las rutas de señalización desencadenadas por las integrinas estén sobreactivadas tras la inducción por estímulos externos. En consecuencia, se evade el control negativo sobre las rutas de señalización FAK, ERK y Akt y, en consecuencia, se activa la migración celular, la supervivencia, el crecimiento y la proliferación de las células tumorales (B.-H. Jiang and Liu (2009), Chalhoub and Baker (2009), You et al. (2015)).

- *Factor de crecimiento tumoral β (TGF β)*: este factor es clave en la patogenicidad de muchos tipos de cáncer. Además, tiene un papel dual, pues es tanto supresor como promotor tumoral. Por un lado, en estadios tempranos del desarrollo tumoral tiene un papel supresor del tumor a través de la ruta de señalización canónica. Este actúa en forma dimérica como ligando activando a su receptor TGF β R2 para, posteriormente, reclutar a su otro receptor TGF β R1 el cual fosforila a los factores de transcripción SMAD2 y SMAD3 que heterotrimerizan con SMAD4 (SMAD(2/3)-SMAD(2/3)-SMAD4) y se traslocan al núcleo, donde reconocen y se unen a regiones concretas del ADN para regular la expresión de genes relacionados con la diferenciación, proliferación, supervivencia, transición epitelio-mesenquimal (EMT, del inglés *epithelial-mesenchymal transition*) y apoptosis (Vander Ark, Cao, and Li (2018)). Por ejemplo, SMAD3 se une al factor supresor E2F4 y E2F5 y se reprime la expresión del oncogén MYC.

Por el contrario, cuando esta ruta $TGF\beta/SMAD$ está inactivada entonces se favorece su ruta no canónica independiente de SMAD la cual activa otras rutas de señalización corriente abajo, como las RAS/MAPK o PI3K/Akt. Por lo tanto, esta ruta promueve la proliferación, diferenciación, apoptosis y transición epitelio mesenquimal; es decir, que esta ruta alternativa favorece el desarrollo de tumores. La activación de la ruta no canónica puede darse bajo diferentes condiciones: reducción en los $TGF\beta R$; reducción de la biodisponibilidad de SMAD2y SMAD3 -fosforilación vía MAPK que previene su activación por $TGF\beta R$ -; mutaciones en SMAD4, o sobreactivación de RAS. Es más, la evidencia muestra que en estadios avanzados de la enfermedad $TGF\beta$ podría tener un papel importante en la progresión y evolución tumoral a través de la ruta de señalización mediada por RAS (Vander Ark, Cao, and Li (2018)).

1.2.3. Resistencia a la muerte

La apoptosis es un proceso fisiológico por el cual se eliminan células que, por lo general, presentan daños críticos que no pueden ser reparados -durante el desarrollo también se da este proceso pero con otros fines-. Durante años se ha estudiado el papel anticancerígeno de este proceso, pues se activa ante muchas de las situaciones de estrés que experimentan las células cancerígenas -desequilibrios en la señalización debidos a señalizaciones hiperactivadas por algunas oncoproteínas como Myc, daño en el ADN debido a la elevada tasa de proliferación, aumento excesivo del estrés oxidativo...- tanto a lo largo del proceso tumoral como cuando las células son sometidas a algún tratamiento anticancerígeno. Este mecanismo puede ser activado por dos vías: intrínseca y extrínseca:

- *Vía intrínseca:* En respuesta a daños en el ADN, ATM es capaz de activar p53 directa o indirectamente: por un lado, puede fosforilar a CHK2 que, a su vez, fosforila a p53; por otro lado, ATM puede fosforilar directamente a p53. En ambos casos, MDM2 es incapaz de unirse y ubiquitinar a p53 para su degradación. En consecuencia, p53 se acumula en el núcleo y activa la expresión de genes proapoptóticos. Entre otros, se activa la expresión de genes que codifican para las proteínas Bid, NOXA y PUMA que, junto con BIM, se unen a diferentes complejos proteicos que tienen secuestradas proteínas antiapoptóticas como Bcl2 y Bcl-xl. Como resultado, se liberan proteínas pro-apoptóticas, entre ellas Bax y Bak que son capaces de formar dímeros que viajan hasta la mitocondria para crear o activar poros no específicos en ella (poro de transición de permeabilidad mitocondrial -PTPm¹-) a través de los cuales se liberan diferentes factores proapoptóticos al citosol, como el citocromo c.

¹La estructura del PTPm está en debate. Algunos autores proponen que se forma mediante el VDAC de la membrana mitocondrial externa y el transportador del nucleótido de adenosina de la interna y que proteínas como Bak y Bax tendrían un efecto positivo en su apertura. Otros, por el contrario, proponen que son estas mismas proteínas las que forman el PTPm (Bonora and Pinton (2014)).

También, la sobrecarga de Ca^{2+} mitocondrial da lugar a la hinchazón de la mitocondria que, como resultado, conlleva la ruptura de su membrana externa y, en consecuencia, se liberan al citosol factores proapoptóticos mitocondriales (Giorgi et al. (2012)). Así, 7 citocromos C se unen a 7 *factores activadores de proteínas apoptóticas 1 (Apaf1)* y forman un heptámero que recluta a 7 procaspasas-9, se autocataliza para convertirse en caspasa-9 y se forma el apoptosoma. Posteriormente, el apoptosoma activa la procaspasa-3, que da lugar a la caspasa-3, lo que conlleva el inicio de la fase de ejecución de la apoptosis (M. H. Kang and Reynolds (2009), Junttila and Evan (2009), Q. Chen, Kang, and Fu (2018)).

A mayores, también podría considerarse otra segunda vía intrínseca desencadenada por el estrés del RE por la que, a partir de una desregulación de la homeostasis del Ca^{2+} intracelular o la acumulación excesiva de proteínas sin plegar o mal plegadas en el RE, la escasez de nutrientes o condiciones de hipoxia puede inducir la activación de la caspasa-12 que, a su vez, se trasloca desde el RE al citosol para acabar activando la caspasa-9 la cual activa la caspasa-3. En la activación de esta ruta intrínseca alternativa del RE participan complejos proteicos formados por el factor 2 asociado al receptor de TNF (TRAF2) o calpaínas (Q. Chen, Kang, and Fu (2018)).

- *Vía extrínseca:* esta segunda ruta es activada por el sistema inmune y otras señales de muerte celular. Así, el ligando FASL, que se expresa en las células T citotóxicas, se une a su receptor FAS de la superficie de otra célula. En la cara citosólica de FAS se une FADD el cual une dímeros de procaspasa-8 para generar su forma activa, caspasa-8, capaz de generar la forma activa de la procaspasa-3, caspasa-3. No obstante, la caspasa-8 también es capaz de activar la ruta intrínseca mediante el truncamiento de Bid a tBid, que se une a Bax/Bak. Como resultado, se produce una amplificación de la señal de muerte (Q. Chen, Kang, and Fu (2018)).

Independientemente de cual de las dos rutas tenga lugar, el resultado es la activación de la fase de ejecución de la apoptosis. La caspasa-3 activa CAD el cual viaja hasta el núcleo para catalizar la hidrólisis de las uniones entre los nucleótidos internucleosomales, por lo que el material genético se fragmenta. De forma simultánea, la fosfatidilserina, localizada en condiciones normales en la cara citosólica de la MP, se expone en su cara externa lo que promueve la atracción de los macrófagos con el fin de que estos fagociten los cuerpos proapoptóticos, vesículas con doble membrana que empaquetan los componentes celulares, para evitar que se desencadene una respuesta inmune.

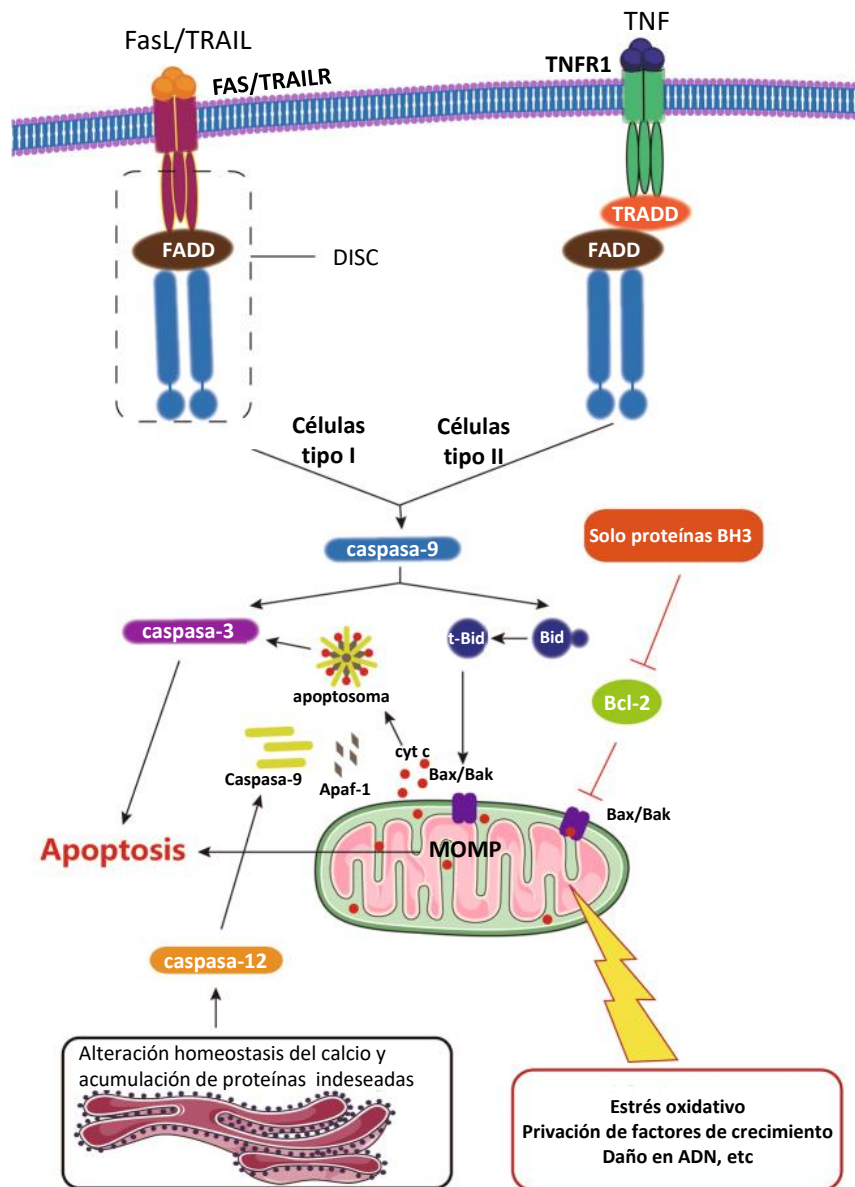


Figura 1.3: Mecanismos de la apoptosis. Existen dos mecanismos por los cuales las células mueren por apoptosis: la vía intrínseca y la extrínseca. La primera puede ser activada por daños en el material genético, falta de factores de crecimiento, estrés oxidativo o diferentes tóxicos pueden dar lugar a la dishomeostasis mitocondrial a través de la expresión de diferentes genes, como Bid, NOXA, PUMA y Bim, que liberan diferentes proteínas secuestradas por proteínas antiapoptóticas como Bcl2 o Bcl-xl, de tal manera que se liberan proteínas pro-apoptóticas como Bax/Bak que forman el PTPm en la mitocondria, incrementa la permeabilidad de la membrana exterior mitocondrial (MOMP) y libera citocromo c al citosol. Este citocromo c puede interactuar con Apaf-1 y la caspasa-9 para acabar activando la caspasa-3 que, finalmente activa la fase de ejecución de la apoptosis. También podría considerarse otra segunda vía intrínseca desencadenada por el estrés del RE por la que, a partir de una desregulación de la homeostasis del Ca^{2+} intracelular o la acumulación excesiva de proteínas en el retículo endoplasmático induce la activación de la caspasa-12 que, a su vez, se trasloca desde el RE al citosol para acabar activando la caspasa-9 la cual activa la caspasa-3. En cuanto a la vía extrínseca, tiene lugar a través de la unión de FASL, TNF- α o TRAIL a su correspondiente receptor que, como resultado, transforma la procaspasa-8 en caspasa-8. En este punto, pueden distinguirse otras dos vías en función del tipo celular. Por un lado, la caspasa-8 puede activar la caspasa-3. Por otro lado, la caspasa-8 puede hidrolizar Bid para dar tBid que, a su vez, interacciona con Bax/Bak y provoca un cambio conformacional que da lugar a su oligomerización y formación del PTPm (modificado de Chen, Kang, and Fu (2018)).

Las células cancerígenas tienen diferentes mecanismos por los que son capaces de evadir las señales apoptóticas. Algunos de los más importantes se resumen a continuación:

- *Mutaciones en las caspasas*: Una de las estrategias que tienen las células tumorales para evadir la apoptosis y promover la supervivencia son las mutaciones de pérdida de función de las caspasas; por ejemplo, mutaciones en la caspasa-9 conlleva la no activación de la caspasa-3 por la vía intrínseca. No obstante, las células del sistema inmune todavía son capaces de activar la vía extrínseca a través de la activación del receptor FAS. Sin embargo, si las mutaciones tienen lugar en la caspasa-8, entonces no tendrá lugar la activación de la apoptosis por la vía extrínseca, por lo que sólo quedará la posibilidad de la vía intrínseca. El problema es mayor cuando es la caspasa-3 la que presenta las mutaciones de pérdida de función, pues CAD no podrá ser liberada de su inhibidor ICAD y, en consecuencia, la fase de ejecución de la apoptosis no tendrá lugar.
- *Bloqueo de la apoptosis en células con daño en el ADN*: algunos tumores sobreexpresan Bcl2 y Bcl-xl, por lo que las proteínas proapoptóticas BAX, BIM, NOXA o PUMA acaban por unirse al exceso de estas en lugar de a las que tienen secuestradas a BAX y BAK, por lo que se evita la activación de la apoptosis por la vía intrínseca. Del mismo modo, las células tumorales también pueden infrarregular estas proteínas proapoptóticas (Adams and Cory (2007), Hanahan and Weinberg (2011)). También se ha descrito que el silenciamiento del gen supresor de tumores TP53 bloquea la apoptosis en presencia de daño en el material genético.
- *Inhibición de la activación de la apoptosis vía PI3K/Akt/PTEN*: esta ruta de señalización se encarga de regular la supervivencia y muerte celular. Puede ser regulada por diferentes señales extracelulares como, por ejemplo, el *factor de crecimiento epidérmico (EGF)* que, tras unirse a su receptor, activa PI3K. Este fosforila el lípido de membrana PIP2 para dar PIP3 al cual se une la oncoproteína Akt para ser activada por PDK1. Posteriormente, Akt fosforilado, a través de la inhibición de las proteínas proapoptóticas BAD y procaspasa-9, inhibe la vía intrínseca de la apoptosis. Además, Akt también activa las rutas de supervivencia y proliferativas a través, por ejemplo, tanto de la inhibición de CHK1 -inhibidor de rutas de supervivencia- y p21 -inhibidor de la proliferación- como de la activación de CDK2 -activador de la proliferación-. Cabe destacar que esta ruta Akt puede ser inhibida por la acción de PTEN, pues defosforila PIP3 para dar PIP2 y, en consecuencia, se previene el reclutamiento de Akt. En muchos tumores, PTEN está ausente, mutado o inactivado, por lo que se evade la inhibición de la ruta de señalización Akt y, por ende, se promueve la supervivencia celular. Así, la ruta Akt puede activarse en las células tumorales debido a un aumento en la expresión de los receptores de crecimiento o de mutaciones de ganancia de función de estos, por un aumento en la expresión de PI3K y Akt o por mutaciones de pérdida de función de PTEN.

En conclusión, alteraciones en los mecanismos que promueven la entrada en la apoptosis conllevan la acumulación de errores a pesar de los cuales la célula tumoral continua proliferando.

No obstante, la apoptosis y alteraciones en su maquinaria no son la única manera por la cual las células tumorales se vuelven resistentes a la muerte. También hay otras como la autofagia, o la necrosis.

- *Autofagia*: es un proceso auto catabólico altamente regulado cuyo fin es reciclar moléculas y orgánulos dañados. Este se inicia en respuesta a diferentes factores de estrés, como la deficiencia de nutrientes, presencia de orgánulos dañados, hipoxia, aumento de ROS, estrés del RE o tratamientos con diferentes drogas. Tiene lugar en unas vesículas con doble membrana conocidos como autofagosomas los cuales envuelven a las macromoléculas u orgánulos. Posteriormente, estos autofagosomas se fusionan con los lisosomas y dan lugar a los autofagolisosomas, en los cuales tiene lugar la hidrólisis de su contenido mediante las enzimas lisosomales. Finalmente, el producto de degradación generado en los autofagolisosomas puede ser usado nuevamente por la célula como nutrientes en su metabolismo. Este mecanismo puede ser dividido en cuatro pasos (figura 1.4): iniciación, nucleación, expansión y cierre, y degradación (Q. Chen, Kang, and Fu (2018)).

La fase de iniciación, que se corresponde con los estados iniciales de la formación de los autofagosomas, está regulada, en mamíferos, por los complejos ULK y VPS34, localizados en los fagóforos, estructuras que derivan de la membrana del RE. Por ello, uno de los principales reguladores de la autofagia es mTORC1, pues puede inhibirla a través de la inactivación de ULK y VPS34 (Sudarsanam and Johnson (2010), Levine and Kroemer (2008)). Por un lado, mTORC1 puede fosforilar la quinasa ULK1, que forma parte del complejo ULK, de tal manera que previene la activación del este complejo y, por ende, AMPK no puede unirse a ULK1. mTORC1 también puede fosforilar a ATG13, otra proteína del complejo ULK, lo que favorece la separación del complejo. Por otro lado, mTORC1 puede fosforilar a las proteínas Beciclina-1 y VPS34, lo que conlleva la inactivación del complejo VPS34. No obstante, mTORC1 también puede inhibir la autofagia a través de la inhibición de la transcripción de TFEB, el cual activa la expresión de genes relacionados con la autofagia (Q. Chen, Kang, and Fu (2018)).

Tras la activación de los complejos, ULK y VPS34 (iniciación), se inicia la fase de nucleación en la que participa el complejo Beciclina-1, regulado negativamente por Bcl-2, de tal manera que su activación participa en el aislamiento de la membrana para su nucleación. Además, la proteína transmembrana Atg9 y la proteína VMP1 de la membrana de la vesícula participan en el transporte de lípidos para conseguir el aislamiento de la membrana. A continuación, en la fase de expansión y cierre intervienen dos sistemas conjugados con proteínas similares a la ubiquitina, Atg12 y LC3, que tras una serie de interacciones dan lugar, en mamíferos, al LC3 II el cual desempeña un papel clave en la formación de los autofagosomas que se asocia tanto a las membranas internas como externas de estos y, además, es un marcador de autofagia. Finalmente, la fusión entre el autofagosoma y el lisosoma está mediada por la proteína similar a SNARE, lo que da lugar a la formación del autofagolisosoma en cuyo interior las enzimas lisosomales hidrolizan su contenido (Q. Chen, Kang, and Fu (2018)).

En situaciones de estrés, tales como hipoxia, daño celular o deficiencia de nutrientes, los niveles de señalización de autofagia se ven incrementados considerablemente, por lo que esta pasa a ser un mecanismo de supervivencia (E. White and DiPaola (2009)). Es más, cuando el estrés es de origen metabólico, las células pueden obtener energía de sí mismas a través de la activación de AMPK, que es otro regulador clave de la autofagia pero en “sentido” contrario a mTORC1, pues impide que mTORC1 inactive los complejos ULK y VPS34. En particular, AMPK puede fosforilar a ULK1 e impedir que mTORC1 inactive al complejo ULK, y también a Beiclina-1 y VPS34 para activar el complejo VPS34 (Q. Chen, Kang, and Fu (2018)). No obstante, si el estrés de origen metabólico persiste, este mecanismo de autofagia puede no ser suficiente y la célula acaba por morir. En esta situación, Bcl-2/Bcl-xl secuestra al complejo Beiclina-1 y se inhibe así la autofagia y entra en juego la apoptosis ya que se libera Bak/Bax de Bcl-2. Por lo tanto, existe una regulación cruzada entre la autofagia y la apoptosis.

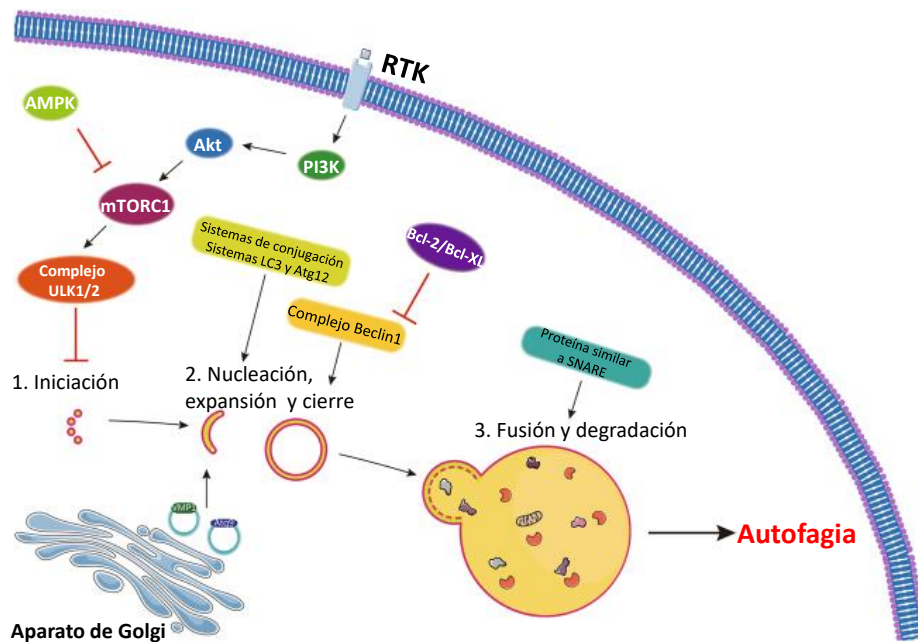


Figura 1.4: Mecanismos de la autofagia. Este proceso auto catabólico por el cual se reciclan moléculas y orgánulos celulares puede ser dividido en cuatro pasos: iniciación; nucleación; expansión y cierre, y fusión y degradación. En los mamíferos, la fase de iniciación está mediada por los complejos ULK y VPS34, los cuales están regulados por mTORC1 que, a su vez, está regulado positivamente por la vía PI3K/Akt y de forma negativa por AMPK. En la fase de nucleación participa el complejo Beiclina-1, regulado negativamente por Bcl-2, de tal manera que su activación participa en el aislamiento de la membrana para su nucleación. Además, la proteína transmembrana Atg9 y la proteína VMP1 de la membrana de la vesícula participan en el transporte de lípidos para conseguir el aislamiento de la membrana. En la fase de expansión y cierre intervienen dos sistemas conjugados con proteínas similares a la ubiquitina, Atg12 (formado por las proteínas Atg5, Atg7, Atg12, Atg12 y Atg16) y LC3. Finalmente, la fusión entre el autofagosoma y el lisosoma está mediada por la proteína similar a SNARE, lo que da lugar a la formación del autofagosoma en cuyo interior las enzimas lisosomales hidrolizan su contenido (modificado de Chen, Kang, and Fu (2018)).

Las células tumorales presentan una elevada demanda energética debido a un metabolismo ineficiente de la glucosa de las células tumorales (*efecto Warburg*, Otto (1956), Woolf, Syed, and Scheck (2016), Poff et al. (2019)), a la elevada tasa proliferativa o al microambiente tumoral en el que hay escasez de nutrientes y condiciones de hipoxia. Así, para hacer frente a su elevada demanda energética, estas hacen uso de la autofagia, lo que les permite cumplir con sus requerimientos energéticos y, en consecuencia, evitar la necrosis y volverse tolerantes a situaciones de estrés (E. White and DiPaola (2009)).

- *Necrosis*: en un principio, se pensó que la necrosis era un proceso no regulado, a diferencia de la apoptosis. Sin embargo, se ha podido comprobar que este proceso también está regulado. Es más, se han descrito múltiples mecanismos regulados no apoptóticos que dan lugar a la muerte celular: necroptosis, partanatos, ferroptosis, piroptosis o NETosis. De entre todas ellas, la necroptosis comparte diferentes componentes de vías de señalización con la apoptosis. Este mecanismo puede ser inducida por daño mitocondrial, por el aumento de ROS, por la isquemia e incluso por factores externos como $TNF\alpha$. Pueden distinguirse 4 vías de señalización clave en la activación de este mecanismo: señalización por TNFR, por Fas/TRAILR, por TLR3/4 o por DAI. De ellos, la activación de la necroptosis vía $TNF\alpha$ es la más estudiada. Por lo tanto, $TNF\alpha$, aparte de jugar un papel clave en el proceso inflamatorio, también es capaz de desencadenar la muerte celular a través de la activación de su receptor TNFR, a través de mecanismos no apoptóticos o apoptóticos en función de si está o no bloqueada la apoptosis.

Así, la activación de TNFR1 por la unión de $TNF\alpha$ induce la formación del complejo I dependiente de RIP1 en la cara citoplasmática de este receptor a partir del reclutamiento de una serie de proteínas: proteína asociada al dominio de muerte de TNFR1 (TRADD), RIP1, TRAF2, E3 ubiquitina ligasa, cIAP1/2 y LUBAC. En condiciones normales, este macrocomplejo recluta diferentes complejos con el fin de activar señales proinflamatorias o de supervivencia, como por ejemplo el complejo TAB (proteína de unión a la quinasa 1 activada por TGF) o el complejo quinasa IKK. No obstante, CYLD desubiquitina a RIP1 este complejo I se desestabiliza y RIP1 se disocia de este. En consecuencia, se forma otro complejo conocido como IIa constituido por TRADD, FADD, procaspasa-8 y FLIP, capaz de prevenir la necroptosis, de tal manera que se activa la caspasa-8 que, a su vez, escinde RIP1 y activar la apoptosis (Pan et al. (2019)). Además, cuando la función de inhibidores de la apoptosis, TAB e IKK está bloqueada, la señalización $TNF\alpha$ -TNFR puede inducir la formación de otro complejo, IIb, formado por RIP1, RIP3, FADD, procaspasa-8 y FLIP_L, en el cual la caspasa-8 activada induce la apoptosis dependiente de RIP1 al escindir RIP1 y disociarlo del complejo IIb, a la par de también escindir e inactivar a RIP3 y CYLD.

Sin embargo, cuando la caspasa-8 está inhibida, como puede suceder ante situaciones como la isquemia, tiene lugar un aumento de las proteínas RIP3 y MLKL; así, el complejo IIb puede transformarse en el necrosoma (Pan et al. (2019)). En este sentido, las proteínas RIP1, RIP3 y CYLD ya no son escindidas por la caspasa-8 y permanecen activas, por lo que CYLD se puede unir al RIP1 asociado a TNFR y activarlo, de tal manera que RIP1 se disocia del complejo y se forma el complejo IIb que, bajo estas condiciones, es el nucleosoma. Así, RIP3 se fosforila

en el nucleosoma que, a su vez, recluta a MLKL y lo fosforila, de tal manera que varios MLKL fosforilados son capaces de formar un poro en la MP y, en consecuencia, aumenta la permeabilidad de la célula, se incrementa la presión osmótica intracelular y aumenta ROS, lo que causa daño mitocondrial. Como resultado, la célula revienta y vierte todo su contenido al medio extracelular. El MLKL fosforilado también puede interactuar con PGAM5 de la membrana mitocondrial, activar el factor de fisión mitocondrial Drp1 e inducir la necrosis a través de la fragmentación mitocondrial (Q. Chen, Kang, and Fu (2018)).

En cuanto a la señalización por Fas/TRAILR, en condiciones normales FasL o TRAIL se unen a su receptor y se induce la formación de DISC, activa la caspasa-8 y se inicia la fase de ejecución de la apoptosis. Sin embargo, en ausencia de cIAP o cuando la caspasa-8 está inhibida, RIP1 se trasloca a la membrana y se forma el complejo IIB y se induce la necroptosis. La señalización por TLR3/4, activada por el lipopolisacárido o por ácido ácido poli-inosínico/policitidílico, respectivamente, induce la necroptosis mediada por TRIF el cual puede interactuar con RIP1 y RIP3. Por ello, en presencia del inhibidor de la apoptosis zVAD-fmk la activación de estos receptores induce la necroptosis mediada por TRIF. Por último, la señalización de DAI en respuesta a ADN bicatenario viral media la necrosis dependiente de RIP3 (Q. Chen, Kang, and Fu (2018)).

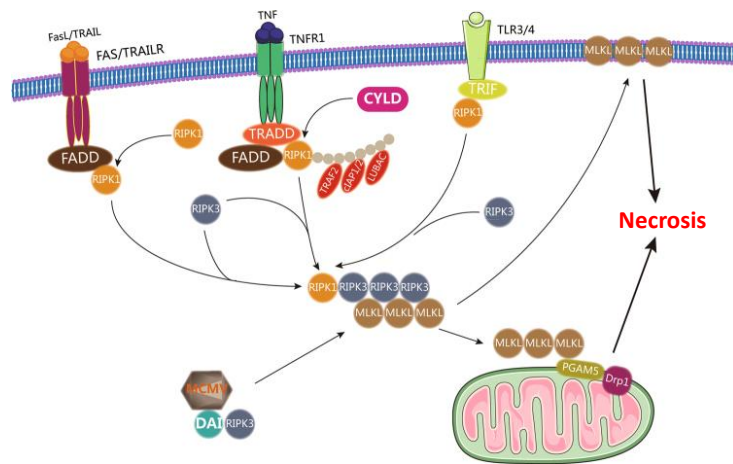


Figura 1.5: Mecanismos de la necrosis. La necrosis puede ser activada por diferentes mecanismos: señalización por TNFR, Fas/TRAILR, TLR3/4 o DAI. La activación de TNFR conlleva la formación del complejo I (TRADD, RIP1, TRAF2, E3 ubiquitina ligasa, cIAP1/2 y LUBAC). Este recluta y activa a la procaspasa-8. En situaciones como la isquemia este complejo se desestabiliza por la acción de CYLD sobre RIP1, que lo desubiquitina. Esto conlleva la formación del necrosoma, aumento de RIP3 y MLKL, y la inhibición de la caspasa-8. Posteriormente, el RIP3 del necrosoma oligomeriza con RIP1 y es fosforilado el cual, a su vez, recluta a MLKL y lo fosforila. Este se transloca a la MP, forma un poro y aumenta su permeabilidad, la presión osmótica, ROS, el daño mitocondrial y, finalmente, la célula revienta vertiendo todo su contenido al medio extracelular. El MLKL fosforilado también interactúa con el PGAM5 de la membrana mitocondrial para activar al factor de fisión mitocondrial Drp1 el cual induce necrosis a través de la fragmentación mitocondrial. En cuanto la señalización por Fas/TRAILR, en ausencia de cIAP e inhibición de caspasa-8 la activación de esta ruta induce necrosis. La señalización por TLR3/4, activada por el lipopolisacárido o por ácido ácido poli-inosínico/policitidílico, induce la necroptosis mediada por TRIF en presencia del inhibidor de la apoptosis zVAD-fmk. Por último, la señalización de DAI en respuesta a ADN bicatenario viral, media la necrosis dependiente de RIP3 (modificado de Chen, Kang, and Fu (2018)).

A pesar de que en los estadios iniciales del desarrollo tumoral hay una pequeña actividad angiogénica, en estadios más avanzados, debido a la baja eficiencia de los nuevos vasos formados respecto de los de los tejidos sanos, puede dar lugar a una situación de isquemia en el tejido tumoral y, en consecuencia, tiene lugar la *necrosis*. En tal situación, hay una alteración general de los orgánulos lo que conlleva que la célula se hinche hasta que revienta y libera su contenido al medio extracelular, lo que promueve un estado inflamatorio. En consecuencia, se reclutan células inflamatorias del sistema inmune cuya función será promover la supervivencia del tejido dañado y eliminar los residuos necróticos, No obstante, estas células inmunes proinflamatorias reclutadas pueden promover la tumorigénesis fomentando la angiogénesis, la proliferación de las células cancerígenas y su capacidad invasiva (Q. Chen, Kang, and Fu (2018)).

Los procesos anteriores, aunque son mecanismos independientes, están interconectados entre sí, tal y como se puede observar en la figura 1.6.

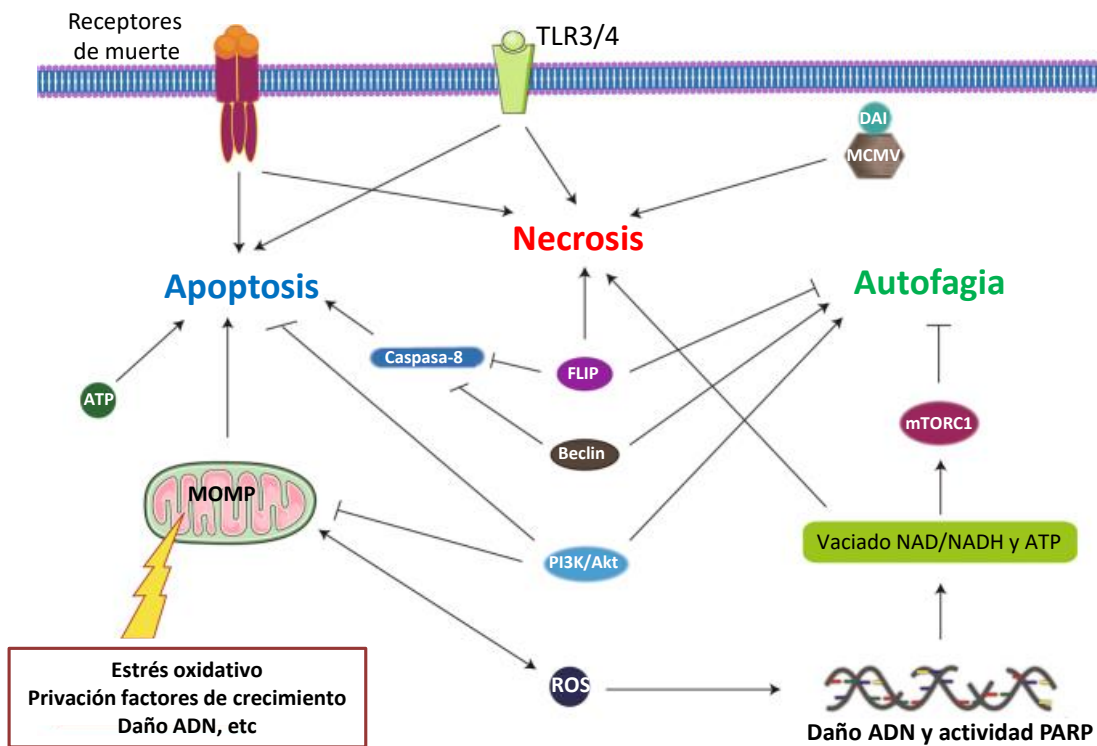


Figura 1.6: Relación entre la apoptosis, autofagia y necrosis. Altos niveles de ATP promueven la apoptosis y bajos la necrosis. mTORC1 detecta los bajos niveles de ATP y relaja la inhibición de la autofagia para activar la necrosis. Akt puede activar mTORC1 e inhibir la autofagia, pero también puede inhibir la apoptosis a fosforilando factores apoptóticos. FLIP puede interaccionar con la caspasa-8, Atg3 y LC3 puede regular el camino de muerte celular que tomará la célula. La caspasa-8 puede inhibir la necroptosis hidrolizando RIP1, RIP3 o CYLD. En cuanto a la Beciclina-1, puede intervenir en la decisión entre la apoptosis o la autofagia, por ejemplo a través de su combinación con las proteínas antiapoptóticas Bcl-2 o Bcl-xl. (modificado de Chen, Kang, and Fu (2018)).

A grandes rasgos, los niveles de ATP son un factor importante para la célula a la hora de escoger entre la apoptosis (altos niveles) o la necrosis (bajos). De hecho, mTORC1 puede detectar estos bajos niveles de ATP, relajar la inhibición de la autofagia y activar la necrosis. Además, Akt, aparte de activar mTORC1 e inhibir la autofagia, también puede inhibir la apoptosis a través de la fosforilación de factores apoptóticos como Bad o la caspasa-9. FLIP, a través de la interacción sus diferentes isoformas con la caspasa-8, Atg3 y LC3 puede regular el camino de muerte celular que tomará la célula. Por ejemplo, cFLIP se une a Atg3 e inhibe la formación del autofagosoma al impedir la unión de Atg3 con LC3. Además, el heterodímero formado entre la procaspasa-8 y cFLIP_L inhibe la necroptosis al escindir RIP1 y es capaz de activar la apoptosis, pero a elevadas concentraciones de cFLIP_L la apoptosis también se inhibe (D. W. Chang et al. (2002)). Sin embargo, el heterodímero de la procaspasa-8 con las isoformas cFLIP_S o cFLIP_R puede desencadenar la necroptosis. En cuanto a la Beciclina-1, puede intervenir en la decisión entre la apoptosis o la autofagia, por ejemplo a través de su combinación con las proteínas antiapoptóticas Bcl-2 o Bcl-xl (Q. Chen, Kang, and Fu (2018)).

1.2.4. Potencial ilimitado de replicación

En las células normales, los cromosomas tienen una zona terminal denominada telómeros. Estos son secuencias repetidas de ADN que protegen los extremos de los cromosomas los cuales se van acortando con cada división y, en consecuencia, la longitud del telómero se acorta. Por lo tanto, el número de replications que puede tener una célula viene determinado por dicha longitud. Finalmente, cuando la célula alcanza un mínimo de longitud del telómero -alcanza el límite *Hayflick*²- esta entra en senescencia, situación en la que la célula es metabólicamente activa, pero ni se divide ni prolifera (Hanahan and Weinberg (2011), Shay and Wright (2011)).

Esta situación de senescencia se mantiene gracias a la activación de genes supresores de tumores tales como *TP53* y *RBI* (gen del Rb). Por lo tanto, alteraciones en estos genes resultan en la proliferación de estas células que han alcanzado su límite *Hayflick*. Por norma general, cuando se da esta situación, tras varias rondas de replicación la célula entrará en crisis y morirá por apoptosis. Sin embargo, puede suceder que estas células adquieran la capacidad de expresar la telomerasa, enzima que sintetiza ADN telomérico durante la replicación, y, en consecuencia, se vuelvan inmortales (Shay and Wright (2011)). No obstante, esta enzima está inactivada en la mayoría de las células humanas somáticas. Esta enzima tiene una subunidad catalítica, TERT, que es la encargada de sintetizar el ADN telomérico. La transcripción de esta subunidad está regulada por varios factores como: Myc y MAX, que activan su transcripción, y MAD1, que la reprime (D. Xu et al. (2001), Toaldo et al. (2010)).

Es más, en el 90 % de las células immortalizadas, incluidos las procedentes de tumores, se ha observado una activación de esta enzima (Hanahan and Weinberg (2011)). En muchos de ellos se ha observado el aumento de expresión de genes *MYC* y la disminución de MAD1, por lo que MAD1 se

²El límite de *Hayflick* es el número de replications que puede llevar a cabo una célula antes de entrar en senescencia.

desplaza del complejo inhibitorio de la transcripción de TERC y se une el dímero Myc-MAX. Este dímero se une a coactivadores y desencadena la transcripción de TERT. Cabe destacar que muchas de las células cancerígenas son capaces de expresar telomerasa debido a la inhibición de genes supresores de tumores o a la activación de oncogenes (D. Xu et al. (2001), Toaldo et al. (2010)).

Curiosamente, en los últimos años se han ido descubriendo nuevas funciones de esta enzima. Por ejemplo, se ha observado que la subunidad TERT es capaz de amplificar la señalización por la vía Wnt, importante en muchos tipos de cáncer, como el de colon, y en el mantenimiento de la señal proliferativa a través de la acumulación de β -catenina (Park et al. (2009)).

1.2.5. Inducción de la angiogénesis

De forma general, para que una célula sea capaz de sobrevivir ha de poder disponer de O_2 y nutrientes. Por ello, es necesario que en sus cercanías haya vasos sanguíneos. No obstante, de forma normal no hay un vaso sanguíneo alrededor de cada célula de un tejido, puesto que es posible el intercambio de nutrientes y catabolitos a través de un mecanismo de difusión simple a distancias de máximo 1 mm respecto del vaso sanguíneo más cercano. No obstante, el crecimiento del tejido conlleva que las células se alejen cada vez más de los vasos de tal manera que a distancias superiores a 50 μm estas no reciban suficiente O_2 y, por ende, entren en un ambiente hipóxico. Es más, a partir de 70 μm comienza la zona de anoxia (Carmeliet and Jain (2011b)).

A medida que disminuye la concentración de O_2 , también lo hace el pH, el ATP y la glucosa, y aumenta el lactato debido al incremento de la glucólisis anaeróbica. Ante esta situación, las células hipóxicas del tejido y otras células del microambiente como los macrófagos y los fibroblastos liberan ciertos factores de la familia VEGF los cuales inducen la formación de nuevos vasos sanguíneos o neovascularización. Además, los fibroblastos, células endoteliales, epiteliales e inflamatorias también liberan metaloproteasas (*MMP*, del inglés *matrix metalloprotease*) para degradar la matriz extracelular, la cual juega un papel fundamental en la regulación de la angiogénesis (Carmeliet and Jain (2011a), J. Liu et al. (2012)).

En condiciones normales, este proceso da lugar a nuevos vasos que adquieren una morfología similar a la de los vasos preexistentes y, además, finaliza cuando las demandas de O_2 cesan. Existen tres mecanismos de neovascularización fisiológica, los cuales pueden darse de forma simultánea (J. Liu et al. (2012), Stacker et al. (2014)):

- *Angiogénesis y linfoangiogénesis*: procesos de germinación y ramificación a partir de vasos sanguíneos y linfáticos ya existentes, respectivamente.
- *Vasculogénesis*: es un proceso por el cual se desarrollan nuevos vasos sanguíneos a partir de células endoteliales progenitoras, que son atraídas hacia la zona hipóxica desde de la médula ósea por los factores angiogénicos liberados.

Obviamente, debido a su crecimiento desmesurado, en los tumores existen condiciones de hipoxia, por lo que son necesarios los mecanismos de neovascularización ya que, de no existir, ningún tumor sería superior a 1 o 2 cm. Por ello, tanto las células tumorales como las *macrófagos asociados al tumor* (TAM, del inglés *tumor-associated macrophage*) y los *fibroblastos asociados al cáncer* (CAF, del inglés *cancer-associated fibroblast*) liberan factores angiogénicos como VEGF, RGF β , FGF o angiopoyetinas (Carmeliet and Jain (2011a), J. Liu et al. (2012)). Además, las células tumorales y estromales también sintetizan otras moléculas capaces de regular la angiogénesis como, por ejemplo, los factores de crecimiento (GCSF, HGH, GFG...) que se unen a TKR de las células endoteliales. Sin embargo, la neovascularización tumoral, a diferencia de la fisiológica, presenta una morfología caótica que da lugar a una circulación ineficiente. En particular, estos vasos son dilatados y de calibre irregular con múltiples ramificaciones e interconexiones y, además, tienen una menor cantidad de pericitos, lo que hace que sean menos estables, más propensos a fugas y más permisivos a la extravasación (Carmeliet and Jain (2011b)).

Algunas de las consecuencias de esta circulación caótica e ineficiente es el aumento del flujo de líquido hacia el espacio intersticial, el colapso de los vasos linfáticos, la extravasación de eritrocitos, o el aumento en la expresión de factores angiogénicos y de MMP que inducen la degradación de la matriz extracelular y la transformación de las células tumorales en células similares a las mesenquimales que adquieren capacidad invasiva. Esta situación junto con la discontinuidad que hay presente en la capa endotelial de los nuevos vasos tumorales favorece considerablemente la migración y diseminación de las células tumorales a otros tejidos a través del torrente sanguíneo y, por ende, aumenta la probabilidad de metastatizar. En consecuencia, la neovascularización tumoral, aparte de favorecer la supervivencia y proliferación tumoral, da lugar a un fenotipo más agresivo (Zijl, Krupitza, and Mikulits (2011), Carmeliet and Jain (2011a)).

Al igual que existen factores proangiogénicos, también hay factores antiangiogénicos, tales como las trombospondinas o el factor derivado del pigmento epitelial -disminuye con el envejecimiento y en los tumores-. Es más, también existe ARN no codificante capaz de regular la angiogénesis, tanto miARN, que se unen al ARNm para degradarlo o bloquear la traducción, como lncARN, que están involucrados en la expresión de genes, la estabilidad del ADN y la traducción del ARNm. En ambos casos, en función del ARNm o gen (lncARN) sobre el que actúen, serán pro o antiangiogénicos. Es evidente que la angiogénesis está sumamente regulada y, además, que el destino de esta viene determinado por el balance entre los factores pro y antiangiogénicos (Carmeliet and Jain (2011a), Khorshidi, Dhaliwal, and Yang (2016)).

1.2.6. Activación de la invasión y la metástasis

El proceso de invasión y metástasis comienza con la transformación de algunas de las células tumorales las cuales pasan de un fenotipo epitelial a uno mesenquimal, lo que se conoce como EMT, y, en consecuencia, se vuelven más invasivas. Hay dos tipos de señalización que favorecen esta transición: la señalización autocrina de las células tumorales y la señalización paracrina por parte de las células estromales (CSF-1, EGF, IGF2). Los cambios que tienen lugar están regulados por diferentes rutas de señalización como, por ejemplo, aquellas en las que participan los miembros de la familia SMAD o la ruta de las MAPK. Por un lado, TGF β , entre otros factores, activa las proteínas SMAD las cuales incrementan los niveles de los factores de transcripción Zeb2 y Snail en el núcleo. Por otro lado, la activación de ERK a través de la ruta de las MAPK induce el aumento en la expresión de Snail (Hanahan and Weinberg (2011), Zijl, Krupitza, and Mikulits (2011), Yifan Wang et al. (2013), Stuelten, Parent, and Montell (2018)).

Tanto ZEB2 como SNAIL modifican los patrones de expresión de diferentes proteínas relacionadas con la adhesión. Por ejemplo, la acción de estos disminuye la expresión de marcadores epiteliales como la E-cadherina, cuya pérdida causa el debilitamiento de la adhesión célula-célula entre las células tumorales y las de su entorno. Esto conlleva el aumento y acumulación de β -catenina libre en el citosol y, en consecuencia, el aumento en la expresión de genes que favorecen la activación de la transición epitelio mesenquimal. Entre otros cambios, estas células aumentan la expresión de vimentina (Richardson et al. (2018)), modifican la organización del citoesqueleto y pierden su polarización. Además, la acción de las MMP y otras proteasas provocan la degradación tanto de la matriz extracelular como de ciertas proteínas lo que causa la liberación de factores de crecimiento. Todos estos cambios favorecen su motilidad y su capacidad migratoria (Yifan Wang et al. (2013), Stuelten, Parent, and Montell (2018)).

A continuación, se explicarán brevemente algunos de los procesos clave de la metástasis:

- *Motilidad*: la motilidad está regulada por un complejo sistema de proteínas y cascadas de señalización que, a su vez, lo está por la activación local de *GTPasas pequeñas* de la familia *Rho*: Rho, Rac y Cdc42. Las células tumorales pueden presentar dos tipos de movimientos: *mesenquimal* y *ameboide*. Por un lado, la ruta de señalización Ras/Tiam1/Rac está involucrada en el movimiento mesenquimal, mientras que la ruta Rho/ROCK lo está en el ameboide. Por lo tanto, en función del balance entre ambas rutas las células tumorales pueden alternar entre ambos tipos de movimiento (B. C. Jones et al. (2017), Stuelten, Parent, and Montell (2018)).
- *Invasión*: en cuanto a los tipos de invasión, se distinguen tres tipos: *mesenquimal*, en el que sólo unas pocas células que se encuentran en el foco de la invasión inicial se infiltran en el tejido de una en una; *colectiva* en el que las células tumorales no han perdido del todo la capacidad de adhesión y presentan un fenotipo híbrido endotelial-mesenquimal, lo que les permite extenderse y avanzar como un grupo para invadir de forma colectiva -a nivel clínico se correlaciona con un mal pronóstico-, y *ameboide*: en el que los lamelopodios que presentan

estas células tumorales les facilita su paso a través de los elementos de la matriz extracelular sin necesidad de ser destruida por metaloproteasas (Zijl, Krupitza, and Mikulits (2011), B. C. Jones et al. (2017), Stuelten, Parent, and Montell (2018)). Independientemente del mecanismo invasivo, sólo unas pocas células invasivas lograrán pasar a través de las paredes de los vasos sanguíneos y entrar en el torrente sanguíneo. Una vez en el torrente sanguíneo, las células invasivas se verán sometidas a diferentes factores que desafiarán sus oportunidades de supervivencia, tales como la acción del sistema inmune (pueden ser atacados por las células *natural killer (NK)*) o de diferentes factores físicos y químicos a los que son sometidos en el interior de los vasos (Zijl, Krupitza, and Mikulits (2011), Labelle and Hynes (2012)):

- *Factores físicos y químicos*: cabe destacar que estas células pueden viajar por el torrente sanguíneo tanto de forma individual -*células tumorales circulantes (CTC)*- como formando *agregados celulares*. Entre los diferentes factores físicos y químicos destacan: el *esfuerzo de cizalladura*, diferencia de velocidad entre las CTC que viajan en el centro del vaso -mayor velocidad- respecto de las que lo hacen más cercanas a las paredes de los vasos; *colisión*, las CTC colisionan con las células sanguíneas y las células endoteliales, lo que compromete su probabilidad de supervivencia y su capacidad para adherirse a las paredes de los vasos; *estrés oxidativo*, algunas células tumorales son capaces de protegerse frente al estrés oxidativo generado por ROS a través de la activación de la ruta del ácido fólico, que estimula la producción de NADPH, una molécula que controla el efecto dañino de los ROS.

-*Plaquetas*: cabe destacar que todos estos factores físicos y químicos afectan en menor medida a los agregados de células tumorales circulantes. En este sentido, las plaquetas tienen un papel clave en la supervivencia de las células metastásicas ya que participan en la formación de dichos agregados a través de la acción de diferentes proteínas: *subunidad α de la proteína Gq*, *protrombina* y *fibrinogeno*. También surgen los *coágulos tumorales* los cuales también reclutan otros tipos de células que protegen a las células tumorales de los linfocitos NK tales como células del sistema inmune que, aparte de incrementar la resistencia de estos agregados frente al sistema inmune (escape inmune), también activan las rutas de supervivencia de las células tumorales.

- *Células inmunes*: En particular, algunas de estas células del sistema inmune son: *Breg*, que participan en la protección del coágulo tumoral induciendo la diferenciación de las Thelper a Treg, *Treg* inactiva las *células T citotóxicas (CTL)*, del inglés, *cytotoxic T cell*) expresando IL10 y TGF β , suprime la acción citotóxica de las NK con TGF β , activa la apoptosis de los NK con la proteína de unión β galactosidasa (BGBP, del inglés *beta galactosidase binding protein*) e incrementa la supervivencia de las CTC a través de la proteína de membrana RANKL unida a su receptor RANK de las células tumorales, y *macrófagos asociados al coágulo* que incrementan el reclutamiento de Treg e inhiben la acción anti-tumoral de las CTL a través de la interacción PD1 de la membrana de las CTL con el PDL1 de la suya.

En cuanto a las células tumorales de los agregados, son capaces de activar las células endoteliales de los vasos sanguíneos para expresar moléculas de adhesión tales como VCAM1 y VAP1. En consecuencia, estas células endoteliales son capaces de reclutar los monocitos circulantes que se diferencian a TAM. Estos TAM tiene una gran relevancia en el proceso tumorigénico, pues aumentan la supervivencia de las células tumorales mediante la producción de citoquinas y también a través de la interacción con las proteínas de membrana señalizadoras de rutas de supervivencia.

- *Nicho metastásico*: antes de colonizar otros tejidos y metastatizar, las células tumorales secretan y liberan al torrente sanguíneo unas vesículas de unos 30-100 nm de diámetro denominadas exosomas, en cuyo interior hay ARN (ARNm, ARNi...) y proteínas citosólicas de la propia célula tumoral. Estos juegan un papel importante en la metástasis, pues sus proteínas de membrana, sobretodo las integrinas, determinan el tejido de destino. Así, los exosomas pueden viajar a órganos distantes y preparar los nichos premetastáticos para la llegada de las células metastásicas. Sus efectos en el tejido diana dependen de las características del tejido primario del cual fueron originados y del órgano donde se han establecido estos nichos. Así, estas vesículas causan cambios en los tejidos diana que, a la larga, acaban por favorecer la supervivencia y la proliferación de las células tumorales (W. Tian, Liu, and Li (2019)).

Las células tumorales del foco primario también liberan diferentes factores, como VEGFA, $TNF\alpha$ y $TGF\beta$, que alcanzan los nichos premetastáticos, en los cuales inducen la expresión de moléculas quimiotácticas que atraen células mieloides a este nicho y, además, también secretan otras moléculas que aumentan la adhesión de estas. De igual modo, los exosomas también tienen un efecto similar. Como resultado, estas células mieloides acabarán por ser uno de los principales componentes de estos nichos. Cabe destacar que estas células establecen un entorno en el cual se inhibe el sistema inmune (NK, células T...) y promueve la metástasis. Además, secretan MMP, como MMP9, que facilitarán la posterior invasión de las células tumorales (W. Tian, Liu, and Li (2019)).

No obstante, también otras células del sistema inmune juegan un papel importante en la adecuación de los nichos metastásicos. Por ejemplo, los *Thelper* participan activamente en la metástasis tumoral en diferentes órganos. Por un lado, en los tejidos óseos sanos estas secretan el ligando RANKL el cual mejora la diferenciación de los osteoclastos, células responsables de la degradación de la matriz ósea, cuya proliferación promueve la metástasis ósea en células que fueron originadas en tumores primarios de mama. Por otro lado, en la metástasis de pulmón originada a partir de tumores de mama primarios, la llegada de moléculas originadas por el tumor primario inducen la expresión de la quimioquina CCL22 por parte de las células no mieloides del estroma pulmonar, molécula responsable del reclutamiento de *Treg* en el nicho (Kitamura, Qian, and Pollard (2015), Jin et al. (2018)). Estos linfocitos disminuyen la actividad de las células inmunes y crean un nicho seguro en el cual las células tumorales están protegidas frente al ataque del sistema inmune (Nakamura et al. (2006)).

- *Destino de la metástasis*: algunos tipos de cáncer muestran preferencia por ciertos órganos para metastatizar. Para explicar este hecho, se han propuesto diferentes teorías (Stuelten, Parent, and Montell (2018)):
 - *Hipótesis Soil and Seed*: sugiere que los tejidos tienen diferentes características que les hacen más o menos favorables a su colonización, a ser un entorno amigable para las células tumorales extravasadas.
 - *Hipótesis quimiotáctica*: postula que ciertas señales químicas producidas por diferentes órganos, tales como quimioquinas, atraen y facilitan el reclutamiento de las CTC. Así, sustancias como las citoquinas, producidas por ciertos tejidos, promueven la metástasis en órganos particulares.
 - *Hipótesis sobre la anatomía del sistema circulatorio*: esta hipótesis se basa en que para su extravasación y el destino, las CTC necesitan pegarse a las células sanguíneas de forma estable. Las diferencias en el diámetro de los vasos conlleva que las CTC se adhieran a los primeros capilares con los que se topan, por lo que la frecuencia de metástasis en aquellos órganos irrigados por una gran densidad de capilares, tales como el pulmón o el hígado, es mayor. No obstante, también existen otras diferencias estructurales entre los vasos de los diferentes órganos que facilitan o limitan la metástasis. Por un lado, en el hígado y la médula ósea la probabilidad de metástasis es superior debido a la discontinuidad de sus vasos y a una membrana basal incompleta. Por el contrario, los vasos pulmonares son continuos y poseen una membrana basal completa -capa endotelial- entre las células capilares y las células alveolares, por lo que son más resistentes a la extravasación. Por otro lado, los vasos sanguíneos del cerebro son continuos y están rodeados por astrocitos y periocitos, por lo que constituyen una barrera todavía más difícil de ser sobrepasada por las CTC.
 - *Hipótesis sobre la genética de las células tumorales*: esta afirma que la especificidad de la metástasis está determinada por la genética de las células tumorales y no por factores relacionados con los tejidos diana.
 - *Hipótesis sobre los exosomas secretados por las células tumorales*: propone que los tumores primarios también sientan las bases para la metástasis en los órganos diana. Ellos secretan exosomas que contienen proteínas que reclutan células que formarán los nichos premetastáticos que, a su vez, atraerán CTC (W. Tian, Liu, and Li (2019)).

A pesar de que el hígado es el destino más común de metástasis de tumores sólidos, existen diferencias en los patrones metastásicos, tanto a nivel de localización, secuencial y temporal, en función del órgano donde se localice el tumor primario. Por ejemplo, el cáncer de mama suele tardar en metastatizar y, de hacerlo, lo hace en huesos, pulmones, hígado y cerebro; en contrapartida, el de colon suele hacerlo con un patrón secuencial: primero en el hígado y, posteriormente, en el pulmón. Además, también el daño causado en el tejido huesped puede ser diferente en función del tumor primario, como por ejemplo en el tejido óseo en el cual el

cáncer de mama y de pulmón tienen un efecto osteolítico -rotura de los huesos-, mientras que el de próstata es osteobástico -formación de huesos- (Nakamura et al. (2006)).

- *Células asociadas con la metástasis y extravasación:* existen diferentes tipos celulares que se han relacionado con la metástasis y la extravasación de las CTC, como los *macrófagos asociados a la metástasis (MAM)*, los monocitos, los neutrófilos o las plaquetas (Zijl, Krupitza, and Mikulits (2011), Kitamura, Qian, and Pollard (2015)):
 - *MAM:* estos se originan a partir de los monocitos reclutados de la circulación atraídos por citoquinas como CCL2 secretadas por las células de los nichos premetastásicos y coexisten con otros macrófagos residentes en el tejido huésped. Estos son esenciales para la angiogénesis -secretan VEGFA-, la extravasación, la supervivencia y el crecimiento de las células tumorales. En concreto, MAM se une al VCAM1 de las células tumorales mediante la Integrina- α -4, lo que activa la señalización por Akt en las células tumorales y promueve su supervivencia.
 - *Neutrófilos:* las CTC encalladas en las paredes de los vasos sanguíneos atraen a los neutrófilos mediante la secreción de IL8, que se activan y secretan las conocidas como trampas extracelulares de neutrófilos, compuestos principalmente por enzimas y ADN, que capturan las CTC cercanas para incorporarlas al foco metastásico e incrementa la masa del coágulo tumoral. Posteriormente, secretan IL1B para activar a las células endoteliales y MMP para facilitar la extravasación de las CTC.
 - *Plaquetas:* cuando están asociadas a los coágulos tumorales contribuyen a su extravasación a través de diferentes mecanismos. Por un lado, secretan ATP para incrementar la permeabilidad de las células endoteliales de los vasos sanguíneos, pues actúa sobre el receptor purinérgico P2RY2 de estas células y facilita la extravasación. De igual modo, también son capaces de facilitar la extravasación mediante su unión a la Integrina α V β 3 de las CTC, lo que induce cambios morfológicos que facilitan su extravasación. Por otro lado, las plaquetas actúan sobre las propias células tumorales a través de la secreción de TGF β el cual induce la EMT.
 - *Monocitos:* estos son reclutados al endotelio activado a través del acoplamiento con la glicoproteína E-selectina endotelial, cuya sobreexpresión es inducida por las CTC. Este acoplamiento produce la retracción de las células endoteliales de los vasos, lo que conlleva la exposición de la membrana basal y disminución de la fuerza de las *tight junctions (uniones estrechas)* entre las células, lo que promueve la extravasación de las CTC.
- *Movilidad:* la migración de las células tumorales también depende de diferentes mecanismos. Uno de ellos son los gradientes de concentración de factores de crecimiento tales como HGF y EGF o de ligandos como CXCR4 o CXCR3-A, de tal manera que las células tumorales (tanto del propio tejido como extravasadas) se dirigen a zonas de mayor concentración de estos. También hay componentes de la matriz extracelular que favorecen la movilidad, como

laminina y tenascina C. Además, las proteasas también juegan un papel clave, pues actúan sobre diferentes proteínas y liberan o activan factores que promueven la movilidad, como, por ejemplo, algunas MMP son capaces de degradar quimioquinas antimigratorias -ligandos CXCLX-, mientras que otras lo son de degradar la epirregulina -unida a la membrana-, lo que provoca la escisión y liberación de su dominio extracelular el cual es similar a EGF (Zijl, Krupitza, and Mikulits (2011)).

- *Inactividad de las células tumorales*: los tumores pueden permanecer en estado de latencia durante largos periodos de tiempo. Además, el microambiente en el que las células tumorales residen impacta directamente tanto en su inactividad como en su “despertar.” Este hecho desconcierta en muchas ocasiones a nivel clínico, pues, de forma eventual, puede pasar que, tras quitar el tumor primario, las células tumorales durmientes residentes en sitios secundarios se reactiven y produzcan metástasis. Para explicarlo, se han descrito diferentes modelos (Shaked et al. (2014), Neophytou, Kyriakou, and Papageorgis (2019)):

- *Latencia celular individual (quiescencia)*: se basa en la inactividad o latencia individual de las células, de tal manera que cada célula tumoral entra en un estado de quiescencia proliferativa y, por ende, no se divide. Este estado de quiescencia proliferativa es debido a los altos niveles de la MAPK p38 activa, bajos niveles de ERK y la presencia de mTOR activada. En cuanto a su supervivencia, esta podría estar relacionada con la presencia de ciertas integrinas en la membrana de las células tumorales, altos niveles de la MAPK p38 y p53, junto con bajos niveles de ERK activo y la inhibición de C-Jun. En contraposición, cuando estas células salen de su estado durmiente y recuperan su capacidad proliferativa, se observa una actividad opuesta en la ruta de las MAPK, pues se activa la ruta de ERK y se inhibe la de p38.
- *Latencia de la masa tumoral*: en este modelo se postula que, aunque las células tumorales continúan proliferando, existe un balance entre el crecimiento y el control de la proliferación como consecuencia de una irrigación insuficiente o por la acción del sistema inmune, por lo que el efecto neto es una masa tumoral que no crece. En este modelo, el crecimiento de las células tumorales es controlado por el sistema inmune: las células T cooperadoras Th1 secretan $INF\gamma$ que ayuda a mantener la latencia (proceso llamado inmunovigilancia). Además, la angiogénesis está inactivada debido al balance entre las señales pro-angiogénicas (como la ruta Ras o la hipoxia) y anti-angiogénicas (p38, MAPK o p53), por lo que el crecimiento tumoral está limitado, a pesar de que las células tumorales en estado de latencia pueden sobrevivir así durante años, incluso hasta décadas.

1.2.7. Inestabilidad genómica y mutación

En una célula normal, si durante la síntesis del ADN se detecta alguna mutación o daño, el ciclo celular se detiene para reparar la mutación o el daño y, a continuación, reentrar en el ciclo celular. Los genes encargados de detener el ciclo celular en caso de detectar alguna mutación son los genes supresores de tumor. Es decir, mantienen la estabilidad genómica. Sin embargo, las células cancerosas, a diferencia de las células normales, presentan mutaciones y/o alteraciones cromosómicas, por lo que se dice que tienen inestabilidad genómica. De hecho, la mutación o delección de genes supresores de tumor durante la progresión tumoral contribuyen a esta inestabilidad genómica. No obstante, también hay otros genes que actúan en el sentido opuesto y que están sobreexpresados o han sufrido mutaciones de ganancia de función, conocidos como oncogenes, los cuales conllevan que las células tumorales proliferen de forma descontrolada.

En las células tumorales existen dos tipos de inestabilidad genómica: inestabilidad cromosómica e inestabilidad en el ADN. Por ejemplo, en los cánceres colorrectales suele manifestarse una inestabilidad de microsatélites (MSI) cromosómica (CIN). La primera consiste en mutaciones en secuencias repetitivas cortas presentes en el genoma conocidas como *microsatélites*, los cuales tienden a formar bucles que propician la aparición de estas mutaciones, y que son consecuencia de defectos en los sistemas *reparadores de apareamientos erróneos del ADN (MMR)* que, como su propio nombre indica, son los encargados de reparar errores de apareamiento en el ADN recién replicado antes de la mitosis. Por lo tanto, las mutaciones en alguno de ellos puede dar lugar a la inestabilidad genómica. Un ejemplo de cáncer donde se observa MSI es el *síndrome de Lynch*, un tipo de cáncer hereditario colorrectal que presenta alguna mutación en uno de los genes MMR: MLH1, MSH2, MSH6 o PMS2. (H. Yamamoto and Imai (2015), Vasen, Tomlinson, and Castells (2015), Kudryavtseva et al. (2016)). No obstante, la inestabilidad genómica debido a estas mutaciones se explicará más adelante en el apartado 1.3.2.2.

En cuanto a la inestabilidad cromosómica, suelen observarse alteraciones numéricas cromosómicas. Por un lado, pueden ganarse o perderse cromosomas enteros, lo que se conoce como aneuploidía, e incluso conjuntos de ellos, poliploidía. Por otro lado, también pueden observarse alteraciones en sus estructuras, como translocaciones, inversiones o delecciones. Por ejemplo, en la figura 1.7 se muestra el cariotipo de células de una leucemia mielógena crónica en el que se aprecian reordenamientos genómicos, de tal manera que se intercambia ADN entre los cromosomas 9 y 22 (Aplan (2006)). Así, a diferencia de las normales, las células cancerígenas son capaces de continuar dividiéndose a pesar de la existencia de este tipo de mutaciones y reordenamientos genómicos. En el apartado “1.3.2.1 Inestabilidad cromosómica” se comentará brevemente este tipo de inestabilidad relacionada con el cáncer de colon.

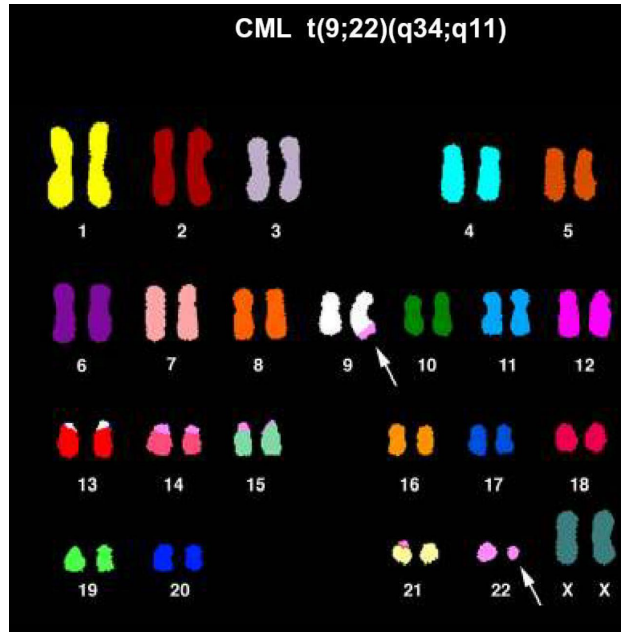


Figura 1.7: Cariotipo de leucemia mielógena crónica. En la figura se puede observar la inestabilidad genómica a través de los reordenamientos genómicos, de la traslocación entre el cromosoma 9 y el cromosoma 22, mostrando en rosa la porción traslocada (modificada de Aplan (2006)).

En resumen, las células tumorales son capaces de ir acumulando progresivamente mutaciones y alteraciones epigenéticas a lo largo del proceso tumoral. Esto se debe a la propia selección natural, pues en un tumor puede surgir una célula con una mutación que la confiera cierta ventaja frente al resto de células tumorales del mismo tejido tumoral. Este proceso de selección de un clon frente al resto puede tener lugar multitud de veces a lo largo de la progresión tumoral, lo que puede ser interpretado como una sucesión de expansiones clonales tal que en cada una de ellas tenga lugar la adquisición de un genotipo “superior” al resto.

1.2.8. Reprogramación metabólica

Una de las principales diferencias entre una célula cancerígena y una normal es su metabolismo, pues se ha descrito una reprogramación metabólica en las tumorales con el fin de que estas células puedan abastecer sus requerimientos energéticos derivados de la elevada tasa proliferativa en un microambiente tumoral en el que hay escasez de nutrientes y condiciones de hipoxia. Por lo tanto, las células tumorales necesitan buscar fuentes alternativas de energía así como otras rutas que les permita metabolizarlas (Hanahan and Weinberg (2011)).

En condiciones normales, las células metabolizan la glucosa, proceso que comienza en el citoplasma, de tal manera que a partir de una molécula de glucosa se obtienen dos de ATP y otras dos de piruvato. Posteriormente, el piruvato atraviesa la mitocondria para, en la matriz mitocondrial, convertirse en Acetil-CoA y entrar en el Ciclo de Krebs. En este ciclo se producen dos tipos de moléculas donadores de electrones: NADH y FADH₂, las cuales, en la membrana

mitocondrial, entran en la cadena de transporte de electrones (formada por cuatro complejos proteicos y la ATP sintetasa). En estos complejos tanto el NADH como el FADH₂ ceden los electrones para que se transfieran al O₂ y así obtener H₂O metabólico. Además, en este proceso también se traslocan protones al espacio intermembrana, lo que genera un gradiente electroquímico que es aprovechado por la ATP sintetasa para generar más energía en forma de ATP (fosforilación oxidativa) (Nelson and Lehninger (2016)).

No obstante, en las células tumorales el requerimiento energético es mucho mayor que en las normales y, para más *inri*, el metabolismo de la glucosa es ineficiente, puesto que en muchas ocasiones las mitocondrias no realizan correctamente sus funciones. En particular, alteraciones en diferentes rutas de señalización, como por ejemplo las relacionadas con MYC y AKT, pueden favorecer la captación de glucosa a la par de un cambio metabólico. En este sentido, el piruvato es procesado, principalmente, en el citoplasma a través de la *glucolisis anaeróbica*, lo que genera dos moléculas de ATP, etanol y lactato, el cual acaba por acidificar el medio. A este cambio metabólico se conoce como *efecto Warburg*, y se caracteriza por anomalías mitocondriales, situación de hipoxia, entre otras, en células proliferativas. Es más, el propio Warburg, en 1924, planteó la *hipótesis de Warburg* la cual afirma que este cambio metabólico, que estaría causado por un daño mitocondrial, sería la causa de la carcinogénesis (Warburg, Posener, and Negelein (1924), Otto (1956), Iyikesici (2019), Woolf, Syed, and Scheck (2016), Klement (2017), Poff et al. (2019)).

Cabe destacar que este cambio metabólico, si bien es cierto que es más rápido, también es menos eficiente que la fosforilación oxidativa, en la que se obtienen 36 moléculas de ATP por cada molécula de glucosa. Es por ello que las células tumorales tienen una mayor demanda de glucosa. Puesto que la captación de glucosa es una etapa limitante en el metabolismo de la glucosa y que está mediado por las proteínas transportadoras de glucosa GLUT, su expresión está aumentada en células tumorales de diversos tejidos. De hecho, se conoce que muchos de los principales oncogenes aumentan la expresión de GLUT de proteínas glucolíticas, como son c-Myc, K-Ras y BRAF. Aunque hay 4, las isoformas más sobreexpresadas asociadas con la transformación maligna son la 1 y la 3, puesto que estas son las que tienen una Km baja y, por ende, permiten la entrada de glucosa a la célula tumoral aun cuando la concentración de glucosa es baja en sangre (T. Yamamoto et al. (1990)). Cabe destacar que, en condiciones de hipoxia, se produce una expresión de GLUT1 inducido por HIF 1 α (Denko (2008)). Por el contrario, en los tejidos normales a isoforma más habitualmente expresada es la dependiente de insulina GLUT4.

A pesar de la ineficiencia de este cambio metabólico, este nuevo mecanismo permite acumular intermediarios clave en el crecimiento y división celular. Es más, por norma general, las células tumorales son capaces de regular su metabolismo energético con el fin de mantener su crecimiento y su elevada tasa de proliferación. Entre los diferentes mecanismos pro-proliferativos que genera este efecto, destacan: el provisionamiento de ATP en situación de hipoxia, la acidificación del microambiente tumoral, la regeneración de antioxidantes endógenos, o la provisión de fuentes carbonadas para la producción de biomasa (Ferreira (2010)).

En este punto, es interesante introducir brevemente el concepto de estrés oxidativo, pues, a pesar de haberlo mencionado con anterioridad, no ha sido oportuno presentarlo hasta el momento. Una de las características del fenotipo tumoral es el aumento de la producción basal de especies reactivas de oxígeno (ROS), lo que genera un estrés oxidativo y que, en cierto modo, está relacionado con la disfunción mitocondrial característico del fenotipo tumoral (Poff et al. (2019)). Sin embargo, este aumento del estrés oxidativo se mantiene en unos niveles subletales gracias a la regulación positiva de una maquinaria antioxidante. No obstante, a pesar de que la situación redox que se alcanza es tal que podría favorecer la supervivencia y el desarrollo del fenotipo tumoral, tales niveles de ROS en un fenotipo sano serían tóxicos. Obviamente, existe un umbral o punto de no retorno a partir del cual el nivel elevado de ROS induciría la apoptosis también en las células tumorales, lo que apoya las terapias anticancerígenas basadas en inducir que las células tumorales sobrepasen dicho punto, tales como la radioterapia o la quimioterapia las cuales aumentan la producción de ROS. Por lo tanto, por un lado, el aumento de la producción de ROS propicia el desarrollo y la supervivencia de los tumores pero, por otro lado, si se sobrepasa un cierto umbral de producción de ROS, éstas entran en apoptosis de forma irreversible.

Entre las diferentes causas que pueden aumentar ROS se encuentran la inflamación crónica derivada de la hiperglucemia, infecciones o la exposición a ciertos productos como el tabaco. Curiosamente, una característica del fenotipo glucolítico en el cáncer es el aumento del flujo a través de la ruta de las pentosas fosfato, lo que genera NADPH el cual se empleará para regenerar la glutatión reducida (GSH), siendo este último la mayor fuente para la protección antioxidante en los tumores en respuesta a la generación de ROS inducida por radio o quimioterapia. Por ello, si se reduce el flujo glucolítico y, por ende, de las pentosas fosfato, el tumor se hace más propenso a presentar un estrés oxidativo más acelerado. No obstante, el tejido sano parece no sufrir efecto adverso frente a estas terapias.

A nivel clínico, el efecto Warburg ha inspirado técnicas diagnósticas como la topografía de emisión de positrones mediante el uso de fluorodesoxiglucosa, pues permite visualizar tumores en el cuerpo humano como zonas donde hay un gran consumo de esta sustancia. Además, también se han diseñado estrategias terapéuticas basados en este efecto tan pronunciado en las células tumorales conocidas como *quimioterapia apoyada metabólicamente* donde, tras un ayuno de 12 horas, se inyecta al paciente una dosis de insulina, lo que genera una situación de hipoglucemia. Posteriormente, se trata al paciente con quimioterapia, aumentando así considerablemente el estrés metabólico en las células tumorales al no disponer de la suficiente glucosa que tanto requieren y, por ende, aumentando el efecto quimioterápico del tratamiento al incrementar la permeabilidad de la membrana (Iyikesici (2019)). En el mismo sentido, también se han planteado estrategias basadas en dietas cetogénicas que han mostrado buenos resultados (Woolf, Syed, and Scheck (2016), Klement (2017), Poff et al. (2019)).

No obstante, en los últimos años se han identificado más reprogramaciones metabólicas que tienen lugar en el fenotipo tumoral. Algunas de ellas se mencionan a continuación (figura 1.8) (Woolf, Syed, and Scheck (2016), Lewis and Abdel-Haleem (2013)):

- La glutamina supone otro sustrato fermentable por el cual también optan las células proliferativas tumorales. Además de como combustible fermentable, la glutamina también sirve como fuente anaplerótica del ciclo de Krebs, el cual alimentará las vías de síntesis de lípidos y proteínas.
- El metabolismo de la Serina y de un carbono³ se ve aumentado para contribuir a la síntesis de nucleótidos y modificaciones epigenéticas.
- Precursores de Acetil-CoA, como la glucosa, la glutamina y el acetato constituyen una fuente de metabolitos para la síntesis de ácidos grasos, ruta que es crítica en algunos tipos de cáncer. El Acetil-CoA también es importante como donador de grupo acetyl en la acetilación epigenética para la regulación en la expresión de genes supresores tumorales y oncogenes.
- Algunas de las mutaciones clásicas de oncogenes también confieren un fenotipo metabólico que ayuda a la supervivencia y proliferación de las células tumorales. De hecho, cada una de los Hallmarks de Hanahan y Weinberg del cáncer pueden relacionarse con alguna alteración metabólica específica.
- Cabe destacar que muchos tumores, aparte de la clara disfunción mitocondrial, también carecen de ciertas enzimas necesarias para el metabolismo de los cuerpos cetónicos. Así, estas células, como mecanismo de supercompensación, pasan a un metabolismo basado en la producción de energía a través de sustratos fermentables, lo que se ha observado en numerosos estudios mediante el aumento del lactato en tumores respecto del tejido sano (Klement (2017)).
- El aumento del flujo glucolítico incrementa la vía de las pentosas fosfato, una ruta metabólica que diverge de la glucólisis y que proporciona precursores de antioxidantes endógenos que permiten la supervivencia de las células tumorales aún en situación de elevado estrés oxidativo. De hecho, se ha visto que, bajo una dieta cetogénica, se reducen los niveles el marcador tumoral TKTL-1 de la vía de las pentosas fosfato, lo que está ligado a la reducción de la glucólisis en células tumorales bajo una dieta cetogénica (Jansen and Walach (2016)).

³El metabolismo de un carbono engloba los ciclos del folato y de la metionina. Este permite a las células generar grupos metilo (de ahí su nombre) para que estas los utilicen en la biosíntesis de precursores anabólicos o para reacciones de metilación.

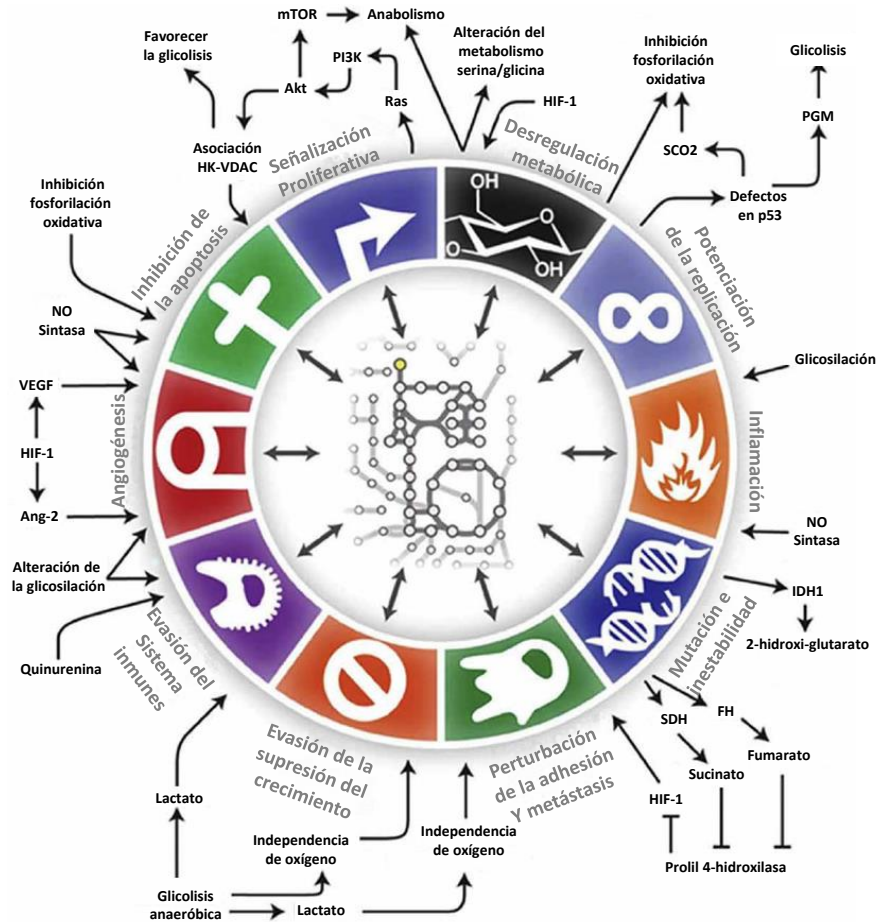


Figura 1.8: Reprogramaciones metabólicas implicadas en diferentes *hallmarks*. En la figura se puede observar cómo algunos de los *hallmarks* afectan al metabolismo de la célula tumoral e incluso cómo el propio metabolismo modula a alguno de ellos (Lewis and Abdel-Haleem (2013)).

1.2.9. Evasión del sistema inmune

Este es uno de los sellos más novedosos que explica cómo las células cancerígenas son capaces de vivir en el organismo huésped. En condiciones normales, el sistema inmune detecta una amplia variedad de agentes, desde virus hasta parásitos, debido a su capacidad para distinguir lo propio de lo extraño. Está formado por diferentes órganos: el timo, los ganglios linfáticos, bazo, apéndice, médula ósea y vesículas linfáticas. Cada uno de ellos es capaz de producir diferentes tipos de células: los ganglios linfáticos producen las células T, las cuales maduran en el timo (de ahí su nombre); las células de estirpe mieloide llegan a ser macrófagos y se producen en gran parte en todos los órganos del cuerpo; las células dendríticas y las células B, que son las células que presentan los anticuerpos, secretan citoquinas. Entre las diferentes funciones de las células del sistema inmune, está la fagocítica y la destrucción de agentes extraños o ajenos, que pueden ser llevadas a cabo por células B, linfocitos T, macrófagos y células NK (Hanahan and Weinberg (2011), Kitamura, Qian, and Pollard (2015)).

En particular, hay unas proteínas transmembrana que juegan un papel de suma importancia en la supresión del sistema inmune conocidas como el *ligando 1 de muerte programada (PDL1*, del inglés *programmed death ligand 1*) o el 2, que pueden aparecer en células tumorales o macrófagos. Estas juegan un importante papel a la hora de suprimir el sistema inmune en situaciones tales como el embarazo, aloinjertos de tejido, enfermedades autoinmunes entre otras. Por lo general, el sistema inmune reacciona a los antígenos y células extrañas con el fin de activar la proliferación de células T CD8 antígeno–específicas las cuales expresan el receptor de membrana de *muerte celular programada 1 (PD-1*, del inglés *programmed death 1*). Cuando este receptor entra en contacto con PDL1 manda una señal a células T para que detengan su proliferación con el fin de evitar auto–inmunidad (Kitamura, Qian, and Pollard (2015)).

Así, las células cancerígenas usan este mecanismo como estrategia para eludir al sistema inmune a través de la sobreexpresión de PDL1 para evitar ser detectadas como “extrañas” por las células T y que estas no las ataquen. En este sentido, PD-1 es una potencial diana terapéutica, por lo que se están investigando fármacos que son antagonistas de estos receptores con el fin de que las células T sean capaces de detectar a las células tumorales. No obstante, existen otros mecanismos. Por ejemplo, las células tumorales pueden dejar de expresar antígenos inmunogénicos y moléculas presentadoras MHC-I, o experimentar mutaciones en proteínas apoptóticas o de supervivencia que les hagan menos susceptibles a las acciones citotóxicas, entre otros (Kitamura, Qian, and Pollard (2015)).

Bajo este contexto, aparecen dos teorías o términos ampliamente usados. Por un lado, surge la teoría de la *inmunovigilancia* la cual defiende que el sistema inmune es capaz de reconocer los tumores como cuerpos extraños y eliminarlos. Sin embargo, la heterogeneidad de las células tumorales puede conducir a clones celulares con inmunogenicidad reducida y, además, capaces de protegerse del sistema inmune. Por otro lado, surge el término *inmunoedición*, que se refiere a la adaptación de las células tumorales en respuesta al sistema inmune (Hanahan and Weinberg (2011), Kitamura, Qian, and Pollard (2015)).

1.2.10. Estado inflamatorio inducido por el tumor

El estado inflamatorio que se genera en el microambiente tumoral, tanto en lesiones premalignas como malignas, puede llegar a ser protumoral y favorecer el crecimiento tumoral y la metástasis. Como se vio en el apartado “1.2.6, Activación de la invasión y la metástasis” , el microambiente tumoral presenta infiltración de células inmunes (TAM, CAF...) que permiten a los tumores generar un estado inflamatorio. Es más, este estado inflamatorio se observa en etapas tempranas de varios tipos tumorales. La inflamación en estadios tempranos del proceso tumoral puede liberar sustancias químicas en el microambiente tumoral que favorecen ciertas mutaciones y favorecen el proceso tumorogénico e, incluso, fomentan un fenotipo invasivo y más agresivo mediante la EMT (Zijl, Krupitza, and Mikulits (2011), Kitamura, Qian, and Pollard (2015)).

Por un lado, algunos tipos de células estromales del microambiente tumoral que pueden incrementar la agresividad de las células tumorales son los TAM, macrófagos presentes en tumores que secretan diferentes tipos de proteínas solubles. Entre estas proteínas están las citoquinas con perfil protumoral, como $TGF\beta$ o IL10, que bloquean las funciones proinflamatorias de los macrófagos e inducen a los monocitos extravasados a adquirir un fenotipo TAM. Otro tipo de productos metabólicos, como la prolina o las poliaminas, que forman la matriz extracelular, juegan un papel importante en la consolidación de la matriz extracelular del tumor. Estos TAM también secretan factores angiogénicos que desencadenan el inicio y el crecimiento de nuevos vasos sanguíneos. En consecuencia, todos estos mecanismos activan las rutas de supervivencia y proliferativas de las células tumorales. Es más, los TAM, *células T reguladoras (Treg)* y *células supresoras derivadas de origen mieloide (MDSC, del inglés myeloid-derived suppressor cells)* son capaces de suprimir al sistema inmune e incrementar la agresividad del tumor (Hanahan and Weinberg (2011), Zijl, Krupitza, and Mikulits (2011), Kitamura, Qian, and Pollard (2015)).

Por otro lado, las células tumorales son capaces de reclutar o de promover la diferenciación de las Treg. Estas células inducen un microambiente regulatorio que inhibe la actividad citotóxica de las células inmunes (CTL...) y, por lo tanto, son capaces de favorecer el desarrollo tumoral. Es más, las MDSC pueden inhibir la actividad antitumoral de las células T a través del contacto MDSC-célula T y también mediante la secreción de citoquinas (Kitamura, Qian, and Pollard (2015)).

En resumen, las células del sistema inmune producen y secretan ciertas citoquinas capaces de regular la proliferación y la acumulación de células inmunitarias. Además, también secretan diferentes proteasas. Por otro lado, las propias células cancerígenas también pueden secretar factores como quimiocinas que conllevan que las células inmunes puedan secretar estos factores que, a su vez, estimulan a las células cancerígenas favoreciendo el aumento de la motilidad celular, movimiento celular, supervivencia celular y la angiogénesis. Todo esto permite a las células tumorales crecer, sobrevivir e invadir. Por lo tanto, estrategias terapéuticas podrían dirigirse hacia las propiedades protumorales de las células inmunes y mejorar los mecanismos antitumorales. Una combinación de *inmunomoduladores* y de *agentes citotóxicos* podría ser la opción más razonable.

1.3. Cáncer de colon

Tras haber presentado una visión general del cáncer y de sus *hallmarks*, se procede a presentar el cáncer de colon (CCR). En este sentido, se va a realizar una breve introducción en la que se hablará de la incidencia y de los factores de riesgo. A continuación, se hará una visión más fisiopatológica en la que se hablará de las bases de la biología molecular del mismo.

1.3.1. Incidencia del cáncer de colon y sus factores de riesgo

En 2020, se diagnosticaron a 1 931 590 personas de CCR en el mundo, lo que le situó en el tercer puesto de incidencia. Además, murieron 935 173 personas a causa de esta enfermedad, convirtiéndolo en la segunda causa de muerte por cáncer, por detrás del de pulmón. Sin embargo, aunque en España ocupa el primer puesto en incidencia, con 40441 personas diagnosticadas, dos puestos por encima respecto del resto del mundo, también ocupa el segundo en mortalidad con 16470 muertes en 2020 (figura 1.9). Cabe destacar que este cáncer afecta de forma diferente a hombres que a mujeres, pues tanto la incidencia como mortalidad es superior en hombres. En particular, en 2020 el 55.19 % de los casos diagnosticados y el 55.14 % de muertes por CCR fue para los hombres, mientras que en España estos porcentajes fueron del 60.85 % y de 58.53 %, respectivamente (Sung et al. (2021)).

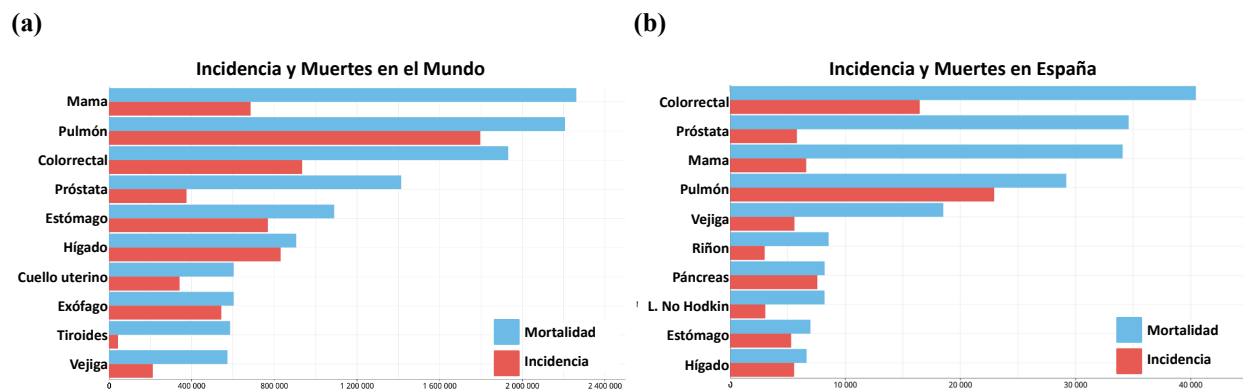


Figura 1.9: Incidencia y mortalidad de los cánceres más comunes (a) a nivel mundial y (b) en España (Datos extraídos de Organización mundial de la salud (2021), Observatorio Global del Cáncer (2021) y Sung et al. (2021)).

A pesar de que esta enfermedad era poco habitual en la década de los 50 del siglo XX, es obvio que ha escalado puestos en estos últimos 70 años, tanto en incidencia como en mortalidad. Esto puede explicarse por la aparición esporádica de este como consecuencia de una serie de factores de riesgo: el mayor envejecimiento de la población, tabaquismo, hábitos alimenticios pobres, el exceso de consumo de carnes rojas, sedentarismo, obesidad, padecer enfermedades intestinales inflamatorias e, incluso, la exposición a agentes infecciosos como *Helicobacter pylori*. Es más, en relación a la dieta, este tipo de cáncer también se conoce como *enfermedad del oeste* o del *Nuevo Mundo*, pues su incidencia en zonas no urbanas en las que no se consume comida procesada ni

carne es prácticamente nula (Dixon et al. (2004)). Además, en cuanto al sexo, si bien es cierto que la incidencia y la mortalidad es superior en hombres, también se ha observado un aumento en la incidencia en mujeres postmenopáusicas. No obstante el aumento de la incidencia no sólo se da en el CCR esporádico, sino también en algunos de los síndromes de CCR hereditario como la enfermedad de Lynch, un tipo hereditario de CCR no polipomatoso (Capelle et al. (2010), Vasen, Tomlinson, and Castells (2015)).

A la contra, se han descrito también diferentes estrategias para prevenir esta enfermedad: terapia de reemplazo hormonal, tratamiento con ácido acetil salicílico (aspirina), cambio en los protocolos alimenticios basados en dietas ricas en fibra (≈ 50 g/día, Mann, Sidhu, and Gowin (2020), Van Blarigan et al. (2018)), realización de actividad física... En relación a la aspirina, esta ha demostrado reducir alrededor de un 25 % el riesgo de sufrir CCR e, incluso, hasta un 50 % en algunos de los síndromes de CCR hereditario (Burn et al. (2011)).

En cuanto a la terapia de reemplazo hormonal, el aumento de la incidencia del CCR observado en mujeres postmenopáusicas causado por la disminución en la expresión del ligando E2 del *receptor de estrógenos* (*ER*, del inglés *estrogen receptor*) ha promovido el uso de estas. En consecuencia, se ha observado que esta genera protección frente a esta patología, puesto que se reduce el tamaño y la incidencia de los adenomas (Butcher et al. (1985), Chlebowski et al. (2004)). En este sentido, también se ha observado la importancia de la dieta en la prevención de esta, pues estudios epidemiológicos como el de Messina et al. (1994) han mostrado una reducción en la incidencia del CCR en la población femenina asiática con una dieta rica en soja, alimento rico en el fitoestrógeno isoflavona genisteína -ligandos exógenos de ER-, puesto que son capaces de mimetizar la acción de E2 y modular la actividad de los ER β (Messina et al. (1994), Newmark (1996), Hébert-Croteau (1998), Jang et al. (1997)).

1.3.2. Vías de desarrollo del CCR

Como se ha dicho anteriormente, el CCR puede ser causado tanto por factores genéticos como ambientales que, en cierto modo, favorecen que las células de las que parte vayan adquiriendo progresivamente los *hallmarks* tratados en el apartado “1.2”. Entre otras, se observa la acumulación progresiva de mutaciones y modificaciones epigenéticas que conllevan a su autonomía proliferativa a través de la activación de oncogénos y la inactivación de genes supresores de tumores. También se observa una elevada inestabilidad genómica y epigenómica en la inmensa mayoría de las lesiones neoplásicas de colon -focos de cripta aberrantes, adenomas y pólipos serrados-, la cual podría jugar un papel clave en el inicio del CCR. En consecuencia, estas células acumulan errores de forma progresiva que, a la larga, conducen a la transformación maligna de estas a través de rondas de expansión clonal en las que se seleccionan las células de carácter más agresivo. No obstante, a día de hoy todavía existe el debate de qué tipo celular es la que origina el CCR: célula madre o similar a una célula madre localizada en la base de la cripta (Zeki, Graham, and Wright (2011)).

En cuanto a la aparición del CCR, hay que diferenciar entre aquellos pacientes que tienen mutaciones heredadas -síndromes de CCR familiar- o los que, por el contrario, no tienen un historial de antecedentes familiares ni modificaciones del material genético heredadas -esporádicos-. Por un lado, los primeros representan entre el 10 y el 15 % de los CCR y pueden clasificarse en dos grupos que, a su vez, pueden ser divididos en función de la patogenicidad de los pólipos y de la mutación que presenten (Vasen, Tomlinson, and Castells (2015)):

- *Síndromes polipomatosos*, en los que se produce un desarrollo de múltiples pólipos en el colon. Según la mutación que presenten, algunos de los más importantes son: la *polipomatosis APC*, que es la poliposis adenomatosa familiar clásica; *polipomatosis MUTYH*; *polipomatosis adenomatosa asociada con mutaciones bialélicas en MMR*; *polipomatosis SMAD4*; *polipomatosis PTEN*, *polipomatosis serrada*... Por ejemplo, la polipomatosis APC es un síndrome autosómico dominante hereditario debido a mutaciones en APC que favorece el desarrollo de cientos o miles de pólipos entre la segunda y tercera década de vida del individuo. No obstante, la severidad de esta patología depende en gran medida de la localización de la mutación, por ejemplo, si se localiza en el fin del gen o en el exón 9 entonces tiene un fenotipo *atenuado*. Además, el tratamiento temprano de estos es de suma importancia, pues la probabilidad de desarrollar CCR es superior al 80 %.
- *Síndromes no polipomatosos*, en los que los pólipos que se desarrollan son pocos o ninguno. Estos pueden dividirse, a su vez, en tres que, además, también se subdividen en función de la mutación presentada: *síndrome de Lynch*, que, a su vez, puede ser *síndrome de Lynch MLH1, MSH2*...; *probable síndrome de Lynch*, que es cáncer colorrectal u otros síndromes de Lynch que presentan inestabilidad de microsatélites, y *cáncer colorrectal familiar*, que no presentan inestabilidad de microsatélites. De los tres, el síndrome de Lynch es el más común. Es un síndrome autosómico dominante hereditario que se caracteriza por un inicio temprano no solo de CCR, sino que también puede provocar cáncer del tracto urinario, gástrico, de endometrio... Puede presentar mutaciones en uno de los genes MMR, los cuales son los encargados de reparar errores en el ADN recién replicado antes de la mitosis: MLH1, MSH2, MSH6 o PMS2. Por lo tanto, las mutaciones en alguno de ellos puede dar lugar a la inestabilidad genómica. Estos pacientes se caracterizan por una carcinogénesis acelerada consecuencia de la acción de oncogenes y supresores de tumores mutados como consecuencia de las mutaciones en algún MMR. Como resultado, estos pacientes tardan menos de dos años en desarrollar el cáncer frente a los más de 10 que tarda un CCR espontáneo, por lo que es recomendable que estos pacientes comiencen a los 25 años a realizarse colonoscopias, así como de los otros tipos de cáncer que pueden desarrollar.

Por otro lado, la mayor parte de los CCR, en concreto entre el 85 y 90 %, son esporádicos. Por lo tanto, la causa principal de aparición de estos son los factores de riesgo anteriormente comentados. En general, el desarrollo del cáncer de colon puede darse a través alguna de las dos vías siguientes: la *vía clásica*, que involucra a adenomas capaces de progresar hacia adenocarcinomas, y la *vía*

serrada, en la que a partir de pólipos serrados -grupo heterogéneo de lesiones con potencial de transformación- se puede desarrollar cáncer colorectal serrado. En ambas tiene lugar la progresión de células epiteliales normales de colon hacia focos de criptas aberrantes, seguido de pólipos tempranos y avanzados con la consecuente progresión hacia el fenotipo cancerígeno temprano y avanzado (Ernst J. Kuipers (2015)).

Se han descrito una serie de alteraciones en diferentes genes, muchas de las cuales son comunes a ambas vías mientras que otras son exclusivas de cada una de ellas, como por ejemplo la mutación de BRAF o el fenotipo metilador de las islas CpG de la vía serrada o la alteración de APC de la clásica. No obstante, las alteraciones en la ruta de señalización Wnt es la forma más frecuente de inicio del CCR (figura 1.10, Mundade et al. (2014), Ernst J. Kuipers (2015)). También cabe destacar que los *receptores nucleares* (NR, del inglés nuclear receptor) tienen un papel activo en el CCR, parte de ellos como supresores tumorales y otros favoreciendo el desarrollo tumoral (figura 1.12, D'Errico and Moschetta (2008)).

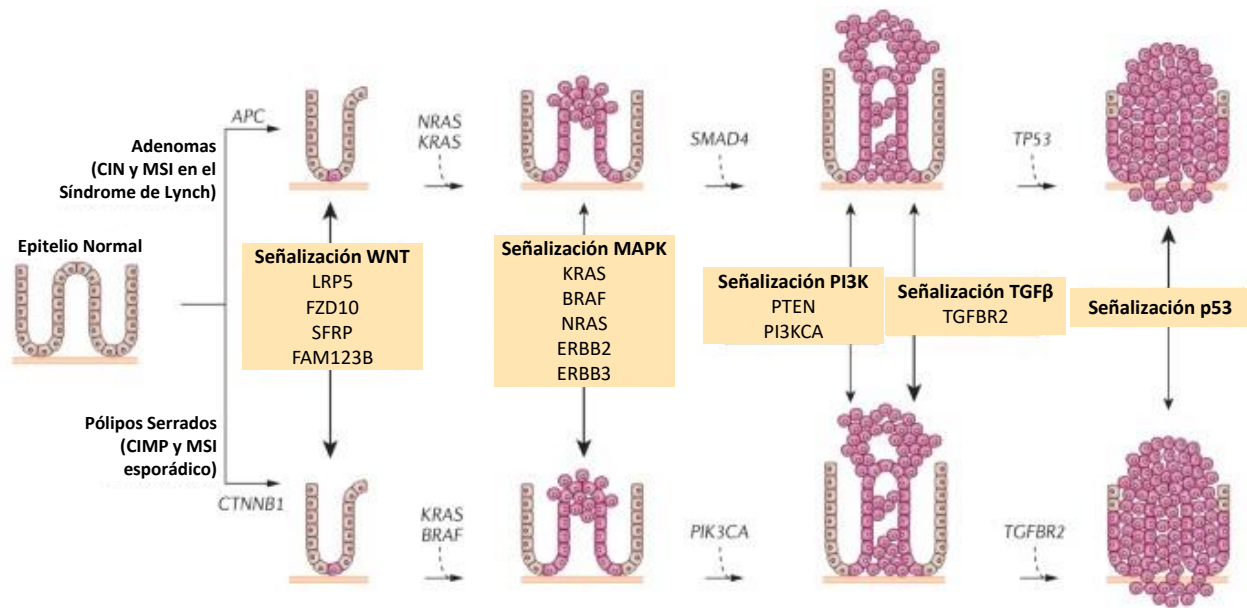


Figura 1.10: Vías implicadas en el desarrollo de cáncer colorectal. En la figura se muestran las dos vías principales del desarrollo del cáncer colorectal, en cada una de las cuales se muestran sobre las flechas horizontales algunos de los genes mutados o modificados epigenéticamente más relevantes en esta enfermedad. En las cajas amarillas se indica la ruta de señalización más importante en cada paso del desarrollo de la enfermedad, de tal manera que el grosor de las flechas verticales indica la importancia de cada una de estas sobre cada una de las vías del desarrollo tumoral. Abreviaturas: APC, poliposis coli adenomatosa; CIN: inestabilidad cromosómica; CTNNB1, β -catenina1; FAM123B, familia con similaridad de secuencia 123B o AMER1; FZD10, receptor frizzled 10; LRP5, proteína relacionada con la lipoproteína de baja densidad 5; MAPK, proteína quinasa activada por mitógenos; MSI: inestabilidad de microsatélites; PI3K, fosfatidilinositol 3 quinasa; PI3KCA, subunidad cataítica α de la fosfatidilinositol-4,5bifosfato quinasa 3; PTEN, fosfatasa y homólogo de tensina; SFRP, proteína relacionada con frizzled secretada; SMAD4, miembro 4 de la familia SMAD; TGF β , factor de crecimiento transformante β ; receptor 2 de TGF β ; CIMP: fenotipo metilador de islas CpG (Modificado de Ernst J. Kuipers (2015)).

Con el fin de arrojar luz sobre la figura anterior (figura 1.10) y mejorar la comprensión de las vías implicadas en el desarrollo del CCR, a continuación, se explican brevemente las rutas de señalización Wnt/ β – catenina y MAPK, pues la *regulación negativa de PI3K/AKT por PTEN*, la *señalización por TGF β* y la *señalización por p53* han sido explicadas en “1.2.2 Evasión de las señales inhibitorias del crecimiento” (figura 1.11).

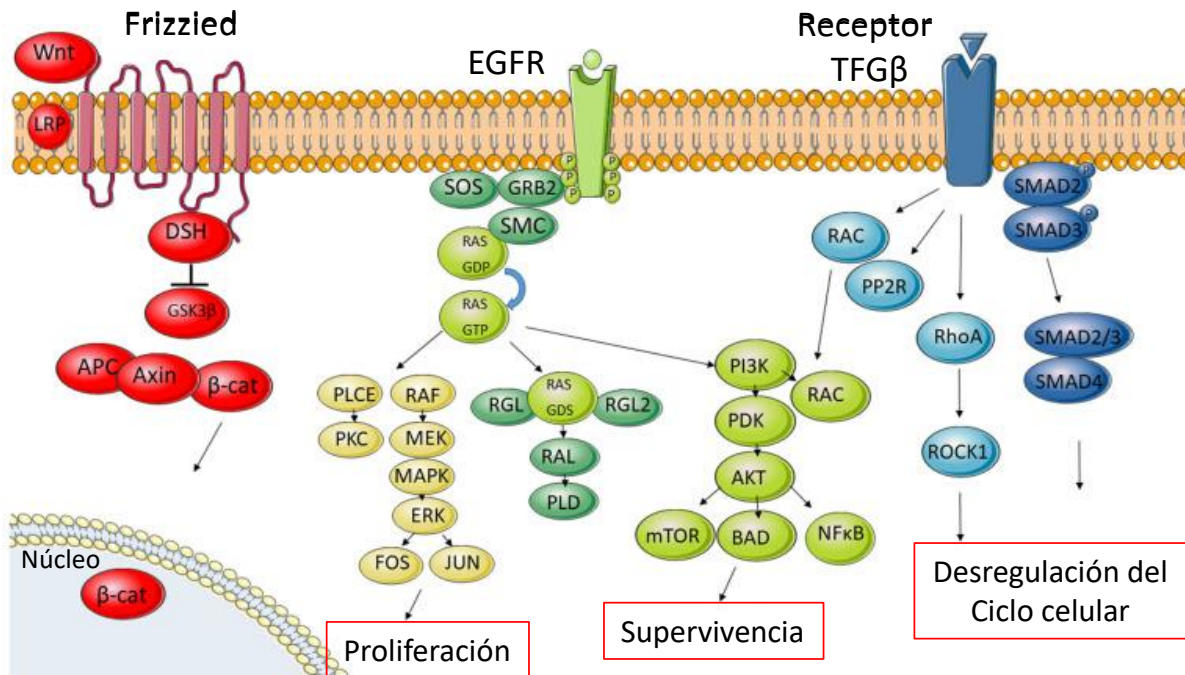


Figura 1.11: Rutas clave en el cáncer de colon. En el cáncer de colon una de las rutas más relevantes es la Wnt/ β – catenina, pues la mayoría de los tumores de colon comienzan por alteraciones en esta ruta. Esta, en condiciones normales y sin ser activada, permite la degradación de la β – catenina hasta que su ligando Wnt activa al receptor *Frizzled* y la β – catenina comienza a acumularse en citosol y se transloca al núcleo. Esta acumulación también puede ocurrir sin necesidad de la activación del receptor causado por alteraciones en miembros de esta ruta, como APC. La activación de esta ruta está relacionada con el aumento de la proliferación, de la motilidad y de la polaridad celular. Las rutas de señalización vía receptores de factores de crecimiento a través de la activación de TKR conlleva la activación de la proteína Ras la cual desencadena, corriente abajo, la activación de la ruta de las MAPK o de PI3K-Akt. Alteraciones en esta ruta pueden resultar en el aumento del crecimiento, la proliferación y la supervivencia celular. Otra ruta muy importante en el cáncer de colon es la señalización por TNF – β , de tal manera que la activación de sus receptores da lugar a la formación del complejo SMAD2/3/4 el cual se transloca al núcleo para regular la transcripción de diferentes genes. En las células tumorales, la pérdida de SMAD2 y SMAD4 confiere a las células resistencia a la apoptosis y la desregulación del ciclo celular. Cabe destacar que todas estas vías están interrelacionadas, pues la ruta Wnt está relacionada con mTOR, Notch o Rac/KRas, y la señalización por TNF – β activa las rutas MAPK, PI3K, Notch y Wnt (modificado de Poturnajova et al. (2021)).

1.3.2.1. Inestabilidad cromosómica

Como se dijo anteriormente, la inestabilidad cromosómica es un tipo de alteración genómica a nivel cromosómico que puede manifestarse como alteraciones numéricas (aneuploidía, poliploidía) o estructurales (translocaciones, inversiones, deleciones...). Los tumores que presentan este tipo de inestabilidad suelen presentar mutaciones en diferentes genes supresores de tumores como el *APC* o *SMAD2/4*, y en oncogenes como *KRAS* o *PIK3CA* (da lugar a PI3K). En concreto, en cáncer de colon, CIN es más frecuente (65-85 %) que MSI en los que, precisamente, una de las mutaciones más comunes vía clásica es el gen *APC*. En este tipo de cáncer, los mecanismos más habituales de desarrollo de CIN son: los relacionados con la degregación cromosómica, la formación del telómero y la reparación del ADN; alteraciones en el sistema del centrosoma, incluido la formación de centros adicionales de organización de microtúbulos, o la disfunción de los telómeros, que a su vez causa alteraciones como *puentes de fusión-rotura* que suelen resultar en la amplificación de oncogenes localizados en regiones teloméricas (Kudryavtseva et al. (2016)).

1.3.2.2. Intestabilidad de microsatélites y mecanismos de reparación de apareamientos erróneos

En caso de producirse un apareamiento erróneo durante la replicación del ADN se activan los MMR. La detección de estos apareamientos erróneos comienza con el escaneo por parte del complejo proteico MUTSA, pues es capaz de detectar nucleótidos no complementarios en la doble hebra. Tras detectar el error de apareamiento, otros dos complejos proteicos entran en juego: MUTLA, que corta el fragmento de ADN dañado, y EXO1, que lo degrada. Así, tras la escisión del apareamiento erróneo, es el turno de rellenar el hueco que se ha creado, para lo que entran en juego la ADN polimerasa- δ , que rellena el hueco con el nucleótido correcto, y la ADN ligasa, que se encarga de unir el nucleótido reparado a la hebra de ADN.

No obstante, en algunos tipos de cáncer colorrectal, sobretudo en síndrome hereditario de Lynch, las proteínas MHL1, MSH2 y/o MSH6, que forman parte del complejo MUTSA, están mutadas. En consecuencia, tiene lugar una elevada tasa de mutaciones como, por ejemplo, en oncogenes o en genes supresores de tumores como el APC. Además, las alteraciones en los componentes de MMR también pueden dar lugar a apareamientos erróneos entre las bases de los microsatélites. Por otra parte, están los microsatélites, secuencias de ADN en las que un fragmento de 2 a 6 se repite sucesivamente en la secuencia y suponen un 3 % del genoma. Generalmente, estos se localizan en las secuencias UTR e intrones, y raras veces en los telómeros y exones. En algunos tumores, una de las consecuencias de que los componentes de MMR estén alterados es el apareamiento erróneo entre las bases de los microsatélites, lo que da lugar a un aumento o disminución de la longitud de estos respecto del ADN normal, hecho que se conoce como MSI.

1.3.2.3. Hiperactivación de la ruta de señalización Wnt/ β – catenina

En los mamíferos existe una proteína citoplasmática llamada β – catenina que, cuando se acumula en el citosol, se transloca al núcleo lo que activa la expresión de genes implicados en procesos tales como la proliferación o diferenciación celular. Así, mientras que en ausencia de β – catenina los factores de transcripción *TCF/LEF* (*factor de células T/Factor Estimulador Linfoide*)⁴ están unidos a la proteína co-represora de la transcripción *Groucho* p *HDAC* y, por ende, se inhibe la transcripción de una serie de genes; en presencia de β – catenina las proteínas correpressoras son desplazadas del complejo *TCF/LEF* por la interacción de la β – catenina y, a su vez, se reclutan diferentes co-activadores que favorecen la transcripción, como por ejemplo *CBP*, *p300* o *BCL9*, de tal manera que el complejo β – catenina-*TCF/LEF* actúa como factor de transcripción de diferentes genes relacionados con la proliferación, crecimiento y diferenciación celular, entre los que cabe destacar *MYC* (c-Myc), *JUN* (c-jun), *CCND1* (CycD) u otros que codifican para alguna MMP (Clevers and Nusse (2012), Logan and Nusse (2004)).

En condiciones normales, en las células existe un complejo de destrucción de la β – catenina que evita la acumulación de esta en el citosol, formado por proteínas como *Axina*, *APC* (*poliposis coli adenomatosa*), *GSK3 β* (*quinasa glucógeno sintasa β*) y *CK1* (*Quinasa Caseína 1*) (Ochoa-Hernández et al. (2012), Hart et al. (1998), Amit et al. (2002), C. Liu et al. (2002)):

- *GSK3 β* y *CK1* (*CSNK1A1*): son las quinasas encargadas de fosforilar a la β – catenina, de tal manera que es la quinasa *CK1* la que primero fosforila a β – catenina y la hace así reconocible por la quinasa *GSK3 β* para que esta la pueda fosforilar a continuación. Esta secuencia es necesaria para su posterior poliubiquitinación por medio de la β – TrCP (*E3 ubiquitina ligasa*) con el fin de marcarla para su degradación en el proteosoma y, por ende, se evita su traslocación al núcleo cuando esta se acumula en el citoplasma, por lo que no podrá actuar como factor de transcripción. Cabe destacar que *GSK3 β* también fosforila tanto a la *Axina* como a *APC*, lo que estabiliza a la *Axina* y facilita la interacción de *APC** con la β – catenina (Dianqing Wu and Pan (2010)).
- *APC*: proteína que participa en multitud de procesos, tales como la destrucción de la β – catenina, la adhesión celular o la estabilidad de los microtúbulos. En particular, la fosforilación de *APC* permite su unión con β – catenina que, por un lado, facilita su fosforilación por las quinasas *GSK3 β* y *CK1* y, por otro lado, evita que la proteína *PP2A* (*Fosfatasa 2A*) la defosforile. Cabe destacar que se ha descrito que en el 85 % de los cánceres de colon presentan mutaciones de pérdida de función en esta proteína (en dominios reguladores de β – catenina, tanto de unión a esta como otros), lo que conlleva a la acumulación de β – catenina en el citosol, por lo que es un *gen supresor de tumores* de gran importancia (Polakis (1999)).

⁴Conjunto de factores de transcripción de la ruta Wnt. Las células de mamíferos expresan cuatro tipos: *TCF1* (*TCF7*), *LEF1*, *TCF7L1* (*TCF3*), y *TCF7L2* (*TCF4*).

- *Axina*: proteína del complejo que actúa como ensamblador de las proteínas del complejo al poseer sitios de unión a dichas proteínas, lo que ayuda a la acción del *APC* y las quinasas *GSK3 β* y *CK1* sobre la β – catenina (Hart et al. (1998)).

No obstante, existe una ruta de señalización, conocida como Wnt/ β – catenina, que evita su degradación. El proto-oncogén *Wnt* -se conocen diferentes vías de señalización y es secretado por diferentes tejidos-, cuyo mecanismo de señalización suele ser autocrino, se une a complejos de receptores heterodiméricos; en concreto, a su receptor de siete segmentos *Frizzled* -a través de su dominio extracelular rico en cisteína (CDR)-, que lo activa y este, a su vez, recluta al correceptor *LRP5/6* (*receptor de lipoproteínas de baja densidad relacionados con la proteína 5/6*), de tal manera que el complejo formado genera una zona de unión para la proteína citoplasmática *Dvl/Dsh* (*Dishevelled*). Esta, tras su activación (fosforilación) a través de ciertas unidades de las proteínas G, captura al complejo de destrucción de la β – catenina y lo bloquea. En particular, este complejo se une al *LRP5/6* fosforilado y a *Dvl* a través de la *Axina* y facilita la acción de las quinasas *GSK3 β* y *CK1* sobre *LRP5/6* y, posteriormente, puesto que *Dvl* se une tanto a las *GBP/Frat* (*Proteínas de Unión a GSK3*), que compiten con la *Axina* por la unión con *GSK3 β* , como a la *Axina*, se facilita enormemente la interacción entre estas y, por lo tanto, la unión de *GSK3 β* con *GBP* y su consecuente disociación de la *Axina*, de tal manera que *GSK3 β* se separa de la *Axina* y se reduce así su actividad quinasa⁵, de modo que ya no podrá fosforilar a β – catenina y esta se estabilizará en el citosol (Clevers and Nusse (2012)).

La ruta Wnt/ β – catenina también está estrechamente relacionada con otras de relevantes del cáncer como *Hedgehog*, *mTOR*, *Notch* o *Rac/KRas*. Por lo tanto, la regulación positiva de β – catenina a través de la ruta de señalización Wnt/ β – catenina o su regulación negativa por medio de su degradación por acción del complejo de destrucción de la β – catenina, median y regulan diferentes procesos tales como la proliferación y diferenciación celular, o la regeneración de tejidos y la capacidad invasiva de las células. Es más, la hiperactivación de esta ruta o la represión de la degradación de β – catenina está presente en diferentes tipos de tumores (hígado, melanomas, colon...), lo que puede deberse a la sobreexpresión de Wnt o a mutaciones en β – catenina que evitan su degradación. En concreto, en el CCR se han observado mutaciones oncogénicas en β – catenina; mutaciones con pérdida de función de *APC*, *Axina-1* o *Axina-2*; metilación en FRPs, y polimorfismos y actividad alterada en *TCF7L2* o *TCF4* (Clevers and Nusse (2012), Bass et al. (2011), Logan and Nusse (2004), Reya and Clevers (2005)). Además, se ha descrito una interacción entre la ruta de señalización Wnt/ β – catenina y los NR (ver apartado “1.3.2.5”), de tal manera que la β – catenina acaba por activar un elevado número de estos y puede dar lugar a alteraciones en la proliferación celular y la tumorigénesis (D’Errico and Moschetta (2008), Mulholland et al. (2005)). También se ha encontrado una relación directa entre alteraciones en *APC* y la sobreactivación de la síntesis de poliaminas en el CCR y que estaría relacionada con la proliferación descontrolada de las células tumorales (Russell and Snyder (1968), Giardiello et al. (1997)).

⁵La regulación de *GSK3* puede llevarse a cabo por diferentes mecanismos, desde su fosforilación a través de la activación de la ruta *PI3K/Akt* o por quinasas como *PKA* o *PKC*, hasta por la formación de complejos proteicos como en el caso de la formación del complejo proteico con *DVL/LRP5/6* tras la activación de la ruta de señalización de Wnt.

1.3.2.4. Activación de Ras/MAPK

En muchos tipos de cáncer la elevada señalización proliferativa puede deberse a una activación excesiva de los receptores de factores de crecimiento. Por un lado, esta situación puede deberse a una expresión constitutiva de estos receptores en las células tumorales, por otro lado, pueden tener ciertas mutaciones que les permitan estar activados independientemente de su unión a su agonista. Estos receptores son TKR, que se caracterizan por tener actividad enzimática tirosin quinasa en su dominio citoplasmático. En condiciones normales, tras la unión de su ligando estos receptores dimerizan y se transfosforilan, de modo que se reclutan una serie de proteínas para transducir la señal, como puede ser a través de la ruta de señalización Ras/MAPK que tiene como resultado el incremento en el crecimiento y la proliferación celular. En primer lugar, el dímero activado recluta a SHC el cual, a su vez, recluta a GRB2. A continuación, a GRB2 se le une un factor intercambiador de guanina como Sos que, a su vez, activa al complejo Ras -GTPasa unida a la MP- intercambiando su GDP por un GTP (la *proteína activadora de GTPasa* (*GAP*, del inglés *GTPase activating protein*) puede inactivar el complejo anterior mediante la hidrólisis del GTP anteriormente intercambiado) (Poturnajova et al. (2021)).

Una vez activado Ras, este dispara la señalización por la ruta de las MAPK. Para ello, Ras recluta y activa a la MAPKKK o Raf, la activa por fosforilación y sufre una serie de cambios conformacionales que incrementan su actividad quinasa. Así, Raf es capaz de fosforilar y activar a la MAPKK o MEK que, a su vez, fosforila y activa a la MAPK o ERK. Ahora, ERK puede actuar sobre sustratos tanto citoplasmáticos como nucleares. En concreto, en el núcleo:

- Fosforila y activa a ELK1 que, a su vez, activará la transcripción de *FOS*.
- Fosforila directamente a Fos y Jun, permitiendo que ahora pueda darse la unión entre ambas proteínas fosforiladas y dar lugar al factor AP1 el cual activa Myc.
- También puede activar directamente a Myc y este, a su vez, activar a CycD cuyo fin es entrar en el ciclo celular.

En las células tumorales pueden presentarse diferentes mutaciones en esta ruta de señalización, entre las que destacan:

- Mutación del dominio de unión a GTP de Ras, reduce su sensibilidad por GAP, por lo que se favorece la unión a GTP y su activación constitutiva.
- Mutación de Sos, conlleva una sobreactivación Ras.
- Mutación de GAP que impide la inactivación de Ras.

Por lo tanto, está claro que las células tumorales que presenten alteraciones a nivel de esta ruta de señalización ya sean activadoras de los TKR, de Ras o en las MAPK, pueden resultar en el aumento del crecimiento, de la proliferación y de la supervivencia celular (Poturnajova et al. (2021)).

1.3.2.5. Receptores nucleares (NR)

Esta superfamilia es uno de los conjuntos de factores de transcripción más grandes en los animales, codificados por 52 genes que poseen una estructura proteica similar. Estos, que pueden localizarse bien en el núcleo o bien en el citosol, actúan como factores de transcripción activados por la unión de un ligando. Son capaces de regular la expresión de diferentes genes e integrar señales provenientes de diferentes procesos metabólicos y fisiológicos: crecimiento, desarrollo o diferenciación celular, mantenimiento de la homeostasis e, incluso, actuar como sensores metabólicos al unirse a ciertos intermediarios metabólicos. Sus ligandos incluyen una amplia gama de compuestos lipofílicos como hormonas, retinoides, vitamina D, ácidos grasos o lípidos provenientes de la propia dieta (Chawla et al. (2001), McKenna and O'Malley (2001)).

En cuanto a su papel en el cáncer, se ha descrito que algunos de estos receptores pueden regular diferentes rutas de señalización que dependen de diferentes factores de transcripción sumamente implicados en el cáncer, como son AP1, p53 o NF- κ B (Kushner et al. (2000), Berger, Qian, and Chen (2013), Sas et al. (2012), Xia et al. (2018)) e, incluso, pueden ser el objetivo de otras rutas de señalización, como su fosforilación por quinasas en respuesta a factores de crecimiento. Además, existe una gran evidencia de la interacción entre la ruta de señalización Wnt/ β -catenina y los NR por la que β -catenina que, al traslocarse al núcleo, activa un elevado número de estos receptores y puede dar lugar a alteraciones en la proliferación celular y la tumorigénesis (D'Errico and Moschetta (2008), Mulholland et al. (2005)). En particular, en el CCR se ha observado la participación de varios NR, tales como LRH-1, ER, PPAR, VDR, TR, RXR, RAR y AR; sin embargo, mientras que una parte de ellos actúan como represores tumorales y su expresión disminuye con la progresión del tumor de colon -RXR, RAR, VDR, ER β o TR α 1-, otros favorecen su desarrollo y su expresión se incrementa -LRH-1, PPAR β , RXR α -P, PPAR γ - (Figura 1.12)(D'Errico and Moschetta (2008)):

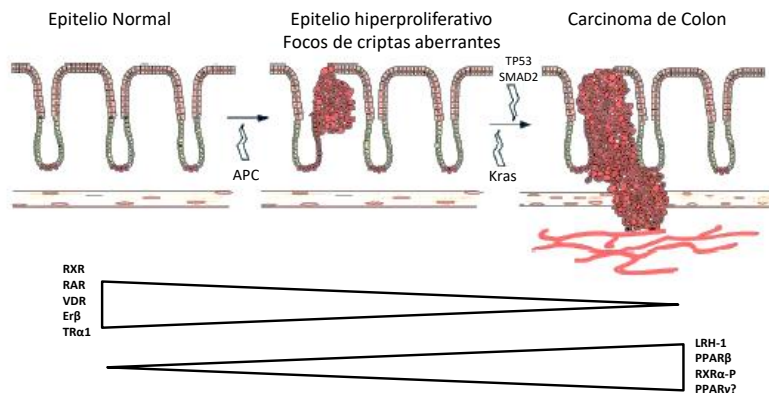


Figura 1.12: Expresión de diferentes receptores nucleares a lo largo de la progresión del CCR. A lo largo del proceso tumoral del CCR, puede observarse cómo parte de los NR disminuye (RXR, RAR, VDR, ER β y TR α 1) mientras que otra parte aumenta (LRH-1, PPAR β , PPAR γ y RXR α -P). Abreviaturas: RXR/RAR, receptor de ácido retinoico; VDR, receptor de vitamina D; ER β , receptor de estrógenos β ; TR α 1, receptor de hormona tiroidea α 1; LRH-1, receptor de homología del hígado 1; PPAR β /PPAR γ , receptor activados por proliferadores peroxisomales β/γ ; RXR α -P, receptor de ácido retinoico α -P (modificada de D'Errico and Moschetta (2008)).

1.3.3. Metástasis de CCR

Alrededor de un 25 % de los pacientes con CCR desarrolla metástasis y suele hacerlo con un patrón secuencial: primero en el hígado y, posteriormente, en el pulmón. No obstante, también existen otros sitios menos frecuentes (L. Sánchez et al. (2015)). A continuación, se van a exponer brevemente algunas de las características de las metástasis en estas dos localizaciones

- *Metástasis de hígado*: el hígado es uno de los órganos más susceptibles para albergar una metástasis tumoral debido a que la mayoría de la circulación sanguínea proveniente del sistema gastrointestinal se dirige a través del sistema venoso portal. La extravasación de las células tumorales es más fácil en el hígado que en los pulmones o en el cerebro ya que los capilares hepáticos están fenestrados y carecen de membrana basal. Sin embargo, cuando logran extravasar, las células tumorales entran en contacto con una gran cantidad de células del sistema inmune que controlan la invasión del tejido. Por ejemplo, las *NK residentes* pueden causar la apoptosis de las células tumorales mediante la participación de TRAIL.

Sin embargo, las células tumorales pueden alterar el efecto que el sistema inmune tiene sobre ellas mediante el reclutamiento de células derivadas de la médula ósea que ayudan en su supervivencia. Para ello, las células tumorales liberan exosomas que presentan el factor inhibitorio de la migración de macrófagos* (*MIF*, del inglés *macrophage inhibitory migration*) en su superficie. En respuesta, las células de *Kupffer* (*macrófagos residentes en el hígado*) secretan $TGF\beta$ que activa las células *estrelladas hepáticas*. Las células estrelladas sobreexpresan y secretan la glicoproteína *Fibronectina* la cual se adhiere a la matriz extracelular y favorece el reclutamiento de células inmunes de origen mieloide que circulan por los vasos sanguíneos.

Además, las células tumorales secretan la quimioquina CCL2 e IL6 las cuales contribuyen a atraer y promover la adhesión de las células mieloides al sitio metastásico. El reclutamiento y la adhesión de las células mieloides protege el tumor y promueve su supervivencia. En paralelo, las células tumorales expresan *glicotransferasas*. Estas alteran el patrón de glicosilación de algunas proteínas de membrana de las células tumorales tales como el *ligando de Galectina 3*, cuyo receptor está en la membrana de las células mieloides, *Galectina 3*, tal que esta alteración incrementa su afinidad por su receptor. Una vez se ha establecido la invasión, las células tumorales y hepáticas interactúan directamente a través de la interacción entre dos proteínas de membrana que se expresa en ambas, *Claudina-2*. Esta interacción mejora el crecimiento y proliferación de las células tumorales.

Finalmente, se han detectado cambios metabólicos en las células tumorales que residen en el hígado. Una menor presencia de *miR-551a* y *miR-483* en estas células tumorales incrementa la producción y secreción de la creatin quinasa B al medio extracelular. Esta quinasa usa la unión con la creatina hepática para generar fosfocreatina, un metabolito rico en ATP que puede ser absorbida y usada por las células tumorales, por lo que se produce una adaptación metabólica hacia el consumo de creatina y, por ende, se favorece el crecimiento tumoral.

- *Metástasis de pulmón*: en los capilares de los pulmones, las CTC pueden entrar en contacto con diferentes factores que favorecen su extravasación tales como SPARC, ANGPTL4, COX2 o MMP2. Además, ellas mismas también pueden presentar diferentes factores que contribuyen a su extravasación en los capilares pulmonares. En cuanto a las proteínas de la matriz extracelular Periostina y Tenascina, estas también favorecen la supervivencia de las células tumorales invasivas mediante su unión a las integrinas. Ambas proteínas también modulan positivamente la ruta de señalización canónica de Wnt en las células tumorales. Además, la Tensacina también puede promover la ruta de señalización de Notch.

Otros protagonistas importantes son los macrófagos reclutados por las células tumorales también pueden contribuir a su supervivencia a través de la unión de las integrinas $\alpha4\beta1$ de la membrana de los macrófagos al VCAM1 de la membrana de las células tumorales. De esta manera, los macrófagos activan la ruta de supervivencia de Akt en las propias células tumorales. Cabe destacar que a pesar de que las células tumorales reciben señales de latencia, en concreto de las proteínas *morfogénicas óseas* (BMP, del inglés *bone morphogenic protein*) secretadas por las células pulmonares en el nuevo nicho, las células tumorales que sobreexpresan Coco, que son proteínas que secuestran BMP, pueden escapar a estas señales. Por lo tanto, la sobreexpresión de Coco en las células tumorales está asociada con un incremento en el crecimiento metastásico.

1.3.4. Prevención y tratamientos

La identificación temprana de los individuos que presentan una predisposición genética al CCR puede favorecer la toma de medidas preventivas. Entre estas destaca la prueba de heces en sangre o la colonoscopia las cuales han supuesto una disminución considerable en la mortalidad por CCR (Vasen, Tomlinson, and Castells (2015)). Además, debido a que no todos los pólipos, a pesar de ser clasificados dentro de un mismo tipo de síndrome, presentan la misma patogenicidad, es imperativo identificar las alteraciones genéticas para su adecuado tratamiento.

El tratamiento del CCR dependerá del tipo, estado, progresión, y severidad del CCR. El principal tratamiento para el CCR primario en estadios del 0 al II es la cirugía, pues la resección permite estadificar adecuadamente el cáncer, previene posibles obstrucciones del colon así como el sangrado del tumor. No obstante, si el CCR es invasivo y/o ha metastatizado (estadios III y algún II), se utiliza, junto con la cirugía, quimioterapia como terapia adyuvante, quimioterapia con terapia multi-farmacológica (estadio II), y radioterapia (CCR recurrentes o en estadios avanzados de la enfermedad).

Los tratamientos quimioterápicos más habituales son FOLFOX y CapeOx (capecitabina y oxaluplatino), que atacan a las células cancerígenas en división. Después de la cirugía, en el caso de que el CCR haya metastatizado, el tratamiento más habitual es la quimioterapia usando 5-fluorouracilo. Además de la quimioterapia, muchos pacientes son tratados con agentes dirigidos. En este sentido, una de las estrategias más habituales es el uso de anticuerpos monoclonales como

el Cetuximab, el Panitumumab o el Eributx, que se dirigen al *receptor del factor de crecimiento epidérmico (EGFR, del inglés epidermal growth factor receptor)*. Además, estos anticuerpos monoclonales en combinación con otros agentes quimioterápicos son también usados para tratar la metástasis de colon. Otra opción son los anticuerpos monoclonales recombinantes humanizados como el Bevacizumab, que se une al *factor de crecimiento del endotelio vascular (VEGF, del inglés vascular endothelial growth factor)*. Cabe destacar que también se están desarrollando tratamientos más novedosos como el uso de vacunas para tratar el CCR. Obviamente, todos los medicamentos mencionados aquí no son más que una pequeña parte de todos los tratamientos existentes (Colomer (2005), Mundade et al. (2014)).

A título informativo, se presentan algunas estrategias de tratamiento en función del tipo de CCR. Por ejemplo, para pacientes con síndrome APC, que hayan desarrollado un número de pólipos elevado en el recto o que presenten un genotipo severo se realiza un tratamiento quirúrgico conocido como *colectomía total*⁶ o *proctocolectomía*⁷. No obstante, si los pacientes con este síndrome presentan un fenotipo atenuado con pocos pólipos (aquellos que tienen una mutación al final del gen del APC o en exón 9) se opta por la *polipectomía endoscópica*⁸. Sin embargo, para pacientes que tienen el síndrome de Lynch, puesto que son pacientes jóvenes con riesgo de desarrollar un segundo CCR el tratamiento quirúrgico es más complejo de plantear. Por ejemplo, aquellos que presentan mutaciones en MSH2, MLH1 o MSH6 se opta por una *histerectomía profiláctica y ooforectomía bilateral*⁹, en lugar de una *colectomía subtotal*¹⁰ (Vasen, Tomlinson, and Castells (2015)).

Actualmente, se está estudiando el efecto del DFMO en el cáncer, pues es un inhibidor de la ruta de las poliaminas la cual, en el CCR, está sobreactivada como consecuencia de l alteraciones en la ruta del APC (Russell and Snyder (1968), Giardiello et al. (1997)).

⁶La colectomía total consiste en la extirpación del intestino grueso a partir de la parte más baja del intestino delgado hasta el recto (todo el colon). A diferencia de esta, la colectomía subtotal es la extirpación de parte del colon.

⁷La proctocolectomía total es la extirpación del colon, el recto y el ano.

⁸La polipectomía endoscópica consiste en la extirpación de los pólipos por vía endoscópica.

⁹La ooforectomía bilateral consiste en la extirpación de los dos ovarios. En contraposición, la histerectomía profiláctica es la extirpación de los ovarios cuando estos son todavía normales pero la cirugía se lleva a cabo de forma preventiva.

¹⁰La colectomía total consiste en la extirpación del intestino grueso a partir de la parte más baja del intestino delgado hasta el recto (todo el colon). A diferencia de esta, la colectomía subtotal es la extirpación de parte del colon.

1.3.5. Inhibición de la síntesis de poliaminas

Las poliaminas son moléculas alifáticas nitrogenadas de bajo peso molecular con grupos amino distribuidos que se distribuyen de forma regular a lo largo de su estructura. Antiguamente, se consideraba que las poliaminas eran aquellas moléculas formadas a partir de la descarboxilación de la ornitina que, en las células de mamíferos son la putrescina, espermidina y espermina. Sin embargo, en la actualidad también se han encontrado otras vías de formación de poliaminas como, por ejemplo, a partir de la arginina (Moinard, Cynober, and Bandt (2005)). En particular, se ha identificado una poliamina en animales que deriva de la arginina, la agmatina, la cual sólo se había identificado en bacterias, plantas y algunos organismos (Lortie et al. (1996)). Otras oligoaminas como la histamina y la cadaverina también aparecen de forma natural en las células, pero no se suelen incluir dentro de las poliaminas. La reserva más importante se encuentra en el citosol y en el núcleo (Moinard, Cynober, and Bandt (2005)).

Estas moléculas, a pH fisiológico, son policatiónicas, por lo que interaccionan con macromoléculas polianiónicas como el ADN, el ARN, proteínas y fosfolípidos ácidos, de tal manera que sólo un 10% de estas se encuentran en su forma libre (Pirinen et al. (2007), Q. Yuan et al. (2001)). En consecuencia, estas intervienen en numerosos procesos celulares como la apoptosis, la división celular y diferenciación, proliferación celular, síntesis de ADN y proteínas, estrés oxidativo, expresión de genes, homeostasis, mantenimiento de la cromatina, regulación de canales iónicos, reprogramación celular, autofagia, eliminación de ROS, y apoptosis (figura 1.13) (Moinard, Cynober, and Bandt (2005), Kaerberlein (2009), Minois, Carmona-Gutierrez, and Madeo (2011), Minois (2014)). En particular, en las células epiteliales de colon las poliaminas tienen un papel fundamental en la restitución epitelial, proceso en el que tiene lugar una activación transitoria de la migración y de la proliferación celular con el fin de reparar el tejido (Gutiérrez et al. (2019)).

En humanos, las poliaminas, que circulan ligadas a estructuras de la membrana celular como los fosfolípidos, provienen de dos fuentes: síntesis de *novo* y exógeno. Por un lado, puede ser sintetizado a partir de la biosíntesis *de novo* e interconversiones entre ellas o de secreciones digestivas (especialmente en el lumen intestinal), que es la fuente predominante en mamíferos. Por otro lado, de origen exógeno mediante la absorción activa a través de la dieta o a través de la producción de poliaminas por parte de las bacterias intestinales. Estas pueden entrar a la célula a través de diferentes transportadores: putrescina a través del transportador de diaminas dependiente de Na^+ el cual también puede sacar esta poliamina al medio extracelular, y las poliaminas espermidina y espermina a través del transportador de poliaminas dependiente de ATP (Minois, Carmona-Gutierrez, and Madeo (2011)).

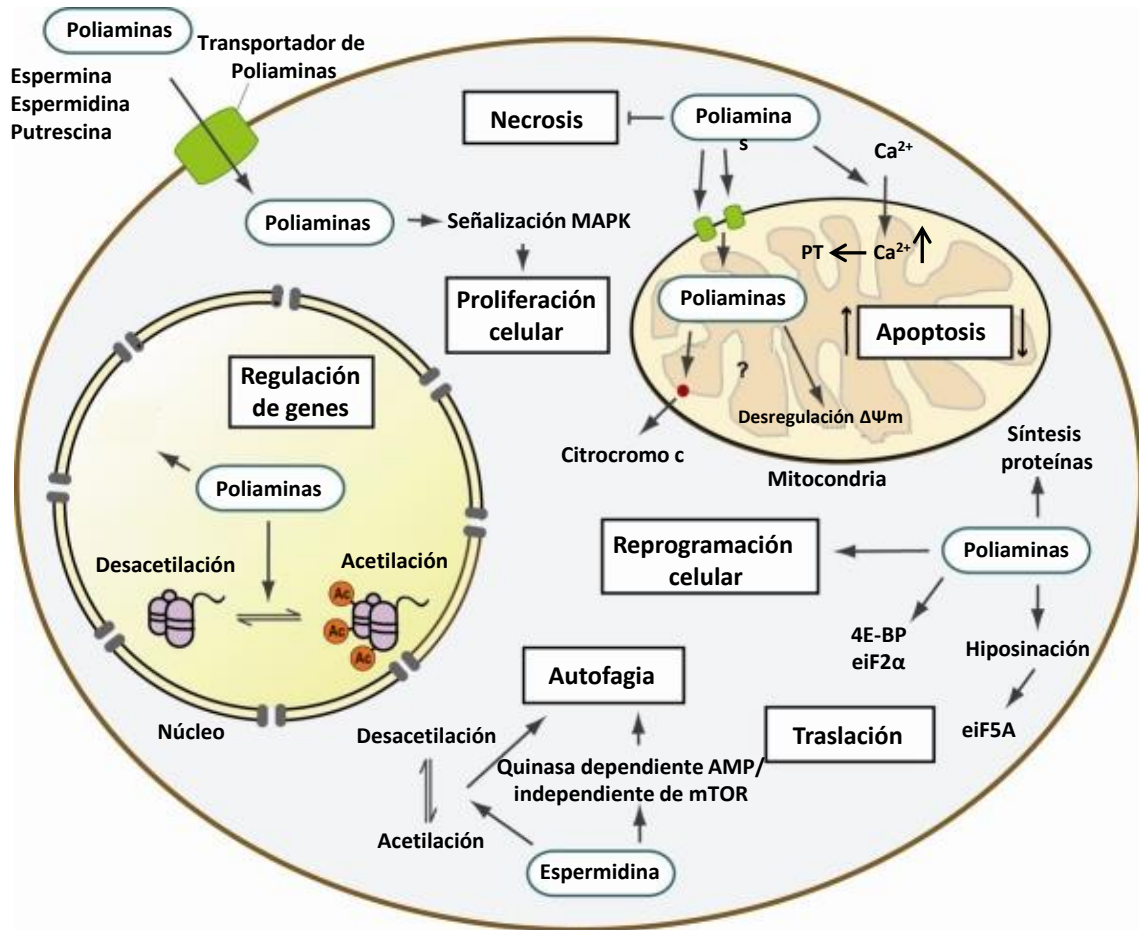


Figura 1.13: Mecanismos de acción de las poliaminas. En este esquema se muestran los mecanismos de acción de las poliaminas y los procesos celulares en los que intervienen a lo largo de diferentes localizaciones celulares: citoplasma, núcleo y mitocondria. Estas pueden entrar a la célula a través de transportadores de diferentes transportadores: putrescina a través del transportador de diaminas dependiente de Na^+ el cual también puede sacar esta poliamina al medio extracelular, y las poliaminas espermidina y espermina a través del transportador de poliaminas dependiente de ATP. Algunos de estos procesos son la regulación de la muerte, proliferación y reprogramación celular, la autofagia, la regulación de la expresión de genes y la síntesis de proteínas. Abreviaturas: PT, transición de permeabilidad, $\Delta\psi$: potencial de membrana mitocondrial. (Carmona-Gutierrez, and Madeo (2011)).

En cuanto a su síntesis, las poliaminas se obtienen a partir de aminoácidos como la L-arginina, L-ornitina y L-metionina. El primer paso es la producción de L-ornitina a partir de L-arginina por acción de la enzima mitocondrial arginasa (figura 1.14). A continuación, la L-ornitina es descarboxilada por acción de la ornitina descarboxilasa (ODC) la cual origina putrescina. Esta enzima tiene una elevada inducibilidad, como por ejemplo a través de estímulos promotores del crecimiento, por lo que su papel en la proliferación celular puede ser clave. En cuanto a su regulación, la tasa de recambio de este enzima es de 10 minutos y, además, puede ser degradado por antizimo. A partir de esta poliamina es posible sintetizar otras dos poliaminas, espermidina y espermina, mediante la acción de las enzimas espermidina sintasa y espermina sintasa, respectivamente. Estas enzimas tienen una baja inducibilidad, a diferencia de ODC (Wallace et al. (2009)).

Puesto que en la síntesis de espermidina y espermina se añaden uno o dos grupos aminopropilo (de forma consecutiva), respectivamente, ha de existir un donador de dichos grupos (figura 1.14). En concreto, este es la S-adenosilmetionina descarboxilada la cual proviene de la descarboxilación de la S-adenosilmetionina por la acción de la S-adenosilmetionina descarboxilasa (SMADC), la cual tiene una elevada inducibilidad, al igual que ODC. Así, la S-adenosilmetionina descarboxilada, tras ceder el grupo aminopropilo, da lugar al producto de reacción 5'-metiltioadenosina (MTA) (Wallace et al. (2009)).

Cabe destacar que la síntesis de poliaminas no es unidireccional (figura 1.14), sino que la espermidina y la espermina también pueden reconvertirse en putrescina por acción combinada del enzima espermidina/espermina N-acetil-transferasa (SSAT), que transfiere un grupo N-acetil a la espermidina o a la espermina, y de la N-acetilpoliamina oxidasa (APAO). No obstante, también se han encontrado otras enzimas, en concreto la espermina oxidasa, capaz de convertir la espermina en espermidina en un único paso (Wallace et al. (2009), Minois, Carmona-Gutierrez, and Madeo (2011)).

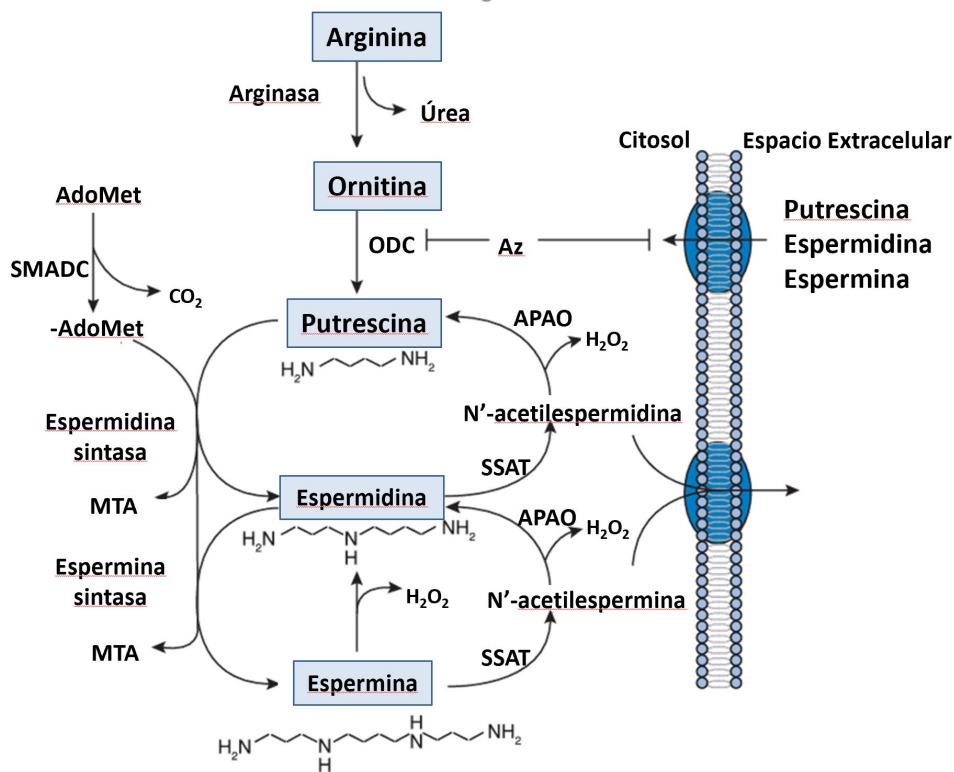


Figura 1.14: Síntesis de poliaminas. A partir de diferentes aminoácidos, como por ejemplo la arginina, ornitina y metionina, tiene lugar la biosíntesis de poliaminas. Por ejemplo, a nivel de absorción la L-arginina, producto ampliamente consumido por sus propiedades vasodilatadoras, tiene una baja tasa de absorción dada su conversión a L-ornitina por acción de la enzima mitocondrial arginasa de los enterocitos. Posteriormente, la L-ornitina sufre una descarboxilación por acción del enzima ODC para dar lugar a la poliamina putrescina. Abreviaturas: AdoMet, S-adenosilmetionina; -AdoMet, S-adenosilmetionina descarboxilada; Az, antizimo (modificado de Wallace et al. (2009)).

Russell and Snyder (1968) describieron que existe una fuerte relación entre los niveles de poliaminas y la proliferación descontrolada. Ambos observaron cómo en células y tejidos tumorales el contenido de poliaminas es muy elevado comparado con las células y tejido sano. El metabolismo de las poliaminas es un componente del mecanismo de carcinogénesis. De hecho, esta ruta de síntesis suele ser un objetivo de muchos oncogenes (Shantz and Levin (2007)). En particular, se ha observado un incremento de ODC el cual se relaciona con un crecimiento mantenido de las células normales que puede dar lugar a neoplasias malignas (Russell and Snyder (1968), Casero and Marton (2007)). Además, también se ha observado un aumento en los niveles de poliaminas en diferentes células cancerígenas como, por ejemplo, en pacientes con cáncer de mama, piel o de colon -poliposis adenomatosa familiar- (Giardiello et al. (1997), Elmets and Athar (2010), Nowotarski, Woster, and Casero (2013)).

Debido a la clara implicación de las poliaminas en el proceso tumoral, estas y sus derivados se emplean como biomarcadores de cáncer y, además, se considera que la ruta de síntesis de estas es una potencial diana terapéutica (Gutiérrez et al. (2019)). A tales efectos, en 1978 se sintetizó el difluorometilornitina (DFMO), un inhibidor suicida del OCD. Es más, se ha visto que los inhibidores de la síntesis de poliaminas son capaces de suprimir la carcinogénesis intestinal y de colon en ratones y humanos (Gerner, Ignatenko, and Besselsen (2003), Ignatenko et al. (2004)). Se ha observado que el principal efecto del DFMO fue una reducción del número de adenomas de alto riesgo (Yerushalmi et al. (2006)). Los inhibidores de poliaminas también se han comenzado a probar en combinación con antiinflamatorios no esteroideos (AINE), compuestos que podrían inhibir el transporte de poliaminas, puesto que estos últimos también han demostrado un efecto beneficioso sobre la carcinogénesis de colon. Además, las combinaciones de estos fármacos podrían requerir de unas concentraciones más bajas de las que se habrían requerido si se hubieran utilizado de forma independiente; es decir, que la combinación de ambas muestra una interacción o efecto sinérgico. En este sentido, varios estudios experimentales han demostrado que el DFMO actúa de forma aditiva e, incluso, sinérgica, con diferentes AINE como, por ejemplo, el agente selectivo de COX1 aspirina (H. Li et al. (1999)), el agente selectivo de COX2 celecoxib e inhibidores no selectivos de COX1 y COX2 como el piroxicam y el sulindac (Lawson et al. (2000)). También existe evidencia clínica de los efectos beneficiosos de muchos AINE para combatir la formación de pólipos en el colon: sulindac en individuos de alto riesgo, como aquellos con PAF, aspirina y adenomas de colon (Giardiello et al. (1993), Sandler et al. (2003)).

Meyskens et al. (2008) demostró la gran eficacia que tiene la combinación del DFMO con el sulindac en el tratamiento del adenoma colorrectal recurrente, para lo que realizó un ensayo en fase III aleatorizado, doble ciego cuyos niveles de tratamiento fueron tratados con DFMO y control (placebo). En particular, se observó una reducción del 70 % del total de pólipos y del 91.5 % tanto en adenomas avanzados como en pacientes con múltiples adenomas recurrentes tras 3 años de tratamiento. En cuanto a la toxicidad del tratamiento, se dió en casos raros; no obstante, el riesgo de eventos cardiovasculares adversos asociados al tratamiento aumentó en aquellos pacientes que presentaban alto riesgo en situación basal, pero no para aquellos con un riesgo previo bajo (Zell et al. (2009)). En otros estudios de mayor tamaño se observaron unos resultados similares. Así, este

ensayo clínico y similares evidenciaron el efecto beneficioso que tiene atacar tanto a la síntesis de poliaminas como a la inflamación y/o el transporte de poliaminas a través de esta combinación de fármacos, lo que supondría una estrategia adecuada para reducir los factores de riesgo previos a la condición cancerígena como son los adenomas de colon, estrechamente relacionados con el desarrollo de CCR en humanos. A nivel molecular, diferentes estudios también han demostrado el papel del DFMO y del sulindac en la reversión parcial del fenotipo tumoral de CCR (Gutiérrez et al. (2019)). Además, estos resultados podrían ser extrapolables a otro tipos de cáncer como el de mama, pulmón y próstata.

Capítulo 2

Homeostasis del Calcio Intracelular

En este capítulo se va a presentar la relevancia biológica del Ca^{2+} intracelular y los complejos sistemas de transporte de este catión que existen a nivel celular.

2.1. El calcio intracelular

A pesar de que cerca del 99 % del Ca^{2+} corporal se encuentra en el hueso, donde realiza una función estructural, el Ca^{2+} restante se distribuye entre el interior celular, 1 %, y el medio extracelular, 0.1. No obstante, este 1.1 % tiene una gran relevancia a nivel fisiológico debido a su gran versatilidad, pues puede actuar tanto como segundo mensajero como como cofactor de una gran panoplia de procesos fisiológicos: desde la expresión de genes, la contracción-relajación muscular, la plasticidad sináptica, la secreción de hormonas, la proliferación y el ciclo celular y la movilidad celular hasta la *diffpoptosis* -concepto acuñado por Whitfield para definir la combinación entre la diferenciación y la apoptosis- (Sharma et al. (2017), Whitfield (1997), Whitfield and Chakravarthy (2001), M. J. Berridge, Lipp, and Bootman (2000), Prakriya and Lewis (2015)). Estas respuestas celulares mediadas por Ca^{2+} pueden llevarse a cabo en cuestión de microsegundos, como sucede en la exocitosis de neurotransmisores, e incluso demorarse minutos e incluso horas, como en la expresión de genes o la proliferación (M. J. Berridge (2012), M. J. Berridge, Lipp, and Bootman (2000)).

El papel fisiológico del Ca^{2+} y su actividad no se limita a una simple disminución o aumento puntual en la $[\text{Ca}^{2+}]_i$, sino que va más allá, pues en la gran mayoría de las células esta concentración muestra un patrón de comportamiento espacio-temporal complejo de carácter tanto puntual como oscilatorio y ondulatorio, que va desde picos estocásticos hasta oscilaciones regulares, ondas periódicas y ondas espirales. Pero este comportamiento tiene su razón de ser, pues de él dependen muchas de las respuestas celulares desencadenadas por este ion. Por ejemplo, la expresión de genes o diferentes mecanismos de señalización pueden depender de la amplitud, frecuencia y duración de las oscilaciones de $[\text{Ca}^{2+}]_i$ en una determinada localización celular (Dupont et al. (2016)).

En cuanto a la distribución del Ca^{2+} en el contexto celular, la concentración en el medio extracelular es alrededor de 1 mM, mientras que en el citosol es cuatro órdenes de magnitud inferior, en torno al $0.1 \mu\text{M}$; además, orgánulos como el retículo endoplasmático (RE) o los lisosomas tienen también una elevada concentración de Ca^{2+} en comparación con el citosol, por lo que, en cierto modo, también actúan como depósitos de Ca^{2+} intracelular. Por lo tanto, es evidente que ha de existir una serie de mecanismos sumamente regulados que permitan mantener esta homeostasis, que van desde complejos sistemas de transporte entre compartimentos celulares o entre el medio intra y extracelular en los que participan *Bombas de Calcio (CP)* y *Canales de Calcio Operados por Receptor (ROCC)* o por *Voltaje (VOCC)*, hasta proteínas de unión a Ca^{2+} capaces de regular los niveles intracelulares del mismo e incluso la participación directa de orgánulos como el RE y la mitocondria (figura 2.1) (Rajagopal and Ponnusamy (2017), Carafoli (2003), Dupont et al. (2016), M. J. Berridge, Lipp, and Bootman (2000), M. J. Berridge (2012)). Además, en el interior celular existen otros sistemas que contribuyen a la regulación de la $[\text{Ca}^{2+}]_i$, presentes tanto en el citosol como en la práctica totalidad de los orgánulos, y que son capaces de unir Ca^{2+} . En concreto, esta regulación de la $[\text{Ca}^{2+}]$ en el interior celular se debe a que una gran parte de este catión permanece unido a alguno de estos sistemas, como por ejemplo las proteínas *Calmodulinas* o *Calsequestrina*, capaces de actuar como amortiguadores de tal manera que cuando la $[\text{Ca}^{2+}]_i$ disminuye estas lo liberan. En consecuencia, cuando se habla de la $[\text{Ca}^{2+}]$, ya sea citosólica, mitocondrial... se hace mención al Ca^{2+} libre, ya que ese es el que es capaz de interactuar con otros sistemas celulares y producir diferentes efectos.

Evidentemente, el mantenimiento de estos gradientes de concentración, generados por las bombas (tanto de la MP como de los depósitos intracelulares), requiere de una gran cantidad de energía; sin embargo, con la simple apertura/cierre de los canales pueden darse unos cambios sumamente rápidos en las concentraciones citosólicas de Ca^{2+} ($[\text{Ca}^{2+}]_{\text{cit}}$). No obstante, el gasto excesivo de energía no es en balde, pues el mantenimiento a la larga de una baja $[\text{Ca}^{2+}]_{\text{cit}}$ es imprescindible para el correcto funcionamiento celular, pues en caso contrario pueden desregularse procesos mediados por moléculas efectoras sensibles a Ca^{2+} , tales como la proliferación, diferenciación, secreción y transcripción de genes, e incluso promover la apoptosis. Sin embargo, la regulación de la $[\text{Ca}^{2+}]_{\text{cit}}$ no es homogénea en toda la célula, pues el hecho de que este sea un segundo mensajero requiere que su regulación sea muy precisa a nivel no solo temporal, sino también espacial. Por ello, existen mecanismos que permiten una regulación local de la $[\text{Ca}^{2+}]_{\text{cit}}$, lo que genera microambientes en los que la $[\text{Ca}^{2+}]_{\text{cit}}$ permanece elevada mientras que en otras zonas está baja: por un lado, existen mecanismos mediados por canales (de la MP o de depósitos intracelulares) que permitan una liberación controlada de Ca^{2+} a nivel local para que lleve a cabo su función señalizadora y, por otro lado, otros mecanismos mediados por bombas e intercambiadores de Ca^{2+} que eliminan rápidamente este incremento de Ca^{2+} para poder alcanzar nuevamente los niveles basales de $[\text{Ca}^{2+}]_{\text{cit}}$ (Dupont et al. (2016), M. J. Berridge, Lipp, and Bootman (2000), M. J. Berridge (2012)).

Obviamente, no en todos los tipos celulares existen ni las mismas maquinarias de regulación de Ca^{2+} ni las mismas rutas de señalización relacionadas. No obstante, ante errores en la señal de Ca^{2+} debidas a defectos en alguno de los mecanismos que la regulan, otros mecanismos pueden compensar estas señales defectuosas para restaurar la normalidad, por lo que se produce una remodelación de la homeostasis del Ca^{2+} intracelular. Para ello, han de existir mecanismos que evalúen la idoneidad de la señal de Ca^{2+} como la *CaN (Calcineurina)*, proteína fosfatasa sensible a Ca^{2+} , capaz de activar factores de transcripción como el *NFAT (Factor Nuclear de Células T Activadas)* (Dupont et al. (2016), M. J. Berridge, Lipp, and Bootman (2000), M. J. Berridge (2012)).

Por lo tanto, está claro que el control de una adecuada $[\text{Ca}^{2+}]_{\text{cit}}$, tanto a nivel general como a un nivel espacio-temporal concreto, es de suma importancia para asegurar el correcto funcionamiento de las diferentes rutas y procesos en los que este catión interviene. Tan es así que un remodelado aberrante de esta homeostasis está relacionado con numerosas patologías, desde enfermedades neurodegenerativas como el Alzheimer hasta el cáncer. Por lo tanto, la correcta homeostasis del Ca^{2+} intracelular, los mecanismos compensatorios cuando esta no se mantiene adecuadamente dentro del contexto celular y tisular concreto, así como el mal funcionamiento de toda esta maquinaria, es clave tanto en el desarrollo como en el tratamiento de diferentes patologías (M. J. Berridge, Lipp, and Bootman (2000), M. J. Berridge (2012)).

2.2. Señalización por Ca^{2+} intracelular

Para entender y comprender la red de señalización del Ca^{2+} intracelular generada por los complejos sistemas de transporte celulares de este catión, es muy útil el concepto de compartimentalización propuesto por M. J. Berridge, Lipp, and Bootman (2000) denominado *caja de herramientas de la señalización del Ca^{2+} (Calcium Signalling Toolkit/Toolbox)*. Esta contiene diferentes piezas que se agrupan en función de su contribución a los sistemas de señalización de Ca^{2+} intracelular, de tal manera que estas pueden mezclarse y combinarse para “crear” un amplio abanico de sistemas de señalización de Ca^{2+} intracelular que darían lugar a señales con diferentes características espacio-temporales. En otras palabras, que las células mostrarán diferentes patrones de expresión de cada uno de los componentes de esta caja de herramientas que, obviamente, podrá variar tanto espacial como temporalmente en función de sus necesidades y de las funciones que deba realizar (figura 2.1) (M. J. Berridge (2012)).

Este sistema se puede complicar aun más, pues muchos de estos componentes tienen diferentes isoformas con ligeras diferencias entre sí. Esto permite un sin fin de rutas de señalización de Ca^{2+} entre las que puede haber desde ligeras diferencias e, incluso, mostrar un comportamiento muy diferente. El concepto de una caja de herramientas compartimentalizada es muy útil a la hora de generar modelos predictivos y explicativos que permitan estudiar su compleja homeostasis, pues es posible generar diferentes modelos a partir de incluir o excluir los diferentes componentes de esta caja (Dupont et al. (2016)).

En cuanto a las agrupaciones en las que pueden dividirse los diferentes componentes de la caja de herramientas, existen cuatro grandes unidades funcionales en función de su aportación a esta señalización (M. J. Berridge, Lipp, and Bootman (2000), M. J. Berridge (2012)):

- *Señales movilizadoras de Ca^{2+}* : En respuesta a ciertos estímulos extracelulares capaces de actuar sobre distintos receptores de la superficie celular (*GPCR -receptores acoplados a proteínas G-* y *TKR*) acoplados a diferentes isoformas de *PLC (Fosfolipasas C)*, las células generan diferentes señales movilizadoras de Ca^{2+} que activan diferentes mecanismos tanto de la MP como de los depósitos intracelulares las cuales movilizan a este catión hacia el citosol a partir tanto del presente en el medio extracelular como en los depósitos intracelulares y, en consecuencia, la $[Ca^{2+}]_{cit}$ aumenta (figura 2.1. Entre estas señales movilizadoras de Ca^{2+} se encuentran el *Inositol 1,4,5-trisfosfato (IP3)* -activa los *IP3R (Receptores de IP3)* del RE- o el propio Ca^{2+} citoplasmático que es capaz de activar la liberación del Ca^{2+} de los depósitos a través de los canales *IP3R* o *RyR (Receptores de Rianodina)* del RE mediante un mecanismo conocido como *Liberación de Calcio Inducida por Calcio (CICR, del inglés, Calcium Induced Calcium Release)*.
- *Mecanismos ON*: Estos mecanismos engloban a aquellos que aumentan la $[Ca^{2+}]_{cit}$ a través de las señales movilizadoras de Ca^{2+} que activan la entrada de este catión al citosol, ya sea desde el medio extracelular o de los depósitos intracelulares, a favor de su gradiente electroquímico, por lo que no requieren de energía para ello. Por ejemplo, a partir de un estímulo extracelular capaz de actuar sobre un receptor de la superficie celular se genera la molécula señalizadora *IP3*, la cual difunde por el citosol hasta alcanzar los *IP3R* del RE y activarlos. Tras su activación, estos sufren un cambio conformacional que les permite liberar Ca^{2+} al citoplasma el cual, a su vez, es capaz de potenciar o inhibir esta salida de Ca^{2+} ; por ejemplo, en ausencia de *IP3*, si $[Ca^{2+}]_{cit} \in [100 - 300]$ nM, estimula la liberación de Ca^{2+} , y la inhibe cuando $[Ca^{2+}]_{cit} > 300$ nM.
- *Procesos Sensibles a Ca^{2+}* : Como ya se ha comentado anteriormente, el Ca^{2+} tiene un papel importante en numerosos procesos, pues, tras el aumento de Ca^{2+}_{cit} tras la acción de los mecanismos ON, y dada su versatilidad, es capaz de actuar como segundo mensajero y actuar sobre diferentes procesos sensibles a Ca^{2+} . En concreto, en esta caja de herramientas hay una serie de componentes con sensibilidad por el Ca^{2+} : *proteínas de unión a Ca^{2+}* , que abarca a las proteínas *amortiguadoras* y a los *sensores de Ca^{2+}* , o elementos de otras rutas capaces de ser reclutados por el Ca^{2+} , como el AMPc.
- *Mecanismos OFF*: A diferencia de los mecanismos ON, los mecanismos OFF engloban a aquellos que retiran el Ca^{2+} del citosol para mantener los niveles basales de este y, en caso de una subida poder restablecer dichos niveles. Los protagonistas de estos mecanismos son las bombas y e intercambiadores, los cuales, para transportar al Ca^{2+} en contra de su gradiente electroquímico, requieren de energía la cual puede ser liberada por la hidrólisis del ATP o por el co-transporte de iones como el Na^+ a favor de su gradiente electroquímico.

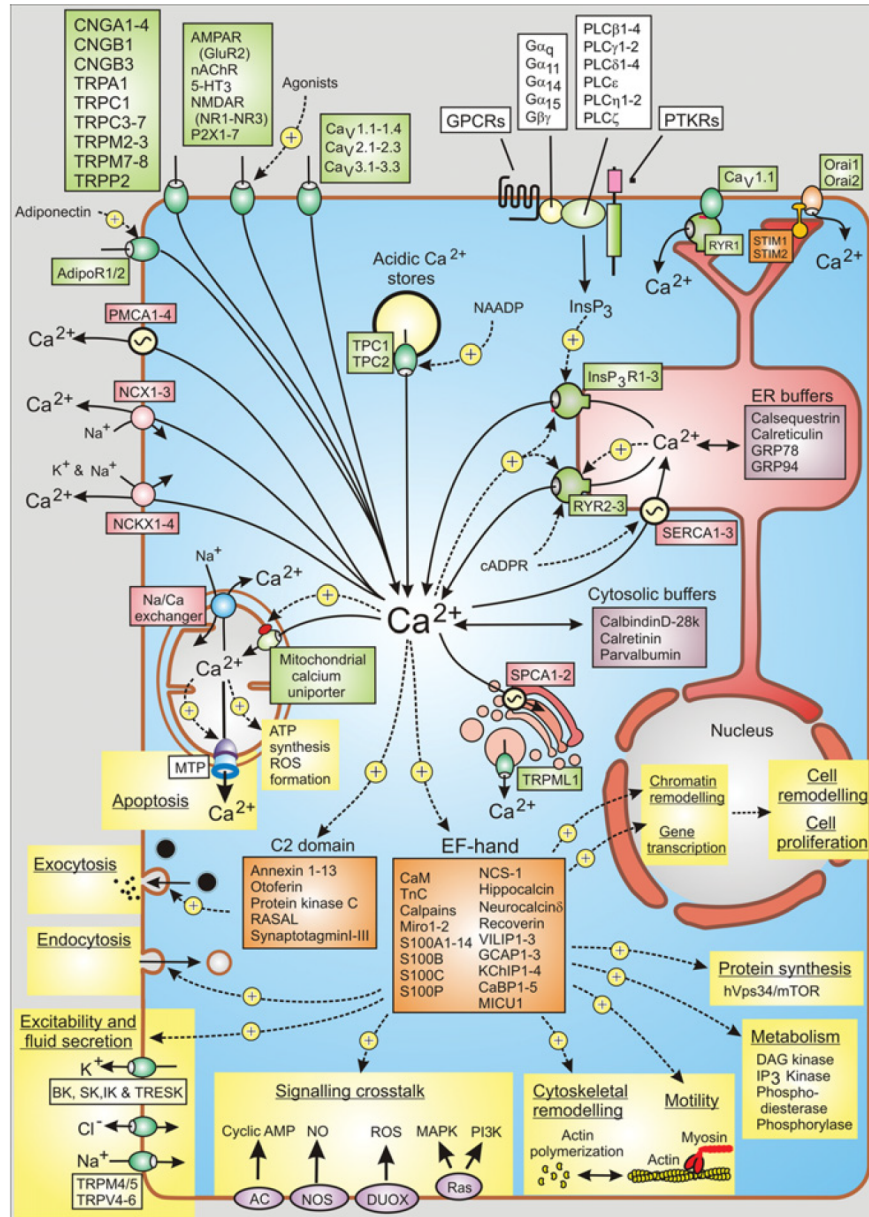


Figura 2.1: Señalización mediada por Ca^{2+} . En esta figura se muestra en complejo entramado de señalización en el que está implicado el Ca^{2+} : las cajas verdes recogen los canales de Ca^{2+} a través de los cuales entra el Ca^{2+} al interior celular; las cajas rojas, las bombas que bombean Ca^{2+} en contra del gradiente -al medio extracelular o al interior del RE-; las cajas naranjas se corresponden con los sensores de Ca^{2+} C2 o Manos-EF que, a su vez, estimulan los procesos celulares señalados mediante cajas amarillas (modificado de Berridge (2012)).

En los siguientes apartados se presentan algunos de los componentes más importantes de la señalización de Ca^{2+} y, para una mejor comprensión, estos se dividen en función tanto de su localización celular como de tipo de mecanismo ON/OFF, es decir, si se localizan en la MP, el RE, la mitocondria y permiten la entrada o expulsan Ca^{2+} del citosol y, además, un último apartado para exponer brevemente a las proteínas de unión a Ca^{2+} .

2.3. La membrana plasmática

La diferencia de concentración de Ca^{2+} entre el medio extracelular y el citosol es muy elevada, de cuatro órdenes de magnitud de diferencia. Por ellos, es de suma importancia la existencia de diferentes sistemas de transporte localizados en esta estructura celular para poder mantener este gradiente electroquímico. En este apartado se van a presentar los diferentes sistemas de transporte de Ca^{2+} que existen en la MP.

2.3.1. Mecanismos *ON* de la MP

El aumento de la $[\text{Ca}^{2+}]_{\text{cit}}$ procedente del medio extracelular está mediado por diferentes canales localizados en la MP. Estos pueden abrirse por diferentes mecanismos. En concreto, se distinguen cuatro tipos según el mecanismo por el cual se activa su apertura:

- Los *canales de Ca^{2+} operados por voltaje (VOCC, del inglés voltage-operated Ca^{2+} channel)* activan su apertura a través de cambios en el potencial de membrana.
- Los *ROCC (canales de Ca^{2+} operados por receptor)* se abren tras la unión de un ligando concreto en el dominio extracelular. No obstante, aquí también pueden considerarse los receptores de la superficie celular GPCR o TKR que, tras la unión a sus correspondientes ligandos, desencadenan la producción de señales movilizadoras de Ca^{2+} , como por ejemplo el *receptor muscarínico de acetilcolina (mAChR)*, un GPCR que activa la producción de IP3R y DAG y, en consecuencia, se activa la liberación del Ca^{2+} de los depósitos.
- Los *canales de Ca^{2+} operados por depósitos (SOCC, del inglés store-operated Ca^{2+} channel)* son aquellos canales cuya activación se debe al vaciamiento del Ca^{2+} almacenado en el RE. Este vaciado es consecuencia de una liberación previa desencadenada por el efecto de las señales movilizadoras de Ca^{2+} , de tal manera que se activa un mecanismo de rellenado del RE con nombre propio, conocido como *entrada capacitativa de Ca^{2+} o entrada de Ca^{2+} operada por depósitos (SOCE)*.
- Los *receptores de potencial transitorio (TRP, del inglés transient receptor potential)* son, por lo general, permeables a Ca^{2+} . En particular, son los TRPC, TRPV, TRPM{1, 2, 3, 6, 7}, TRPA1, TRPP{2, 3, 5} y TRPML{1, 2, 3} los que presentan dicha permeabilidad. Cabe destacar que no tienen un mecanismo único y general que induzca su apertura, pues mientras unos se abren inducidos por el frío, como el TRPM8, otros lo hacen por el aumento del Ca^{2+} .

A continuación, se desarrollan en más detalle cada uno de los tipos de canales anteriores.

2.3.1.1. Canales de calcio operados por voltaje (VOCC)

Por lo general, los VOCC se localizan en las células excitables. Estos, en respuesta a la despolarización de la MP, se abren y, en consecuencia, el Ca^{2+} del medio extracelular entra a la célula a favor del gradiente electroquímico. Es más, debido a la gran magnitud de dicho gradiente, la entrada a la célula de este catión es muy rápida. No obstante, la posterior repolarización de la MP conlleva el cierre de estos canales. Estos canales son capaces de integrar una señal de Ca^{2+} local que favorece funciones celulares tales como la contracción muscular, la secreción hormonal, la transmisión sináptica, la regulación enzimática o la transcripción genética (Tyson and Snutch (2013), M. Berridge (2014)). No obstante, también se ha descrito su participación en procesos tales como mitogénesis, proliferación, diferenciación o apoptosis (Phan et al. (2017)).

Estos canales son un complejo multimérico constituido por las siguientes subunidades: α_1 , $\beta_1 - 4$, $\alpha_2\delta$ y γ . No obstante, para constituir un canal funcional basta con tres de ellas entre las cuales ha de estar α_1 , por lo que el resto de subunidades se consideran como auxiliares (figura 2.2, Tyson and Snutch (2013), M. Berridge (2014)).

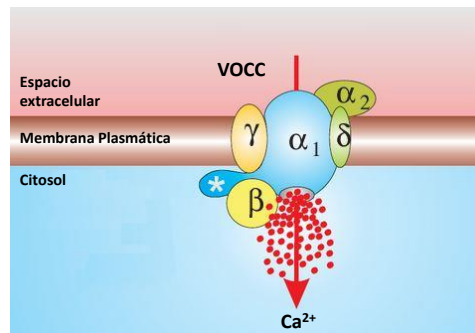


Figura 2.2: Estructura de los VOCC. En la figura se muestra un esquema de la estructura de un VOCC y sus posibles diferentes subunidades (modificado de M. Berridge (2014)).

La subunidad α_1 , que es la de mayor tamaño (entre 170 y 20 kDa), es la formadora del poro, contiene el sensor de voltaje, los mecanismos de compuerta y diferentes regiones de regulación tanto por segundos mensajeros como por fármacos y toxinas. En el genoma humano hay diez genes diferentes conocidos que codifican para ella que, además, a partir de variantes de *splicing* pueden presentar diferencias estructurales y funcionales. Es más, es a partir de esta subunidad que se nombran a los diferentes VOCC. En concreto, su nomenclatura actual consiste en el símbolo del catión al cual es permeable (en este caso Ca , de Ca^{2+}), seguido de un subíndice que indica el mecanismo de apertura/cierre (V de voltaje) y familia genética de α_1 (1 a 3): Cav1, Cav2 y Cav3. Además, dentro de cada una de estas familias se distinguen diferentes canales en función del gen que codifica la subunidad α_1 : Cav1 agrupa a Cav1.1-4, cuyas subunidades son las α_{1S} , α_{1C} , α_{1D} y α_{1F} ; Cav2.1-3, con α_{1A} , α_{1B} y α_{1E} , y Cav3.1-3 para α_{1G} , α_{1H} y α_{1L} (Tyson and Snutch (2013), M. Berridge (2014)).

En cuanto a su estructura, en esta subunidad se distinguen cuatro dominios homólogos unidos entre sí dentro de la misma secuencia proteica (dominios del I al IV) que, a su vez, contienen seis dominios transmembrana cada uno de ellos (TM1 a TM6). De todos ellos, es el TM4 el que presenta las cargas positivas encargadas de detectar las variaciones de voltaje. Respecto al poro, es la unión entre los TM5 y TM6 quien lo forma. No obstante, aunque esta es la estructura general, existen diferencias funcionales entre los miembros de las familias VOCC que, en gran parte, pueden deberse a la variabilidad estructural conservada entre dominios. En este sentido, se han identificado regiones en las que se producen interacciones moduladoras específicas o modificaciones. Es más, algunas de estas regiones son exclusivas de cada familia de VOCC lo que le diferencia a sus miembros de los de otras familias. No obstante, también hay otras regiones que se comparten por más de una de las familias. En la figura 2.3 se muestra la estructura de la α_1 para cada una de las tres familias y diferencias estructurales entre ellas (Tyson and Snutch (2013)).

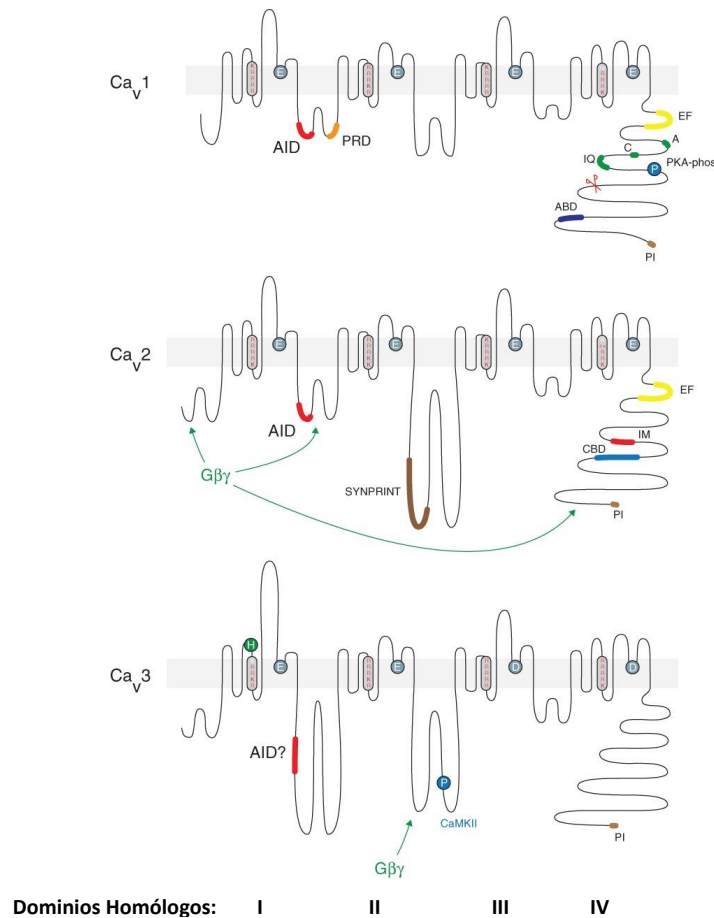


Figura 2.3: Estructura de la subunidad α_1 y dominios reguladores para las diferentes familias de VOCC. La subunidad α_1 está formada por cuatro dominios homólogos (I-IV) unidos en una misma secuencia proteica, cada uno de los cuales contiene seis dominios transmembrana. Se muestra la estructura y dominios regulatorios de α_1 para cada una de las tres familia VOCC. Abreviaturas: AID, dominio de interacción de α_1 ; PRD, dominio rico en proteínas; PKA-phos, fosforilación por PKA; ABD, dominio de unión a AKAP15; A, C, IQ, dominios AC/IQ; CBD, sitios de unión a calmodulina; IM, motivo IQ modificado; PI, dominio de interacción con PDZ domain Interaction; $G\beta\gamma$, subunidad $G\beta\gamma$ de la proteína G (modificado de Tyson and Snutch (2013)).

Como puede observarse en la figura 2.3 (Tyson and Snutch (2013)):

- Al dominio AC/IQ, que está únicamente en los Cav1, se unen de forma competitiva calmodulina (CaM) y otras proteínas de unión a Ca^{2+} , lo que influye en la facilitación e inactivación dependiente de Ca^{2+} de estos.
- Los dominios CBD (dominio de unión a CaM, del inglés *CaM binding domain*) y el IM (motivo IQ modificado, de inglés *IQ modified domain*) están únicamente en los Cav2, los cuales interactúan con CaM para regular tanto la facilitación como la inactivación dependiente de Ca^{2+} . Además, CBD también interactúa con otros sensores de Ca^{2+} .
- Los motivos EF o manos EF, que unen Ca^{2+} , están tanto en los Cav1 como en los Cav2 y podrían tener un efecto modulador.
- La subunidad $G_{\beta\gamma}$ de las proteínas G es capaz de unirse Cav2 (unión I-II) y Cav3 (unión II-III) para ejercer un efecto inhibitorio sobre el canal. No obstante, en Cav3.2 pueden modular esta inhibición a través de una fosforilación por la CaMKII en la región de unión con esta subunidad.

En cuanto al resto de subunidades, que son auxiliares, cada una de ellas tendrá una función concreta en estos canales y de las diferentes combinaciones de estas con α_1 se obtendrán diferentes VOCC:

- Por un lado, la subunidad β puede estar codificada por cuatro genes diferentes (CACNAB1-4), por lo que existen 4 isoformas diferentes. Su función es regular el tráfico y la actividad de α_1 , pues su coexpresión con α_1 tiene efecto en el aumento de la amplitud de la corriente, que el voltaje de activación e inactivación disminuya y que la velocidad de inactivación se eleve (Hanlon and Wallace (2002), M. Berridge (2014)).
- En cuanto a las subunidades α_2 y δ , las cuales derivan de un único gen cuya proteína precursora sufre una escisión proteolítica que da lugar a α_2 y α_1 que, posteriormente, forman un puente disulfuro entre ellas y pasan a actuar como una única subunidad la cual actúa en la parte externa de la MP. Se han identificado cuatro tipos cuyo papel es el de modular los VOCC y es capaz de modificar la interacción con inhibidores (Tyson and Snutch (2013)).
- Por último, para la subunidad auxiliar γ , se han identificado ocho de ellas, cada una de las cuales tiene preferencia por uno u otro tejido. Por ejemplo, γ_1 , que se encuentra de forma exclusiva en el músculo esquelético, reduce el paso de Ca^{2+} a través de los Cav1. Por el contrario, como γ_7 se expresa en el cerebro y parece limitar la expresión de Cav2.2 (M. Berridge (2014)).

Estos canales también pueden clasificarse atendiendo a su *umbral de activación* y al tipo de corriente de Ca^{2+} que presentan (tabla 2.1). En particular, a partir de la caracterización de las corrientes de Ca^{2+} como las llevadas a cabo en un estudio pionero por Fenwick, Marty, and Neher (1982) en el cual estudiaron estas por *patch-clamp*, se han englobado en dos tipos. Por un lado, están los de *alto umbral de activación (HVA)*, que requieren de potenciales superiores a -30 mV y que, además, se subdividen en los *tipo L*, que muestran una corriente de larga duración y una inactivación lenta, y los *No-L (N, P/Q y R)*. Por otro lado, los de *bajo umbral de activación (LVA)* se activan por encima de los -50 mV, lo que supone que se activen ante pequeñas variaciones en el potencial de membrana. Además, presentan corrientes transitorias y una inactivación rápida. Cabe destacar que mientras que para los primeros se han encontrado inhibidores selectivos como las *dihidropiridinas (DHP)* para los tipo L o las toxinas polipeptídicas de serpientes y arañas para los *No L* (conotoxinas para tipo N o agatoxinas para P/Q), para los LVA no se conoce ninguno selectivo por el momento (Tyson and Snutch (2013), M. Berridge (2014)). Además, cabe destacar que la sobreexpresión de STIM1 (sensor de Ca^{2+} del RE y activador de la SOCE) es capaz de inhibir el componente del poro de Cav1.2 y, además, que ORAI1 (canal canónico operado por depósitos) aproxima a STIM1 hacia Cav1.2 (Youjun Wang et al. (2010)).

Por lo general, estos canales suelen expresarse en las células excitables, como por ejemplo las células del músculo esquelético, liso o cardíaco, en neuronas, en células neuroendocrinas y endocrinas. No obstante, también se han encontrado expresadas en células no excitables tales como el hígado o el riñón. En concreto, los patrones de expresión generales son: Cav1 en células musculares, neuronales y endocrinas; Cav2 en neuronas, y Cav3 tanto en células excitables como no excitables. Cabe destacar que una misma célula es capaz de expresar más de una isoforma y, en ocasiones, como sucede en las neuronas, cada isoforma se localiza en regiones celulares diferentes (M. Berridge (2014)).

En cuanto a las posibles patologías que pueden aparecer por alteraciones en estos canales, algunas de las que se han observado, junto con el canal involucrado, son: esquizofrenia (Cav1.2), problemas cardiovasculares como bradicardia (Cav1.3), enfermedades a nivel auditivo como la sordera congénita (Cav1.4), problemas de visión como la distrofia cono-bastón (Cav2.1) o la ceguera nocturna congénita (Cav2.2) o enfermedades epilépticas (Cav3) (Tyson and Snutch (2013)). Además, también se ha observado su participación en diferentes tipos de cáncer, como por ejemplo el de colon, pues a mostrado participar en procesos tales como la mitogénesis, la proliferación, la diferenciación, la apoptosis e, incluso, en la metástasis (X.-T. Wang et al. (2000), Monteith et al. (2007), Taylor et al. (2008), Phan et al. (2017)). En particular, se ha observado que Cav3 está implicado en la proliferación celular en el cáncer de colon. Además, se ha identificado la sobreexpresión de la mayoría de las isoformas de los Cav (M. Berridge (2014), X.-T. Wang et al. (2000), L. Zhang et al. (1997), Toyota et al. (1999), S. G. Kim et al. (2003)).

Estos VOCC presentan ciertas diferencias a nivel señalización entre ellas. Por un lado, Cav1 es la que más fácilmente responde ante la despolarización, la que mejor señalización hacia el núcleo presenta e inducir la expresión de diferentes genes regulados por CREB (*CycA*, *BCL2*, *ERGI*), implicado en la fisiopatología del cáncer (Sakamoto and Frank (2009)), lo que está relacionado

con su colocalización con clústers de CaMKII. Por otro lado, Cav2 carecen de la habilidad para activar una señalización hacia el núcleo y, además, para poder funcionar han de estar a una distancia superior a $1 \mu\text{m}$ de los clústers Cav1-CaMKII, lo que dificulta su funcionamiento y su capacidad para aumentar $[\text{Ca}^{2+}]_{\text{cit}}$. No obstante, el Ca^{2+} introducido por Cav2, frente al Cav1, es preferiblemente captado por el RE y la mitocondria (H. Ma et al. (2012)).

En la tabla 2.1 se muestra un resumen de los diferentes VOCC, sus correspondientes subunidades α_1 , el tipo de corriente, función fisiológica, sus localizaciones predominantes e inhibidores específicos.

Tabla 2.1: Clasificación de los VOCC.

| VOCC | Familia α_1 | Subunidad α_1 | Tipo de Corriente | Funciones | Tejido | Inhibidores Específicos |
|--------|--------------------|----------------------|-------------------|---|---|---|
| Cav1.1 | | α_{1S} | | Acoplamiento Excitación-Constracción | Músculo Esquelético | |
| Cav1.2 | | α_{1C} | | Acoplamiento Excitación-Constracción Regula Actividad Enzimática Regula Expresión Genes | Músculo cardíaco Músculo Liso Neuronal | |
| Cav1.3 | 1 | α_{1D} | HVA: L | Secreción Endocrina Marcapasos Transducción Auditiva | Páncreas cardíaco Neuronal Médula Adrenal | DHP |
| Cav1.4 | | α_{1F} | | Transducción Visual Señalización/Homeostasis Cél.T | Ojo Sistema Inmune | |
| Cav2.1 | | α_{1A} | HVA: P/Q | | | ω – AgaIVA |
| Cav2.2 | 2 | α_{1B} | HVA: N | Liberación de Neurotransmisores | Neuronal | ω – Ctx – GVIA ω – Ctx – MVIIC |
| Cav2.3 | | α_{1E} | HVA: R | | | SNX-482 |
| Cav3.1 | | α_{1G} | | Marcapasos Descargas Repetitivas Proliferación | Neuronal cardíaco | No tiene |
| Cav3.2 | 3 | α_{1H} | LVA: T | Marcapasos Descargas Repetitivas Proliferación | Neuronal cardíaco Riñón | Ni^{2+} |
| Cav3.3 | | α_{1L} | | Proliferación | Neuronal | No tiene |

2.3.1.2. Canales de Ca^{2+} operados por receptor (ROCC)

Los *ROCC* se activan tras la unión de un ligando concreto -hormona, neurotransmisor... - en el dominio extracelular de su receptor. En consecuencia, se activa la apertura del canal iónico asociado a este receptor para permitir el flujo de ciertos iones a favor del gradiente electroquímico (M. Berridge (2014)). Estos canales pueden subdividirse en dos tipos, según el receptor y el canal iónico que formen o no parte de la misma estructura: *ionotrópicos* y *metabotrópicos*, respectivamente. A continuación, se explican brevemente cada uno de ellos.

- *Receptores ionotrópicos u operados por ligando (LGIC, del inglés ligand gated ion channel)*: estos canales se caracterizan porque tanto el receptor como el propio canal forman una sola estructura y suelen estar formados por dos o más dominios: un TM, que forma el poro a través del cual pasan los iones (Ca^{2+} , K^{+} , Na^{+} o Cl^{-}), y otro dominio extracelular al cual se une el ligando. Debido a esta localización receptor-canal en una misma proteína, la activación del receptor a través de la unión del ligando extracelular modula directamente la apertura del canal -algunos pueden activarse por cambios en el potencial de membrana- y, en consecuencia, la respuesta de estos canales al estímulo es rápida (figura 2.4a). Esta superfamilia de canales incluye a cuatro tipos principales: *receptores Cys-loop, de glutamato ionotrópicos (iGluR), canales dependientes de ATP* y la *superfamilia de los TRP*. En cuanto a la respuesta celular que producen, puede ir desde la despolarización de la membrana por el flujo de iones a favor del gradiente hasta la entrada al citosol del propio Ca^{2+} el cual podrá actuar como segundo mensajero.
- *Receptores metabotrópicos, GPCR, u operados por segundos mensajeros (SMOCC)*: a diferencia de los LGIC, estos canales están separados del receptor y, además, pueden encontrarse en diferentes membranas de la misma célula, por lo que ha de existir alguna vía de comunicación entre ambos. En consecuencia, esta separación y la necesidad de una comunicación entre ambas proteínas les confiere la particularidad de dar una respuesta más lenta al estímulo en comparación con los LGIC, pero, a su vez, se consigue una respuesta más duradera y una amplificación de la señal que activa el canal (figura 2.4b). Estos receptores, debido a su estructura, también se conocen como receptores con siete TM, de tal manera que estos dominios están unidos mediante *loops* intracelulares -interaccionan con la proteína G- y extracelulares (implicados en el reconocimiento de los ligandos). Sin entrar en mucho detalle, existen tres familias de estos receptores: A, que es la más numerosa y responde a prácticamente todos los aminoácidos; B, que responde a hormonas y péptidos gastrointestinales como el glucagón, y la C, que responden a neurotransmisores como el Glutamato y el GABA, importantes en la homeostasis del Ca^{2+} intracelular, e incluso responden también al propio Ca^{2+} .

Estos receptores ejercen su actividad a través de las proteínas triméricas G (subunidades α y $\beta - \gamma$) que tienen acopladas. Así, mientras que en reposo el receptor y la proteína G están separados, esta última en forma trimérica y unido a GDP, la activación del receptor favorece la interacción receptor-proteína G y, en consecuencia, se intercambia el GDP por GTP. En función de la proteína G y de sus subunidades, se pueden desencadenar diferentes rutas de señalización, como por ejemplo (Ortega-Gutiérrez (2013)): las subunidades $\alpha_{i,o}$ de proteínas G_i disparan la inhibición de los canales de Ca^{2+} ; las proteínas G_q desencadenan la activación de la Fosfolipasa $C\beta$ (PLC) y, en consecuencia, generan IP3 que, a su vez, activa la liberación del Ca^{2+} de los depósitos (Receptor Muscarínico de Acetilcolina mAChR-); las subunidades $\beta\gamma$ de todas las proteínas G activan los VOCC tipo L, la PLC β ... Debido a su mecanismo de apertura, los receptores muscarínicos son un caso especial de los SMOCC estudiados en el siguiente apartado.

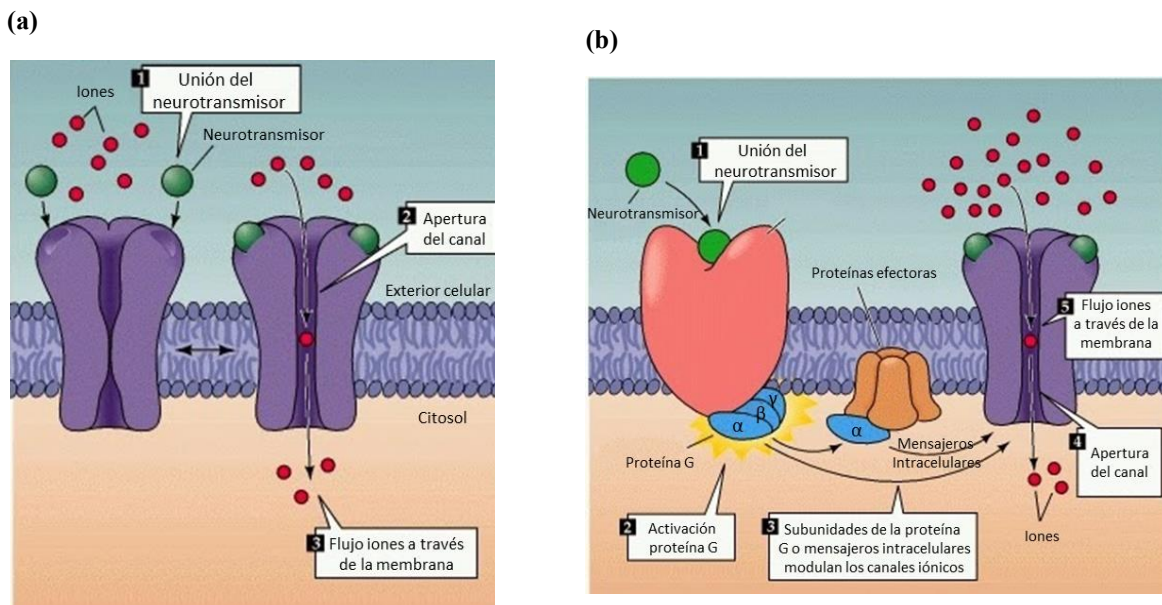


Figura 2.4: Receptores Operados por Receptor. Tras la unión de un ligando concreto a su dominio extracelular, estos receptores se activan y permiten el flujo de iones a través de su canal asociado a favor del gradiente electroquímico. En función de si el receptor y el canal forman parte o no de la misma estructura, se distinguen dos tipos: **a)** receptores ionotrópicos, aquellos cuyo canal asociado forman parte de la misma estructura que el receptor, y **b)** receptores metabotrópicos, cuyo canal es una estructura diferente al receptor. Estos últimos pueden ser GPCR o TKR y activar desde canales hasta diferentes rutas de señalización corriente abajo.

Dentro de estos ROCC hay varios que permiten la entrada de Ca^{2+} al citosol: los receptores de glutamato, tanto ionotrópicos como metabotrópicos; receptores de acetilcolina, que también pueden ser ionotrópicos y metabotrópicos, y los receptores purinérgicos. A continuación, se explicará brevemente cada uno de ellos:

- *Receptores de glutamato*: El glutamato es el neurotransmisor excitatorio por excelencia, bien conocido por su participación, por ejemplo, en el centro de recompensa del cerebro (Goodwani et al. (2017)). Este se libera al medio extracelular en las brechas sinápticas, por lo que activa los *Receptores de Glutamato* de las células postsinápticas y participan en procesos como la plasticidad sináptica la expresión de genes. Estos, a parte de en las células del sistema nervioso, también se encuentran en otros tipos celulares como en el páncreas, los osteoblastos o los osteoclastos, y participan en procesos como la producción de hormonas en la glándula adrenal y el páncreas (Julio-Pieper et al. (2011), Nedergaard, Takano, and Hansen (2002), Goodwani et al. (2017)). Estos receptores pueden dividirse en dos familias, *ionotrópicos (iGluR)* y *metabotrópicos (mGluR)*, de cada una de las cuales se conocen varias isoformas (Julio-Pieper et al. (2011), Blackshaw, Page, and Young (2011), Willard and Koochekpour (2013)):

- *iGluR*, en función del agonista, se han descrito tres: *receptores NMDA (NMDAR)*, activados por el *N-metil-D-aspartato*, y *no NMDA (receptores AMPA -AMPA-*, por el amino-3-hidroxi-5-metil-4-isoxazol-propionato, y *receptores kainato -KAR-*, activados por el ácido kaínico). Estos generan una respuesta rápida y permiten el paso no selectivo tanto de Na^+ (No NMDA) como de Ca^{2+} (NMDA), de tal manera que se despolariza la MP de la neurona postsináptica (Blackshaw, Page, and Young (2011), Willard and Koochekpour (2013), Connaughton (2007), Fouilloux (2004), Simeone, Sanchez, and Rho (2004)).

De los tres, sólo los NMDA dejan pasar Ca^{2+} y, además, también son los que mayor regulación presentan: co-agonistas como la glicina, la D-serina o algunas poliaminas, y voltaje consecuencia del Mg^{2+} que lo bloquea. También presenta regulación negativa por zinc y por pH ácido. En cuanto a su estructura, los NMDAR están formados por diferentes combinaciones de las subunidades NR1, NR2A, NR2D, NR3A y NR3B, de tal manera que haya dos NR1 y, o bien dos NR2 o una NR2 y otra NR3 (Connaughton (2007), Goodwani et al. (2017), L. Sun et al. (1998)). Cabe destacar que los receptores NMDA juegan un papel importante en la plasticidad neuronal de tal manera que según el número de receptores ionotrópicos de glutamato en la membrana de la célula postsináptica, esta plasticidad puede ser de potenciación a largo plazo o depresión a corto plazo, importantes en los procesos de aprendizaje y memoria (Bleich et al. (2003), Simeone, Sanchez, and Rho (2004)). En cuanto a la supervivencia de la célula, estos receptores parecen tener un papel positivo, salvo cuando la concentración del glutamato es tal que da lugar a una sobreactivación de los NMDAR, lo que se conoce como *excitotoxicidad*, proceso clave en patologías como la enfermedad de Alzheimer (Mark et al. (2001), Y. Liu et al. (2007)).

- *mGluR*: estos receptores, que están acoplados a proteínas G, se dividen en 3 subfamilias de acuerdo a criterios de homología de secuencia y de qué agonistas son capaces de activarlos. Por un lado, los receptores del grupo I, compuesto por los mGluR1 y 5, activan la vía del IP3/DAG, por lo que tienen un papel fundamental en la homeostasis

del Ca^{2+} al activar tanto su liberación desde el RE al citosol. Por otro lado, están los receptores del grupo II, que acoge a los mGluR2 y 3, y del grupo III, que agrupa a los mGluR{4, 6, 7, 8}, disminuyen los niveles de AMPc. No obstante, a pesar de que tanto mGluR1 como 5 están acoplados a una proteína Gq, presentan ligeras diferencias en su respuesta. En concreto, mientras que el mGluR1 provoca un único aumento transitorio de $[\text{Ca}^{2+}]_{\text{cit}}$, mGluR5 da lugar a oscilaciones de este catión dependientes de PKC, lo que podría explicarse en parte por la presencia de un único aminoácido diferente en la región C-terminal sobre el que actúa la PKC, Thr-840 para mGluR5 y Asp-854, para mGluR1 (C. H. Kim et al. (2005)).

Estos se expresan tanto en neuronas como en glías y participan tanto a nivel de las células postsinápticas y células accesorias en respuesta a glutamato como en la liberación de glutamato y otros neurotransmisores en las presinápticas. No obstante, también se ha observado su expresión en otros tejidos, tanto con innervación glutamatérgica -células cardíacas, glándulas adrenales...- como sin tal innervación -linfocitos-. Es más, tanto los incrementos como los descensos de glutamato en ciertos tejidos periféricos podrían estar relacionados con diferentes patologías, como por ejemplo el incremento de glutamato con la leucemia mieloide crónica y aguda o el cáncer (Julio-Pieper et al. (2011), Willard and Koochekpour (2013), Blackshaw, Page, and Young (2011), Goodwani et al. (2017), Nedergaard, Takano, and Hansen (2002)).

- *Receptores de acetilcolina o colinérgicos (AChR)*: la acetilcolina, principalmente liberada por las neuronas colinérgicas, está implicada en diferentes procesos tanto en el sistema nervioso central -función motora, percepción sensorial, procesos cognitivos, ritmos circadianos o nocicepción- como en el periférico -ritmo cardíaco, motilidad gastrointestinal o la actividad del músculo liso-. Este neurotransmisor actúa sobre los AChR, que pueden agruparse en dos: los *receptores nicotínicos de acetilcolina (nAChR)*, ionotrópicos -expresados en las uniones neuromusculares, sinapsis neuronales y en células no excitables- y los *receptores muscarínicos de acetilcolina (mAChR)*, metabotrópicos miembros de las GPCR (C. K. Jones, Byun, and Bubser (2012), M. Berridge (2014)):
 - *nAChR*: pertenecen a la familia de los *lazos de cisteína*, cuya estructura común a todos los miembros de esta familia es un pentámero de cinco subunidades que se dispone creando un poro a través del que pasan los iones y, además, presentan un enlace disulfuro en la cara extracelular entre dos cisteínas, característica por la cual reciben su nombre. En concreto, los nAChR están formados por diferentes combinaciones de las diferentes subunidades α , β , γ , ϵ y δ , y de sus isoformas, aunque los patrones de expresión de las subunidades y de sus diferentes isoformas varía entre los diferentes tipos celulares en los que se expresan estos receptores. Estos receptores participan en procesos como la regulación del ACh y otros neurotransmisores o la contracción muscular, y se expresan en células tanto pre como postsinápticas, aunque con diferentes patrones de expresión según el subtipo de nAChR y del tipo celular. Su apertura puede

desencadenarse por la unión de la ACh, aunque la nicotina también puede ejercer esta función y, tras su activación, transmiten información al interior celular mediante dos mecanismos principales, vía despolarización de la membrana y vía señalización de Ca^{2+} (M. Berridge (2014)).

Por un lado, la despolarización de la membrana permite el paso de los cationes Na^+ , Ca^{2+} y K^+ lo cual conlleva la despolarización de la membrana y la activación de mecanismos voltaje-dependientes (M. Berridge (2014)). Por otro lado, la activación de la señalización de Ca^{2+} tiene lugar tanto en células excitables como no excitables -células T o B-, donde se ha observado que diferentes combinaciones de las subunidades y de sus isoformas dan lugar a receptores con diferentes permeabilidades a Ca^{2+} y diferentes afinidades a ACh, por lo que es posible obtener una variedad de respuestas celulares diferentes en el contexto de la señalización de Ca^{2+} (M. Berridge (2014), Mansvelder and McGehee (2000), Mansvelder, Keath, and McGehee (2002)).

- *mAChR*: al igual que ocurría con los mGluR, los mAChR son GPCR cuyas proteínas G también son G_q o $G_{i/o}$. Se han descrito cinco tipos los cuales pueden agruparse en dos grupos en función de qué proteína G tengan acoplada. Por un lado, los *M1*, *M2* y *M3* están acoplados a G_q , por lo que activan $\text{PLC}\beta$ y dan lugar a IP_3 y DAG y, en consecuencia, se libera el Ca^{2+} del RE al citosol y se activa la PKC, respectivamente. Por otro lado, los *M2* y *M4* acoplan $G_{i/o}$, lo que conlleva la inhibición de la adenilato ciclasa, por lo que se reduce el AMPc, y la prolongación de la apertura de los canales de K^+ , catiónicos no selectivos y TRP. Por lo tanto, está claro que cada tipo de mAChR puede tener efectos diferentes a nivel fisiológico.

Por un lado, el mAChR *M1*, que es la forma predominante de los mAChR del sistema nervioso central, desempeña papeles importantes en procesos cognitivos, posiblemente a través de su efecto potenciador de las corrientes a través de los NMDAR tanto en el hipocampo como en el cerebro anterior. Además, su activación incrementa la excitación sináptica de las células piramidales del córtex medial prefrontal y mejora las funciones cognitivas mediadas por esta región. Por el otro lado, el *M2*, que se expresa en las neuronas presinápticas de las terminaciones colinérgicas del cerebro, donde actúan como autorreceptores para inhibir la liberación de ACh, y en los terminales axónicos de las neuronas no colinérgicas, cuyo papel en esta localización es de heterorreceptor para la regulación presináptica de la liberación de otros neurotransmisores. Por lo tanto, agonistas de *M1* y antagonistas de *M2* podrían ser beneficiosos para tratar el déficit cognitivo asociado a ciertas patologías como la esquizofrenia (C. K. Jones, Byun, and Bubser (2012)).

- *Receptores purinérgicos P2X*: estos son canales catiónicos ionotrópicos que se abren en respuesta a ATP extracelular. Están formados tanto por homo como por heterotrímeros con dos hélices transmembrana por subunidad de tal manera que ambos extremos se localizan en la cara intracelular. Se han descrito 7 genes que codifican para estos. En particular, los *P2X*, tras unirse el ATP, se abren e inducen la despolarización de la MP y permiten el paso

de Ca^{2+} , de tal manera que aumenta su concentración en el citosol. En consecuencia, median procesos tanto de respuesta celular rápida como lenta, lo que puede estar determinado por la composición de las subunidades. Por un lado, canales como el P2X2 o el P2X3, que tienen una inactivación rápida, median respuestas rápidas como la exocitosis y la secreción que tiene lugar en las neuronas en la propagación el impulso nervioso o en la contracción muscular. Por otro lado, otros como el P2X7 -inactivación lenta-, intervienen en procesos de respuesta más lenta como el crecimiento celular. Se localizan tanto en neuronas y glías como en músculo, endotelio, epitelio, hepatocitos... Se ha observado la implicación estos canales en algunas patologías tales como neurodegenerativas -Alzheimer y Parkinson-, inflamatorias -inflamación intestinal del hígado- o el desarrollo de cáncer e, incluso, se ha considerado como posible diana terapéutica para otras como la retinopatía diabética (M. Berridge (2014), Territo and Zarrinmayeh (2021), Jain and Jacobson (2021), Tassetto et al. (2021), Suresh, Phasuk, and Liu (2021)).

2.3.1.3. Canales de Ca^{2+} operados por segundo mensajero (SMOCC)

Los SMOCC son aquellos canales capaces de activar su apertura a través de la acción de una señal intracelular en la que participan segundos mensajeros, tales como nucleótidos cíclicos -canales operados por nucleótidos cíclicos (CNGC)-, ácido araquidónico -canales de Ca^{2+} regulados por ácido araquidónico (ARC)-, DAG -algunos TRPC- (M. Berridge (2014)). No obstante, los receptores muscarínicos estudiados en el apartado anterior también podrían ser considerados un tipo de SMOCC, pues se activan por acción de segundos mensajeros tras la acción de un estímulo externo (Potier and Trebak (2008)).

Los CNGC son heterotetrámeros que contienen diferentes combinaciones de las subunidades A y B los cuales se activan por nucleótidos cíclicos, tras lo que permiten el paso al citosol de Ca^{2+} y Na^+ desde el medio extracelular. En total, se han descrito seis genes en los mamíferos que codifican para estas subunidades: cuatro CNGA y dos CNGB, respectivamente. De ellas, las CNGA1-3 son las formadoras del canal, mientras que el resto tienen una función moduladora las cuales presentan un sitio de unión a CaM tipo IQ en su extremo N-terminal capaz de unir apo-CaM y, en caso de tal unión, permitir que el Ca^{2+} actúe como regulador negativo e inactivar el canal. Cabe destacar que los CNGA pueden dar lugar a canales funcionales *per se*, pero no las CNGB. En cuanto a su distribución, se ha encontrado expresión de estos en diferentes tipos celulares, desde aquellas de los órganos sensoriales -visión, olfato y gusto- hasta otras como el cerebro o el riñón. Alteraciones de estos genes pueden dar lugar a enfermedades, como por ejemplo la retinitis pigmentosa en el caso de mutaciones en CNGA1 y CNGB1, o acromatopsia por mutaciones autosómicas recesivas en CNGA3 y CNGB3 (M. Berridge (2014), Pang et al. (2010)).

Otro SMOCC relevante en la homeostasis del Ca^{2+} intracelular de las células parotidias y acinares pancreáticas, es el canal ARC, pues tiene una gran selectividad por este catión. Se han relacionado también canales como ORAI1 u ORAI3 pueden formar canales ARC para permitir la entrada de Ca^{2+} independiente de depósitos. Es más, se ha encontrado que este hecho sucede

en células cancerígenas de mama, colon o pulmón pero no en las normales (M. Berridge (2014), Thompson and Shuttleworth (2013), Tanwar, Arora, and Motiani (2020)). Respecto a los TRPC, que pueden ser considerados como SMOCC por su activación a través de DAG, se explicarán en mayor detalle en el apartado “2.7 Superfamilia de los TRP” .

2.3.1.4. Canales de Ca^{2+} operados por depósitos (SOCC)

El Ca^{2+} libre intracelular se encuentra, por lo general, almacenado en diferentes depósitos intracelulares como el RE, el Golgi o los lisosomas. Es más, gran parte del Ca^{2+} involucrado en la señalización intracelular proviene del almacenado en el RE (M. Berridge (2014)). Como se verá más adelante, existen diferentes mecanismos de liberación de Ca^{2+} desde el RE al citosol. Por ejemplo, mediante respuestas celulares a estímulos externos mediadas por ROCC, los cuales activan la PLC que, a su vez, cataliza la conversión de fosfatidilinositol 4,5-bisfosfato (PIP2) a inositol 1,4,5-trifosfato (IP3) y DAG. Así, el IP3 generado activa los *receptores de inositol trifosfato (IP3R)* de la membrana del RE y, en consecuencia, se libera Ca^{2+} al citosol (Prakriya and Lewis (2015)). No obstante, también existen otros canales a través de los cuales puede liberarse Ca^{2+} desde el RE al citosol: RyR, leak, TRPM8...

En consecuencia, para asegurar la disponibilidad de Ca^{2+} y que este pueda ejercer su función señalizadora, es necesaria la acción de otros mecanismos que, por un lado, contrarresten este aumento de $[\text{Ca}^{2+}]_{\text{cit}}$ y, por otro, rellenen los depósitos intracelulares de Ca^{2+} . En este sentido, entra en juego la SERCA, una bomba del RE que bombea Ca^{2+} a su interior, que sólo es capaz de captar parte de este catión recién liberado. En consecuencia, a través de este único mecanismo el RE sería incapaz de recuperar todo el Ca^{2+} que ha liberado. A la larga, esto conllevaría su total vaciado y supondría la pérdida de la capacidad señalizadora y la activación de diferentes rutas de señalización de *estrés del RE* y de apoptosis (M. Berridge (2014)).

Por ello, existen otros mecanismos a través de los cuales se pueden reestablecer los niveles basales de Ca^{2+} de los depósitos previos a la liberación. En particular, en las células eucariotas se ha identificado una familia de canales denominados *canales de Ca^{2+} operados por depósitos (SOCC)* los cuales, tras el agotamiento de las reservas de Ca^{2+} del RE, se activan y se abren para permitir la entrada a la célula de Ca^{2+} desde el medio extracelular a través de la MP. En consecuencia, este Ca^{2+} incrementará la $[\text{Ca}^{2+}]_{\text{cit}}$ y podrá ser empleado para el llenado de los depósitos de Ca^{2+} intracelulares. Al mecanismo anterior se le conoce como *Entrada Capacitativa de Ca^{2+} (SOCE)*, del inglés *Store-Operated Ca^{2+} Entry*) (Prakriya and Lewis (2015)).

Cabe destacar que este mecanismo de transporte es el mecanismo de entrada de Ca^{2+} más importante en las células no excitables -como es el caso de las células epiteliales de colon-, aunque no es exclusivo de estas, pues también puede estar presente en células excitables como las musculares o las neuronas (Parekh and Putney Jr (2005), Brini et al. (2013), Sobradillo et al. (2014), Villalobos et al. (2017)). En cuanto a la importancia o papel fisiológico de este mecanismo de transporte, más allá del relleno de los depósitos intracelulares, parece estar implicado en

diferentes procesos donde es necesario un mantenimiento a largo plazo de la señalización de Ca^{2+} , como por ejemplo en la proliferación celular o en la expresión de genes. Es más, se ha visto que en células tumorales de colon, respecto de las normales, existe una mayor SOCE y más prolongada en el tiempo la cual está positivamente correlacionada con la capacidad proliferativa de estas, a pesar de que el tamaño de los depósitos es inferior (Sobradillo et al. (2014), M. Berridge (2014)). También se ha descrito su participación en la exocitosis, apoptosis o en la regulación enzimática.

No fue hasta 2005 y 2006 que se descubrió a los primeros participantes culpables de la detección de la disminución de Ca^{2+} del RE -STIM1- y de la entrada de este desde el medio extracelular -ORAI1- (Liou et al. (2005), Feske et al. (2006), Prakriya et al. (2006)). A partir de entonces, se han ido incorporando a estos más componentes de este mecanismo. Por un lado, una isoforma de STIM, *STIM2*, que también actúa como sensor de Ca^{2+} en el RE (Williams et al. (2001), Soboloff et al. (2006), Parvez et al. (2008)). Por otro lado, también se han identificado otros SOCC con diferente selectividad por Ca^{2+} tales como las isoformas de ORAI *ORAI2* y *ORAI3* (Mercer et al. (2006)) e incluso miembros de la familia de los TRP, como por ejemplo el *TRPC1* o *TRPV4,6* (figura 2.5) (M. Berridge (2014)). Cabe destacar que estos se han caracterizado mediante técnicas electrofisiológicas, pues, por ejemplo, ORAI1 es conocido como el canal de *canal de Ca^{2+} activados por la liberación de Ca^{2+} (CRAC)* puesto que presenta una conductancia muy baja y selectiva por este catión (Sallinger, Berlansky, and Frischauf (2020)).

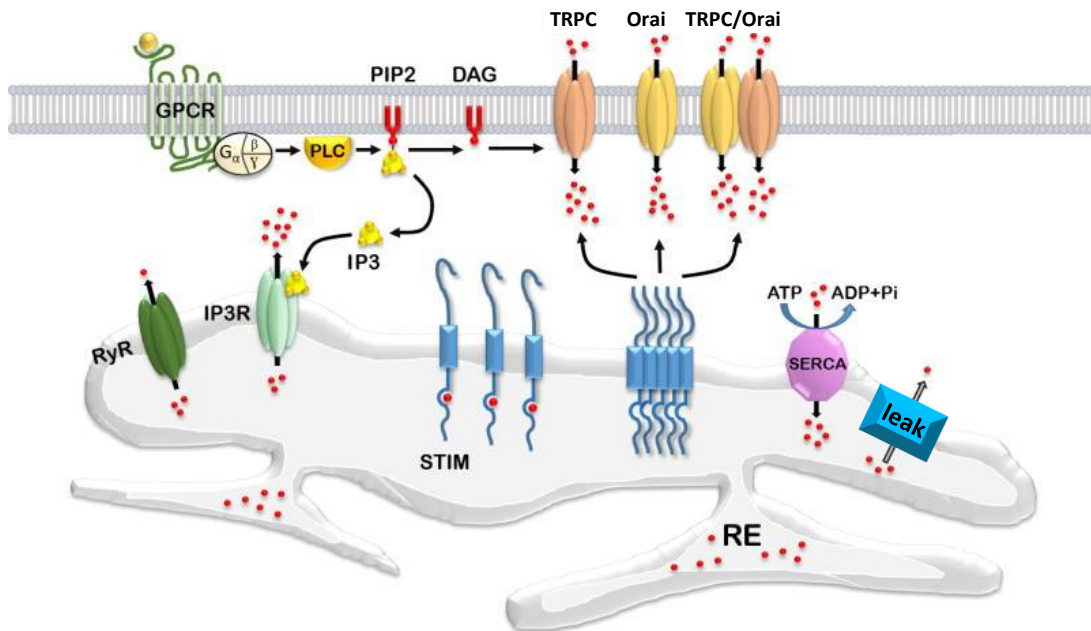


Figura 2.5: Activación de los canales de Ca^{2+} operados por depósitos. En respuesta a un estímulo, como puede ser el carbacol que activa a su GPCR, los depósitos de Ca^{2+} intracelulares y, en concreto, el RE, puede liberar su contenido de Ca^{2+} al citosol y vaciarse. Además, también hay presente unos canales *leak* que de forma constitutiva liberan Ca^{2+} al citosol. Esta disminución de Ca^{2+} es detectada por los sensores de Ca^{2+} STIM los cuales, en estas condiciones, oligomerizan e interaccionan con los SOCC (ORAI, TRPC...) de la MP para permitir la entrada de Ca^{2+} a la célula desde el medio extracelular, de tal manera que este Ca^{2+} puede entrar en el RE a través de la SERCA en contra del gradiente (modificado de Pchitskaya, Popugaeva, and Bezprozvanny (2018)).

La importancia de la SOCE tanto en la homeostasis del Ca^{2+} citosólico como en diferentes procesos celulares dependientes de Ca^{2+} es especialmente evidente cuando existen diferentes mutaciones tanto de ganancia como de pérdida de función tanto en STIM como en ORAI y que se relacionan con diferentes enfermedades. Por un lado, las mutaciones de ganancia de función, que dan lugar a una activación constitutiva de los canales CRAC y activación de la SOCE, conllevan el aumento de $[\text{Ca}^{2+}]_i$ y que, a nivel clínico, puede manifestarse como trombocitopenias o dislexia, por lo que juega un papel importante en enfermedades tales como los Síndromes de Plaquetas York o de Stormorken, entre otras. Por otro lado, las mutaciones de pérdida de función tienen el efecto totalmente opuesto, es decir, se entorpece la activación de los canales CRAC y de la SOCE que, a nivel clínico, puede dar lugar a enfermedades como la inmunodeficiencia combinada severa o la hipotonía muscular (Grabmayr, Romanin, and Fahrner (2021)).

2.3.2. Mecanismos *OFF* de la MP

Tras el aumento del Ca^{2+} que tiene lugar por los mecanismos *ON* tanto de la MP como de otros orgánulos celulares, y una vez este catión haya ejercido su función, es necesario recuperar los niveles basales de este $[\text{Ca}^{2+}]_{\text{cit}}$. Para ello, la MP, así como en otros orgánulos celulares que se tratarán en otros apartados, cuenta con una serie de *mecanismos OFF* que funcionan en contra del gradiente electroquímico (M. J. Berridge, Lipp, and Bootman (2000), M. J. Berridge (2012)): *bombas $\text{Ca}^{2+}/\text{ATPasa}$ de la MP (PMCA, del inglés plasma membrane $\text{Ca}^{2+}/\text{ATPase}$) e intercambiadores $\text{Na}^+/\text{Ca}^{2+}$ (NCX, del inglés $\text{Na}^+/\text{Ca}^{2+}$ exchanger)*.

2.3.2.1. Bombas $\text{Ca}^{2+}/\text{ATPasa}$ de la MP

Las PMCA presentan una baja capacidad y alta afinidad por Ca^{2+} , por lo que son capaces no sólo de recuperar el nivel basal de $[\text{Ca}^{2+}]_{\text{cit}}$, sino que, además, pueden continuar bombeando Ca^{2+} al medio extracelular para mantener su baja concentración en el citosol. Para llevar a cabo este transporte, emplean la energía liberada por la hidrólisis de ATP durante la cual se forma un intermediario aspartil fosfato y, por ello, pertenecen a la familia de las ATPasas transportadoras de iones tipo P y, dentro de esta, a la subfamilia P_2 o subtipo 2B (Strehler and Zacharias (2001), M. J. Berridge (2012), Hegedús et al. (2020)). Estas están codificadas por 4 genes, ATP2B1-4, que codifican para las PMCA1-4. No obstante, se han identificado alrededor de 30 variantes de *splicing* alternativo de estas bombas, muchas de ellas altamente especializadas, con un perfil de expresión específico de tejido y con funciones concretas. Por ejemplo, las isoformas 1 y 4 se expresan de forma ubicua en todos los tejidos, mientras que la 2 y la 3 son exclusivas de las neuronas del cerebro (Strehler and Zacharias (2001), Hwang et al. (2021)).

La estructura de estas bombas consiste en 10 TM cuyos extremos N-terminal y C-terminal se localizan en la cara citosólica la cual puede dividirse en 3 partes diferentes: el lazo citosólico entre los segmentos 2 y 3, el lazo citosólico entre los dominios 4 y 5 y la cola citosólica correspondiente al extremo C-terminal (figura 2.6).

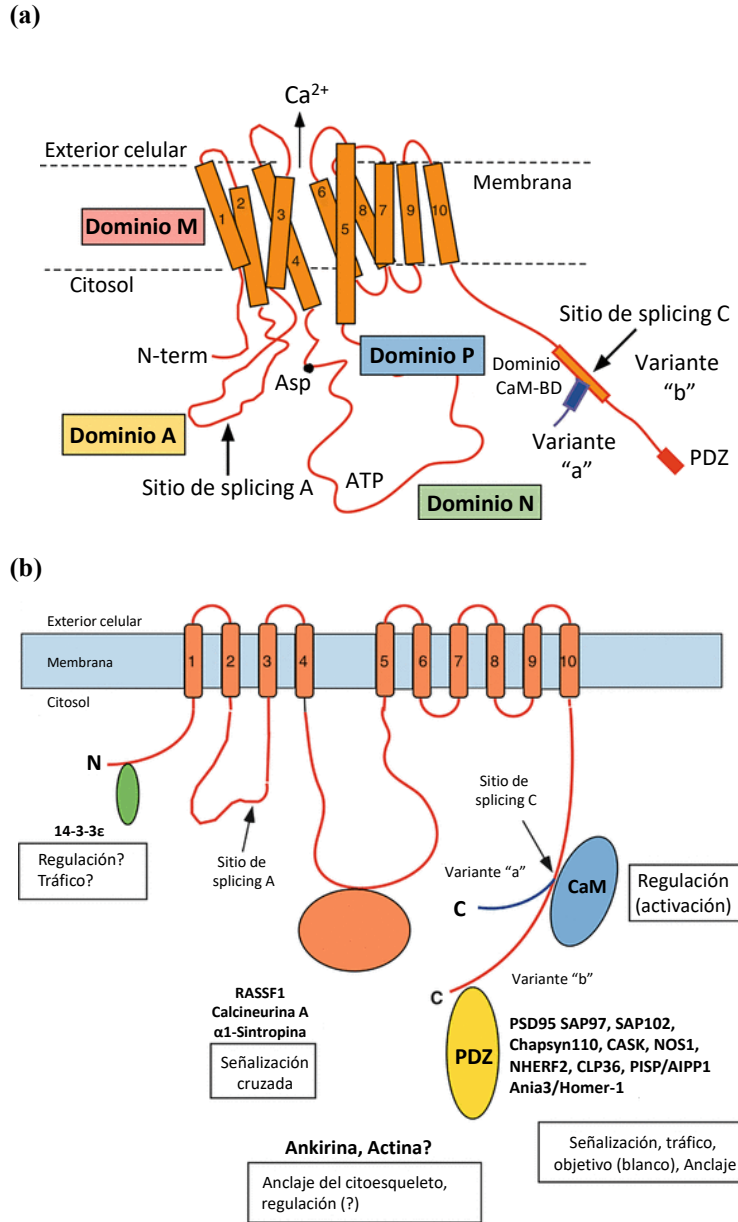


Figura 2.6: Estructura de las PMCA y regulación. **a)** La estructura de la PMCA consta de 10 TM, que forman el dominio M, y 3 lazos intracelulares que participan en diferentes procesos. En estos lazos se distinguen el residuo conservado de aspartato (Asp), que puede ser fosforilado, y el sitio de unión a ATP. Además, en su región citosólica se distinguen tres dominios: A, solenoide; P, de fosforilación, y N, de unión a nucleótidos. Los sitios de *splicing* A, B y C son aquellas regiones que suelen estar afectadas por las variantes de *splicing*. De estas, el sitio C se corresponde con el dominio de unión a CaM (CaMBD) que da lugar a dos isoformas, "a" y "b" que difieren en su extremo C-terminal debido a un cambio en el marco de lectura (mostrado con un extremo adicional azul para "a"). **b)** Estas bombas están altamente reguladas a través de la unión de diferentes proteínas, las cuales se indican como elipses sobre la región de la PMCA a la que se unen, cuyo efecto sobre dicha bomba se muestra en un recuadro negro bajo cada una de ellas. Además, también hay proteínas, como ankIRina y actina, que si bien se sabe que interaccionan con PMCA, aun no se ha descrito el sitio de la bomba con el que lo hacen. No obstante, no todas estas proteínas se unen a todas las PMCA. De hecho, las proteínas que contienen dominios PDZ sólo pueden unirse a la variante de *splicing* "b". Abreviaturas: CaM, calmodulina; CaM-BD, dominio de unión a CaM; PDZ-BD, dominio de unión a motivos PDZ; PL, región sensible a fosfolípidos; P, región susceptible de fosforilación por PKA o PKC (modificado de Strehler and Zacharias (2001)).

El primer lazo es el dominio de transducción o solenoide. El segundo incluye el dominio catalítico de ATP y el residuo de aspartato que forma el intermediario acil fosfato durante la hidrólisis del ATP. Por último, el extremo C-terminal es el dominio regulatorio que contiene un dominio de unión a Ca^{2+} -CaM que actúa como autoinhibidor en ausencia de esta. Tras el aumento de la $[\text{Ca}^{2+}]_{\text{cit}}$, este catión se une a CaM, aumentan los niveles de Ca^{2+} -CaM citoplasmáticos y se unen con gran afinidad al dominio autoinhibitorio de PMCA en el C-terminal y, en consecuencia, se estimula la actividad de esta bomba. Estos sitios de unión a Ca^{2+} -CaM también pueden ser regulados por fosfolípidos. Además, esta región reguladora también contiene zonas de regulación por Ca^{2+} , es susceptible a la fosforilación por las PKA y PKC, puede ser degradado por proteasas como la calpaina o mediar interacciones con dominios PDZ (figura 2.6b). Por lo tanto, las variantes de *splicing* (“a” y “b”) que afectan a esta región tienen implicaciones en la regulación de estas bombas. De hecho, las proteínas que contienen dominios PDZ sólo pueden unirse a la variante de *splicing* “b” (Strehler and Zacharias (2001), Strehler (2013), Hwang et al. (2021)).

Alteraciones en la expresión o variantes genéticas de estas bombas han sido relacionadas con diferentes patologías en humanos tales como enfermedades cardiovasculares, neurodegenerativas o cáncer: PMCA1/ATP2B1 se ha relacionado con hipertensión, esclerosis múltiple o cáncer oral; PMCA2/ATP2B2 con autismo, Parkinson o cáncer de mama; PMCA3/ATP2B3 con esclerosis múltiple o adenomas productores de aldosterona, y PMCA4/ATP2B4 con hipertensión, Alzheimer, diabetes o cáncer de colon (Hegedús et al. (2020)).

2.3.2.2. Intercambiadores $\text{Na}^+/\text{Ca}^{2+}$

Los NCX son transportadores que intercambian Ca^{2+} por Na^+ en función del potencial de membrana y en contra del gradiente electroquímico. De forma general, intercambian 3 Na^+ por cada Ca^{2+} , aunque puede variar entre 4:1 y 1:1 en función de la concentración de cada uno. Para ello, en condiciones fisiológicas, aprovechan el gradiente de Na^+ para obtener la energía necesaria para bombear Ca^{2+} al medio extracelular. No obstante, es capaz de funcionar en modo inverso. Estos, a diferencia de las PMCA, tienen baja afinidad por Ca^{2+} (diez veces inferior a PMCA) y alta capacidad, por lo que son capaces de responder al comienzo del proceso de recuperación para eliminar rápidamente grandes cantidades de Ca^{2+} (M. J. Berridge (2012), G.-C. Sánchez Julio César (2016)).

Se han identificado tres isoformas (NCX1-3) que, en mamíferos, están codificadas por los genes *SLC8A1*, *SLC8A2* y *SLC8A3*, respectivamente. Estos, en un principio, se postuló que su estructura secundaria constaba de 9 hélices transmembrana; no obstante, en estudios recientes se ha observado que podrían ser 10 (Liao et al. (2012)). Además, contienen 4 sitios de unión en su centro: uno para Ca^{2+} y 3 para Na^+ . También presentan un lazo intracelular entre los dominios 5 y 6 que actúa como regulador de la actividad del canal a través de la acción sobre este de: Ca^{2+} , que se une a los dominios CBD1 y CBD2 para activarlo; Na^+ , que activa a NCX rápidamente para, a continuación, inactivarlo; pH; óxido nítrico; PIP2; fosfocreatina; ATP, entre otros. También puede ser inhibido por amiloride o benzamil (M. J. Berridge (2012), G.-C. Sánchez Julio César (2016)).

De los tres, el NCX1 muestra una expresión a lo largo de todos los tipos celulares en mamíferos, mientras que NCX2 se expresa en el cerebro, músculo estriado y miofibroblastos humanos, y NCX3, a parte de la expresión que muestra NCX2, también se expresa en macrófagos y monocitos humanos. Estos transportadores pueden regular la fuerza contráctil del corazón, la memoria a largo plazo, proliferación o apoptosis, entre otros (G.-C. Sánchez Julio César (2016)). Se ha encontrado evidencia de la implicación de estos NCX en diferentes patologías del sistema nervioso tales como enfermedad de Alzheimer, Parkinson, esclerosis múltiple o epilepsia, si bien es cierto que no se conoce bien el mecanismo de implicación de estos en tales enfermedades. También se ha encontrado relación de estos en otras enfermedades cardíacas como arritmias, fallo cardíaco o infarto. En este sentido, se ha barajado la posibilidad de emplear los NCX como dianas terapéuticas en neuroprotección frente a enfermedades cerebrovasculares (G.-C. Sánchez Julio César (2016), Bode and O'Halloran (2018)).

Otro miembro de la superfamilia de los NCX es el $\text{Na}^+/\text{Ca}^{2+}/\text{K}^+$ o NCKX. En humanos, estos están codificados por cinco genes SLC24 diferentes, como por ejemplo el NCKX1, localizado en la retina, o el NCKX2, en el cerebro y, en menor cantidad, en la retina y otros tejidos. Estos, de forma similar a los NCX, emplean el gradiente electroquímico de entrada de Na^+ y de salida de K^+ para sacar Ca^{2+} al medio extracelular en relación 4:1:1 (G.-C. Sánchez Julio César (2016), Bode and O'Halloran (2018)).

2.4. El retículo endoplasmático (RE)

El Ca^{2+} puede ser almacenado en diferentes compartimentos, tales como el RE, el Golgi o los lisosomas. En consecuencia, estos son capaces de actuar como depósitos de este catión y contribuir al mantenimiento de su baja concentración citosólica. Además, también es capaz de liberar Ca^{2+} en respuesta a diferentes estímulos y participar en la señalización de Ca^{2+} . El RE, presente en todas las células eucariotas, es el depósito de Ca^{2+} de mayor tamaño. Es un sistema membranoso formado por una red de sáculos aplanados y túbulos ramificados que se extienden por la célula hasta tal punto que la membrana de este orgánulo supone la mitad de la membrana total de la célula. En su interior se almacena gran parte del Ca^{2+} intracelular, cuya $[\text{Ca}^{2+}]_{\text{RE}}$ libre es similar a la del medio extracelular, si bien es cierto que su distribución no es homogénea. En concreto, alrededor de la mitad del RE muestra una $[\text{Ca}^{2+}]_{\text{RE}} \leq 1 \text{ mM}$ mientras que en otras regiones puede alcanzar hasta 2 mM (Fuente et al. (2013), Alberts et al. (2010)).

El RE participa en la regulación de varias funciones fisiológicas como en la síntesis de las proteínas transmembrana y lípidos de la mayoría de orgánulos, que tiene lugar en su membrana. Además, también participa en otros procesos como el estrés del RE o la apoptosis, donde el Ca^{2+} tiene un papel clave. Por lo tanto, alteraciones en la homeostasis del Ca^{2+} del RE podrían favorecer el desarrollo de distintas patologías. En cuanto al transporte de Ca^{2+} entre el RE y el citosol, existen dos posibles mecanismos: liberación de Ca^{2+} de los depósitos para potenciar la señal del Ca^{2+} (ON) y captación el Ca^{2+} del citosol (OFF) (Morris et al. (2021), M. Berridge (2014)).

2.4.1. Mecanismos *ON* del RE

Como ya se ha comentado anteriormente, diferentes estímulos extracelulares pueden actuar sobre ciertos receptores de la superficie celular capaces de activar diferentes mecanismos de transporte de Ca^{2+} . Entre estos receptores, están los GPCR y los TKR, que, tras ser activados, inducen una cascada de señalización en la que se generan diferentes señales, como el IP3 o entrada de Ca^{2+} desde el medio extracelular, capaces aumentar la $[\text{Ca}^{2+}]_{\text{cit}}$ a través del Ca^{2+} almacenado en el RE. En el RE existen tres vías principales de liberación de Ca^{2+} al citosol según el canal a través del que lo haga: los *receptores de IP3*, que se activan por la señal movilizador de Ca^{2+} IP3; los *receptores de rianodina*, que liberan Ca^{2+} que se activan por un aumento del Ca^{2+} citosólico, y los *canales leak*, que liberan Ca^{2+} al citosol de forma constitutiva.

2.4.1.1. Receptores de inositol trifosfato (IP3R)

En la membrana del RE se localizan los canales conocidos como *receptores de inositol trifosfato* (IP3R) los cuales, tras ser activados por la unión de IP3, se abren y permiten la liberación de Ca^{2+} desde el RE al citosol. La molécula de IP3 se genera como consecuencia de la activación de GPCR o TKR. Como resultado de esta activación, se activa la *fosfolipasa C* (PLC), enzima capaz de generar IP3 y DAG a partir de la hidrólisis del lípido de membrana *fosfatidilinositol-4,5-bifosfato* (IP2) (figura 2.7) (S. W. Li, Westwick, and Poll (2002), Alberts et al. (2010), M. Berridge (2014)).

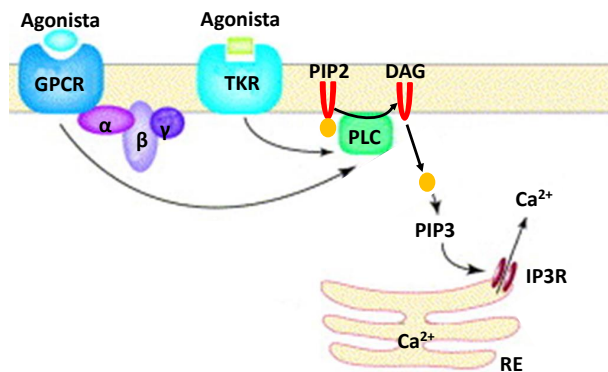


Figura 2.7: Activación de los canales IP3R. Tras la activación de ciertos receptores GPCR o TKR a través de un estímulo externo, se activa la PLC la cual hidroliza a PIP2 para dar DAG e IP3. Este último activa los canales IP3R y libera Ca^{2+} desde el RE. Se han descrito 3 diferentes isoformas de IP3R codificadas por tres genes diferentes cada una de las cuales tiene diferentes afinidades por Ca^{2+} (modificado de Li, Westwick, and Poll (2002)).

Se han descrito 3 isoformas para el IP3R, IP3R1-3, cada una de ellas codificada por un gen diferente: *ITPR1-3*, respectivamente. Además, también se han identificado diferentes variantes de *splicing*. Cada una de estas tres isoformas se diferencian en su afinidad por Ca^{2+} , siendo el IP3R2 la de mayor afinidad e IP3R3 la de menor -10 veces menor-. En cuanto a su localización, las tres isoformas están distribuidas a lo largo de multitud de tejidos (Iwai et al. (2007), M. Berridge (2014)).

Las tres isoformas presentan una estructura común y forman canales tanto homo como heterotetraméricos en los que cada monómero consta de seis TM, siendo los TM 5 y 6 los formadores del poro. En cuanto a los extremos N y C-terminal, ambos se localizan en la cara citosólica (S. Zhang et al. (2011)). Por un lado, el extremo N-terminal, unido al TM1, contiene el sitio de unión a IP3 y, además, también contiene un dominio modulador de la respuesta de liberación de Ca^{2+} . Por otro lado, el C-terminal, que está unido al TM6, presenta un menor tamaño (Baker, Fan, and Serysheva (2017), M. Berridge (2014)).

Existen diferentes moléculas capaces de regular la actividad de los IP3R. Entre las que actúan como agonistas destacan el ATP, especies reactivas de oxígeno (ROS) o el propio Ca^{2+} citosólico y luminal. No obstante, el Ca^{2+} puede actuar como estimulador de la liberación de Ca^{2+} a bajas concentraciones citosólicas, por lo que se activan por el propio Ca^{2+} por un mecanismo conocido como *liberación de Ca^{2+} inducida por Ca^{2+} (CICR)*. No obstante, este catión también puede actuar como inhibidor si la concentración es superior a 300 nM. Otras moléculas capaces de actuar como antagonistas son la cafeína, la calmodulina o la heparina, entre otros. En cuanto a los moduladores, les hay que influyen positivamente sobre su actividad, tales como su fosforilación por PKA (activada por AMPc), o negativamente, como la fosforilación por PKB (M. Berridge (2014)). Además, existe un tipo de señalización por Ca^{2+} muy interesante debido a las fluctuaciones de su $[\text{Ca}^{2+}]_{\text{cit}}$ capaz de controlar diferentes procesos, y en el cual los IP3R tienen un papel clave. De hecho, diferentes combinaciones de expresión de las distintas isoformas de IP3R pueden dar lugar a patrones oscilatorios muy diferentes (S. Zhang et al. (2011)).

En este sentido, la isoforma IP3R2 induce prácticamente una señalización oscilatoria, mientras que IP3R1 puede inducir tanto un patrón oscilatorio como transitorio. Por el contrario, IP3R3 muestra un comportamiento antioscilar, si bien es cierto que en función del tipo celular y del estímulo puede mostrar un pequeño comportamiento oscilatorio. En la figura 2.8 se muestra un resumen del efecto regulatorio de diferentes moléculas y cationes sobre cada una de las tres isoformas. Por ejemplo, a parte del efecto del propio Ca^{2+} sobre estos canales, anteriormente comentado, se tiene que las isoformas IP3R1 e IP3R2 son las más sensibles a IP3, lo que contribuye al comportamiento oscilatorio de su señalización. Además, tanto Ca^{2+} como IP3 pueden tener un efecto sinérgico en la regulación de estos canales, pues, por ejemplo, Ca^{2+} puede alterar la afinidad de estos receptores por IP3, su probabilidad de apertura o la acción sobre estos de otras proteínas asociadas a IP3 como CaM. En particular, CaMKII está involucrada en la regulación dependiente de Ca^{2+} de liberación de Ca^{2+} a través de IP3R inducida por IP3. Por ejemplo, la fosforilación de IP3R2 por CaMKII es capaz de disminuir su probabilidad de apertura. Otro mecanismo de regulación por fosforilación de estos canales es a través de la PKC, la cual ejerce un efecto regulador positivo sobre IP3R1 y negativo sobre IP3R2 e IP3R3, de tal manera que esta capacidad de regulación de PKC puede ser inhibida por Ca^{2+} y CaM.

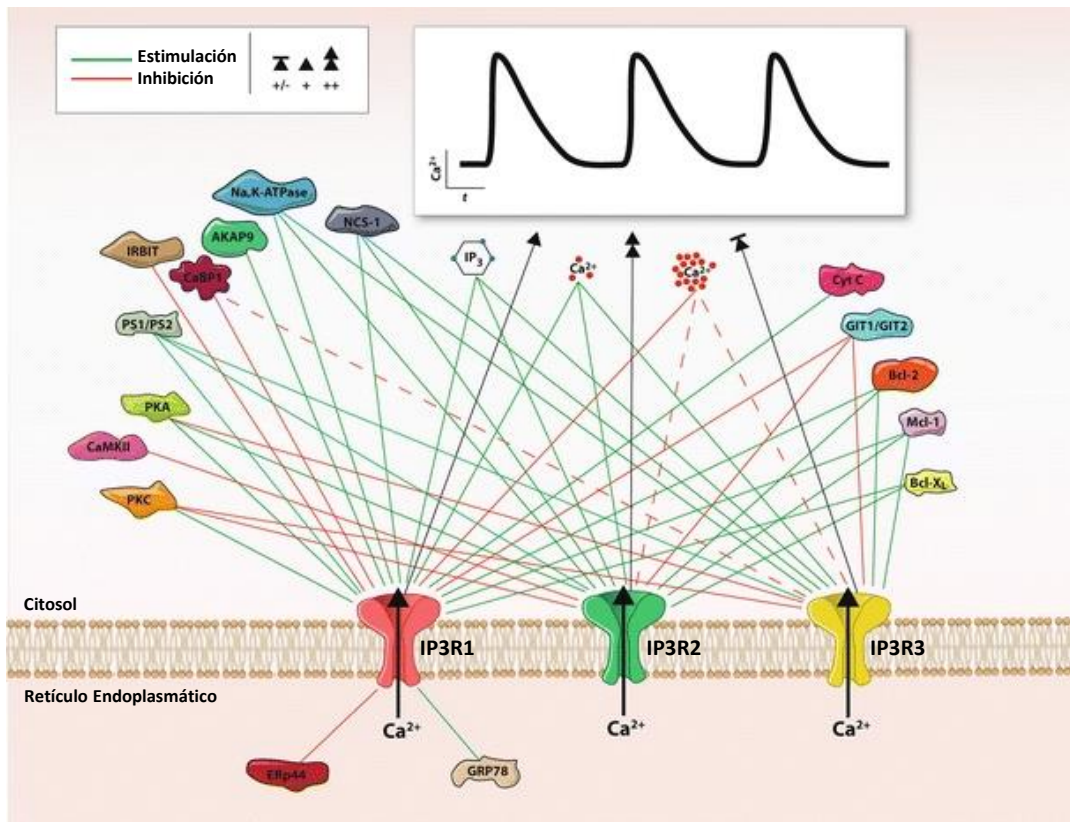


Figura 2.8: Regulación de los canales IP3R. Esquema de la regulación de las diferentes isoformas de IP3R a través de diferentes proteínas tales como proteínas quinasa, proteínas apoptóticas, moléculas como el IP₃ o el catión Ca²⁺ (modificado de S. Zhang et al. (2011)).

La liberación de Ca²⁺ a través de los IP3R desempeña un papel clave en la apoptosis, ya que parte de este Ca²⁺ liberado es captado por mitocondria siendo un paso desencadenante de multitud de señales apoptóticas (Morris et al. (2021)). Además, también se ha descrito que alteraciones en estos canales dan lugar a diferentes patologías, como por ejemplo la ataxia espinocerebelar tipo 15 provocada por mutaciones en el gen que codifica para IP3R1, o el patrón de expresión de IP3R1 y su relación con la enfermedad de Alzheimer (M. Berridge (2014)).

2.4.1.2. Receptores de rianodina (RyR)

Los RyR deben su nombre a su gran afinidad por el alcaloide rianodina. Este alcaloide presenta un papel dual en función de su concentración, pues mientras que a bajas concentraciones es capaz de estimular la liberación de Ca^{2+} desde el RE a través de los RyR, a altas la inhibe. Se localizan casi de forma exclusiva en el RE, aunque también se han encontrado en vesículas secretoras de insulina (M. Berridge (2014), Meissner (2017))

Existen tres isoformas distintas, RyR1-3, codificadas por tres genes distintos, *RYR1-3*, respectivamente. Aunque se expresan en multitud de tejidos, RyR1 lo hace de forma predominante en el músculo esquelético, aunque también en algunas neuronas; RyR2 en el músculo cardíaco, liso, algunas neuronas y células cromafines adrenales, y RyR3 está ampliamente distribuido por diferentes tejidos como el músculo y en neuronas. Estos RyR presentan una gran homología estructural con los IP3R. Además, entre ellas existe una identidad de secuencia del 65 %. Al igual que estos, los RyR son tetrámeros, donde cada uno de ellos consta de 5 o 6 hélices- α , con un extremo C-terminal citosólico que constituye el canal. También en su cara citosólica se encuentra el extremo N-terminal el cual contiene un dominio regulador al que se pueden unir diferentes moléculas para modular la actividad de este canal (M. Berridge (2014)).

Números agonistas son capaces de activar los RyRs como la cafeína, el ATP, la ADP ribosa o el óxido nítrico. También existen otros tipos de activadores de los RyR, como por ejemplo los receptores de dihidropiridinas, CaM, PKA o ciertas FKPB como la FKBP12 (Raymond and Redman (2006), M. Berridge (2014)). No obstante, uno de los moduladores fisiológicos más importantes es el propio Ca^{2+} a través de un mecanismo conocido como *Liberación de Ca^{2+} Inducida por Ca^{2+}* (CICR, del inglés *Ca^{2+} induced Ca^{2+} release*). Este mecanismo consiste en la amplificación de la señal de Ca^{2+} que se genera tras un aumento del Ca^{2+} citosólico provocado tanto por la entrada de Ca^{2+} a través de canales de la MP como por la liberación de Ca^{2+} desde los depósitos a través de los IP3R. En consecuencia, este Ca^{2+} citosólico induce la activación de RyR que, a su vez, estimula la liberación de más Ca^{2+} desde al citosol a través de los RyR. Entre los diferentes mecanismos que pueden llevar a cabo este incremento de Ca^{2+} inicial para activar CICR están, por ejemplo, los VOCC en las células musculares y en las neuronas (M. Berridge (2014)).

La principal función de estos, sobretodo de RyR1, es el acoplamiento excitación-contracción en el músculo esquelético que, a su vez, depende de Cav1.1. Mutaciones en los canales de RyR pueden dar lugar a diferentes patologías. Entre ellas, se han observado mutaciones en *RYR1* que pueden causar hipertermia maligna, o mutaciones en *RYR2* que dan lugar a taquicardia ventricular polimórfica relacionada con catecolaminas (M. Berridge (2014)).

2.4.2. Mecanismos *OFF* del RE

Al igual que hay mecanismos que aumentan la $[Ca^{2+}]_{cit}$ a favor del gradiente electroquímico a través del vaciamiento de los depósitos, también existen aquellos que captan el Ca^{2+} citosólico y lo introducen nuevamente en los depósitos. Por un lado, en el RE están las bombas *ATPasa de Ca^{2+} del retículo sarco/endoplasmático (SERCA)*, del inglés *sarco/endoplasmic reticulum Ca^{2+} /ATPasa*). Por otro lado, en el aparato de Golgi se expresan las bombas *ATPasa de Ca^{2+} de la vía secretora (SPCA)*, del inglés *secretory pathway Ca^{2+} /ATPasa*).

2.4.2.1. ATPasas de Ca^{2+} del retículo sarco/endoplasmático (SERCA)

De la misma forma que la PMCA, la SERCA pertenece al grupo de las ATPasas de tipo P y, en particular, a la subfamilia P_{2A} . Se encarga de la recaptación de Ca^{2+} desde el citosol al lumen del RE en contra de gradiente electroquímico. Para ello, a partir de la hidrólisis de una molécula de ATP son capaces de transportar dos iones de Ca^{2+} al RE. Sin embargo su función no solo consiste en rellenar los depósitos tras su vaciamiento, sino que, además, son capaces de mantener los niveles basales de Ca^{2+} contrarrestando la salida pasiva de Ca^{2+} a través de los canales *leak*. Esto es de suma importancia para evitar, por ejemplo, que se induzcan los mecanismos de estrés del RE (Themistocleous et al. (2021)).

Se han identificado tres miembros de esta familia (SERCA1-3), y un total de 11 formas de *splicing* alternativo. La expresión de las isoformas es dependiente del tipo celular y, además, la respuesta frente a distintas drogas es diferente a pesar de que entre ellas hay un 75 % de homología. Para SERCA1 hay dos transcritos: 1a, que se expresa principalmente en el músculo esquelético adulto, y 1b, presente en el músculo esquelético fetal. En cuanto a la SERCA2, se han descrito también dos transcritos: 2a, que se encuentra en el músculo esquelético y cardíaco adulto, y 2b, que se expresa en estos mismos músculos tanto en adultos como en fetos. Finalmente, la SERCA3 presenta 3 isoformas, todas ellas expresadas en células no musculares (Themistocleous et al. (2021)).

Las tres isoformas están compuestas por 10 TM consistentes en 10 hélices α . Los dominios citosólicos son los responsables de la unión e hidrólisis del ATP. Los TM se encargan de unir y transportar el Ca^{2+} . Al igual que en las PMCA, estas también se pueden dividir en tres partes típicas de las ATPasa tipo P que se localizan en la cara citosólica: el dominio de unión a nucleótidos, formado por la región intermedia del lazo que une los TM 4 y 5; el dominio de fosforilación, formada por las regiones más cercanas a los TM de este mismo lazo, y el dominio de actuación, formado por el extremo N-terminal y el lazo entre los TM 2 y 3 y que cataliza la defosforilación del dominio de fosforilación (Almasi and El Hiani (2020), Themistocleous et al. (2021)).

El bloqueo o inhibición de las bombas SERCA provoca un vaciamiento lento del Ca^{2+} del RE. Entre los inhibidores más utilizados destaca la tapsigargina (inhibidor irreversible), o el ácido ciclopiazónico (CPA) y la 2.5-di(terbutil)-1.4-benzohidroquinona (tBHQ), ambos inhibidores reversibles (Takemura et al. (1989), Themistocleous et al. (2021)).

Alteraciones en SERCA se han relacionado con diferentes patologías. Por ejemplo, puede dar lugar a hipertrofia o fallo cardíaco (Kranias and Bers (2007)). También se ha relacionado con el cáncer. Por ejemplo, la SERCA3 disminuye en los cáncer de mama, colon y pulmón ya desde estadios tempranos de la enfermedad. Por ello, el uso de inhibidores como la tapsigargina en estos tumores no tendría efecto alguno. No obstante, sí que se ha observado que en células de cáncer hepático tratadas con tapsigargina se promueve la apoptosis de estas (Almasi and El Hiani (2020), Themistocleous et al. (2021)).

2.4.2.2. ATPasas de Ca^{2+} de la vía secretora (SPCA)

Las SPCA son bombas que se localizan en todas las regiones del aparato de Golgi. Al igual que la SERCA, la SPCA pertenece al grupo de las ATPasas de tipo P y, en particular, a la subfamilia $\text{P}_{2\text{A}}$. Su principal diferencia es que transporta un solo ión Ca^{2+} por cada molécula de ATP hidrolizada y que, además, transporta Mn^{2+} . Esto se debe a que en su TM contiene solo un sitio de unión a Ca^{2+} con ligeras diferencias que le confieren la capacidad de transportar también Mn^{2+} . Por lo demás, su estructura es prácticamente igual a la de la SERCA en cuanto a dominios se refiere (W. He and Hu (2012)).

En humanos, las SPCA1-2 están codificadas por dos genes: *ATP2C1-2*, respectivamente. Además, también se han identificado 4 variantes de *splicing* para la SPCA1: SPCA1a-d (W. He and Hu (2012)). La expresión de la SPCA1 se extiende por todo tipo de tejidos, mientras que la SPCA2 tiene una expresión más localizada: tracto digestivo, tiroides, próstata... En cuanto a su localización celular, se ha identificado en vesículas de secreción donde su principal función es el transporte de Mn^{2+} , cofactor de numerosas enzimas (Missiaen et al. (2007)). Se ha observado que tiene cierto efecto en la regulación del RE. Es más, deficiencias en SPCA1 resultan en una hipersensibilización de las células al estrés del RE. Alguna de las complicaciones que podrían estar causadas por alteraciones en las SPCA son, por ejemplo, la enfermedad de *Hailey-Hailey* como consecuencia de mutaciones en un solo alelo de *ATP2C1*, o el aumento de expresión de SPCA1 en el cáncer de mama (W. He and Hu (2012)). Cabe destacar que la isoforma SPCA2 ha mostrado participar en la entrada de Ca^{2+} desde el medio extracelular a través de ORAI1 en un mecanismo independiente de los depósitos de Ca^{2+} , de los sensores STIM1-2 y desacoplado de la actividad ATPasa de la SPCA2 a través de la interacción entre el extremo N-terminal de esta isoforma con el extremo C-terminal de ORAI1 (M. Feng et al. (2010), M.-Y. Feng and Rao (2013)). Este mecanismo de transporte de Ca^{2+} a través de ORAI1 independiente de depósitos es importante en diferentes procesos fisiológicos, tales como el transporte de Ca^{2+} en tejidos secretores como el páncreas o las glándulas mamarias y, de hecho, su expresión se induce durante la lactancia, su desbalance conlleva al aumento de la proliferación y la tumorigénesis (M.-Y. Feng and Rao (2013)). Esta isoforma se ha relacionado con desórdenes neurológicos (Dang and Rao (2016)).

2.5. La mitocondria

La mitocondria es un orgánulo subcelular que desempeña diferentes funciones esenciales. Aunque la más conocida quizás sea la de producción de ATP y otros intermediarios a través del ciclo de Krebs y de la fosforilación oxidativa, también participa en diferentes mecanismos como la regulación de las especies reactivas de oxígeno (ROS) o la respuesta ante el estrés celular mediante dos mecanismos clave: la autofagia y la apoptosis (se habló de ellos en el apartado “1.2.3 Resistencia a la muerte”). Además, también participa de forma activa en la homeostasis del Ca^{2+} intracelular al ser capaz de intercambiar grandes cantidades de Ca^{2+} con su entorno (Alberts et al. (2010), Rizzuto et al. (2012), Morris et al. (2021)).

Este orgánulo está formado por dos membranas: la membrana mitocondrial externa (MME) y la membrana mitocondrial interna (MMI). Por lo tanto, todas aquellas moléculas, proteínas e iones que deban alcanzar la matriz mitocondrial deberán atravesar ambas. En la MME se localizan una gran cantidad de *canales aniónicos dependientes de voltaje (VDAC)* o porinas que la hacen permeable a iones como el Ca^{2+} y a proteínas pequeñas inferiores a 5 kDa. Por el contrario, la MMI es impermeable a la mayoría de los iones y presenta una extensión superior a la MME debido a las numerosas crestas que posee en su superficie, por lo que puede alojar una gran variedad de componentes. Por un lado, presenta sistemas de transporte como el *uniportador de Ca^{2+} mitocondrial (MCU)* a través del cual entra Ca^{2+} a la matriz mitocondrial, intercambiadores catiónicos que permiten la salida del Ca^{2+} como el *intercambiador de Na^+/Li^+ / Ca^{2+} mitocondrial (NCLX)* o el *intercambiador mitocondrial $\text{H}^+/\text{Ca}^{2+}$ (mHCX)*... Por otro lado, MMI también contiene complejos enzimáticos como los componentes de la cadena transportadora de electrones o las enzimas que participan en la elongación de los ácidos grasos. En cuanto al espacio entre ambas membranas, hay presentes diferentes enzimas como la creatina quinasa y el citocromo c, este último de suma importancia en mecanismos como la apoptosis (Kann and Kovács (2007), Foskett and Philipson (2015)).

Cabe destacar que la mitocondria, a pesar de no expresar bombas, es capaz de captar grandes cantidades de Ca^{2+} . Esto se debe a la gran diferencia de potencial que existe entre el citosol y la matriz mitocondrial, hecho que tiene su origen en la cadena de transporte de electrones, por lo que se dice que la captación de Ca^{2+} por parte de la mitocondria es electrogénica. El transporte de electrones entre los diferentes complejos de la cadena respiratoria genera un flujo de protones hacia el espacio intermembrana que genera un gradiente el cual es aprovechado por la ATPsintetasa para sintetizar ATP gracias a la energía liberada por el paso del H^+ desde el espacio intermembrana hasta la matriz mitocondrial. Este proceso genera una diferencia de potencial de unos -150 a -180 mV en el interior de la mitocondria. Esta diferencia de potencial tan grande en magnitud es la responsable de la entrada de Ca^{2+} a la mitocondria a través del MCU y, por ende, su particular capacidad para intercambiar grandes cantidades de Ca^{2+} con su entorno (Foskett and Philipson (2015), Morris et al. (2021)).

No obstante, puesto que el MCU se localiza en la MMI y que este tiene una baja afinidad por Ca^{2+} , la mitocondria no es capaz de captar este catión directamente desde el citosol, a menos que se generen microdominios de alta $[\text{Ca}^{2+}]_{\text{cit}}$ en el entorno de los canales. Por lo tanto, para poder mantener sus niveles basales de este Ca^{2+} es necesaria una transferencia directa desde otros depósitos. Afortunadamente, estos orgánulos poseen una gran movilidad y pueden sufrir procesos de fusión y fisión, lo que les permite desplazarse hacia ciertas zonas y actuar como barrera para evitar ciertos incrementos de Ca^{2+} en regiones concretas. También pueden desplazarse a regiones donde la demanda energética es grande, como por ejemplo los gránulos de secreción o zonas próximas a la SERCA (Hollenbeck and Saxton (2005), Rizzuto et al. (2012), Foskett and Philipson (2015), Morris et al. (2021)).

En este sentido, la mitocondria es capaz de regular la liberación de Ca^{2+} desde los depósitos y su entrada desde el medio extracelular a través de los SOCC gracias a la existencia de zonas de proximidad o de contacto con estos orgánulos y a la formación de microdominios de Ca^{2+} . Estos microdominios de Ca^{2+} son regiones en las que la concentración de este catión es muy elevada en comparación con el resto del citosol que pueden formarse, por ejemplo, alrededor de un canal IP3R tras haber liberado Ca^{2+} al citosol. En este sentido, pueden darse interacciones físicas entre los canales de Ca^{2+} del RE o los SOCC y proteínas de la membrana de la mitocondria y que esta actúe como tampón. Por ejemplo, entre la mitocondria y el RE se forma unas estructuras denominadas *membranas del RE asociadas a la mitocondria (MAM)* o *contactos RE-mitocondria (MERC)*, con una distancia entre ambos orgánulos de 10-50 nm, de tal manera que alrededor del 20% de la superficie del RE forma parte de estas estructuras. Existen una serie de proteínas de anclaje que las estabilizan, tales como mitofusina2 (Mfn2) o PDZD8. En la interface que se forma tienen lugar diferentes procesos tales como la síntesis e intercambio de lípidos o la formación de autofagosomas que pueden determinar el destino de la célula. Por ejemplo, a través de la *proteína regulada por glucosa 75 (GRP75)* se establece una interacción entre los VDAC y los IP3R de la mitocondria tal que al estimularse el IP3R el Ca^{2+} liberado se dirige tanto al citosol como al interior de la mitocondria a través de los VDAC y previene la inactivación por Ca^{2+} de los IP3R al evitar el aumento localizado de concentración de este catión alrededor de este canal (Morris et al. (2021), Rizzuto et al. (2012)) (figura 2.9).

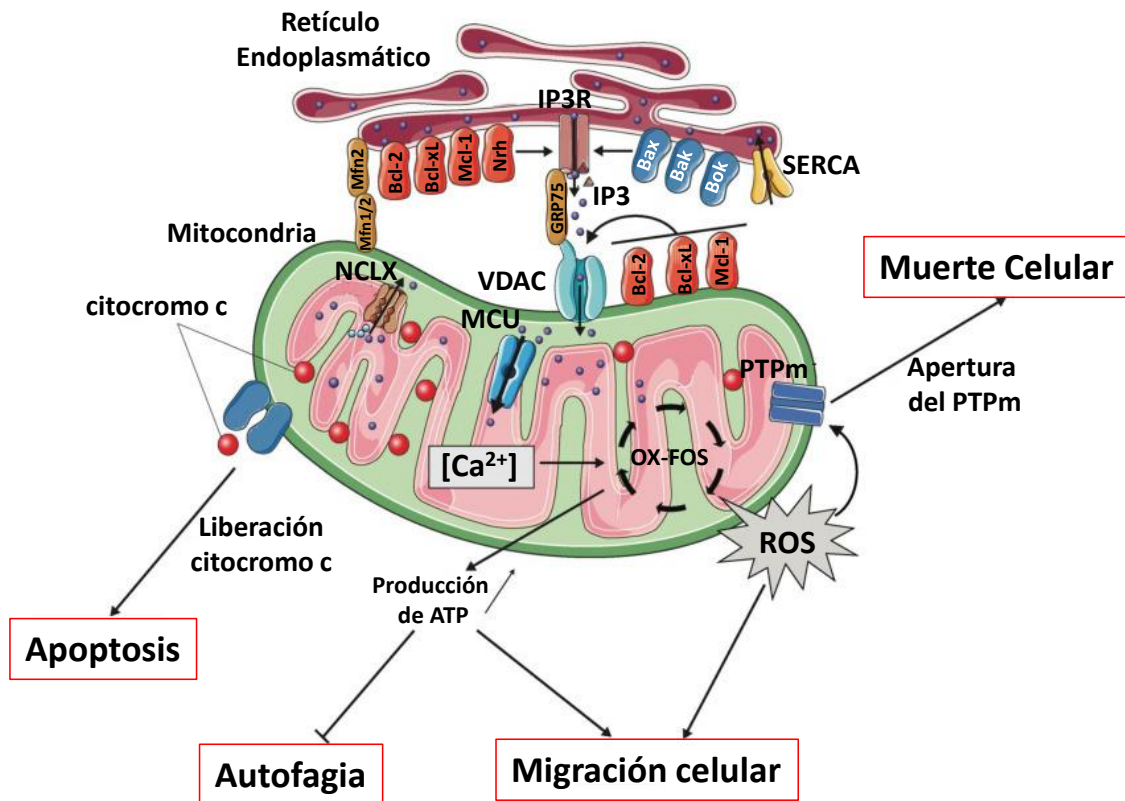


Figura 2.9: Regulación por la mitocondria de la liberación de Ca^{2+} desde el RE. En la interface de las zonas de contacto entre el RE y la mitocondria, estabilizada por proteínas de anclaje como miofusina2 (Mfn2) o PDZD8, las proteínas de la familia Bcl-2 localizadas tanto en la membrana del RE (Bak, Bax, Bok, Bcl-2, Bcl-xL, Mcl-1 y Nrh) como de la mitocondria (Bcl-2, Bcl-xL, Mcl-1) tienen un papel relevante en el transporte de Ca^{2+} desde el RE a la mitocondria a través de la interacción de estas con el IP3R o el VDAC. De esta manera, a partir de la regulación de las acumulaciones mitocondriales de Ca^{2+} , las Bcl-2 controlan procesos tales como la producción de ATP y ROS que repercuten directamente en el destino celular: la apoptosis, supervivencia, autofagia o migración celular. No obstante, también hay proteínas, como GRP75, que actúan como puente entre los IP3R y los VDAC y permiten una interacción entre ellas la cual tiene un impacto directo sobre la prevención de la inhibición dependiente de Ca^{2+} de los IP3R y de la mejora de la captación por la mitocondria de este catión recién liberado por el RE (modificado de Morris et al. (2021)).

En este proceso de regulación también juegan un papel importante las proteínas de la familia Bcl-2 (Bak, Bax, Bok, Bcl-2, Bcl-xL, Mcl-1 y Nrh), tanto pro como antiapoptóticas y localizadas tanto en la membrana del RE como de la mitocondria, pues a través de su interacción con los canales y transportadores de Ca^{2+} localizados en estos orgánulos son capaces de controlar el tráfico. Así, a través de la regulación de las acumulaciones de Ca^{2+} mitocondrial, las Bcl-2 controlan diferentes procesos bioenergéticos como la producción de ATP y ROS y, por ende, en el destino celular: apoptosis, supervivencia, autofagia o migración celular. Por un lado, las proteínas antiapoptóticas Bcl-2, Bcl-xL, Mcl-1 y Nrh pueden interactuar con IP3R para reducir la liberación de Ca^{2+} desde el RE y así evitar que la mitocondria actúe como tampón y lo capte, lo que favorece el mantenimiento de los procesos bioenergéticos mitocondriales y la prevención de la muerte celular. Por otro lado,

las Bcl-2, Bcl-xl, Mcl-1 localizadas en la membrana externa de la mitocondria, en función del tipo celular y del estado metabólico de estas células, pueden promover o inhibir la captación de Ca^{2+} a través de su interacción con VDAC. Por el contrario, las proteínas proapoptóticas Bax y Bak tienen el efecto contrario en el RE, pues promueven la liberación de Ca^{2+} y la muerte celular. Pero todavía hay más proteínas capaces de regular estos procesos, pues el Bok localizado en la membrana del RE también es capaz de regular el número de MERC y de liberar Ca^{2+} a través de su interacción con los IP3R y así favorecer la sobrecarga de Ca^{2+} y la consecuente muerte celular (Morris et al. (2021)). En resumen, las proteínas antiapoptóticas previenen de la apoptosis al reducir la cantidad de Ca^{2+} captada por la mitocondria procedente del RE, ya sea induciendo una menor liberación de este catión o una menor captación y, por el contrario, las proapoptóticas favorecerán la apoptosis al inducir una mayor liberación de Ca^{2+} o una mayor captación (figura 2.9).

Cabe destacar que en la cadena respiratoria tienen lugar diferentes reacciones redox -complejos I y III- que generan ROS. Estos participan en la regulación de procesos tales como la producción de hormona del crecimiento o citoquinas, e incluso en la regulación de transportadores y canales iónicos. No obstante, la producción descontrolada de ROS, conocida como *estrés oxidativo*, conlleva la acumulación de ROS la cual induce el PTPm y la muerte celular por apoptosis (figura 2.9). Por ejemplo, en el sistema nervioso central, la acumulación de ROS podría estar involucrada en la isquemia cerebral, incluso, en enfermedades neurodegenerativas como el Alzheimer y el Parkinson. Además de ROS, la acumulación mitocondrial de Ca^{2+} también puede desencadenar la apoptosis. Esto sucede cuando la captación de Ca^{2+} es tal que se sobrepasa la capacidad tampón de la mitocondria y, en consecuencia, este se acumula en la mitocondria. Esta acumulación induce la apertura del PTPm, poro no selectivo que favorece el colapso del potencial mitocondrial, su inchamiento y la liberación de factores mitocondriales solubles como nucleótidos de pirimidina o citocromo c y, como resultado, se induce la muerte celular. Además, en presencia de ROS la apertura del PTPm se potencia. En este sentido, para contrarrestar la acumulación de Ca^{2+} en la mitocondria existen otros sistemas de transporte que permiten su salida de esta: los NCLX y los mHCx. No obstante, sí que es necesario que el Ca^{2+} esté presente en el interior de la mitocondria, pues participa en diferentes procesos fisiológicos tales como el metabolismo, ya que regula la actividad de tres deshidrogenasas del ciclo de Krebs: piruvato deshidrogenasa, α -cetoglutarato deshidrogenasa e isocitrato deshidrogenasa (Kann and Kovács (2007), Bonora and Pinton (2014), Morris et al. (2021), Rizzuto et al. (2012)).

La mitocondria está implicada en diferentes patologías. En particular, en el cáncer las células modifican la homeostasis del Ca^{2+} mitocondrial a través de cambios en la expresión de diferentes canales y transportadores de Ca^{2+} . De hecho, se ha observado que la expresión del MCU está aumentada en algunos tipos de cáncer. En consecuencia, un incremento en la $[\text{Ca}^{2+}]_{\text{mit}}$ conlleva al aumento de ROS mitocondrial que, a su vez, estimula una la señalización redox protumoral capaz de activar otras vías relacionadas con la proliferación celular, cambios metabólicos y adaptaciones celulares al estrés (Delierneux et al. (2020)). También se ha relacionado que el aumento de ROS mitocondrial puede contribuir a otras patologías tales como enfermedades neurodegenerativas, cardiovasculares o diabetes (Woods et al. (2019), Delierneux et al. (2020)).

En resumen, el transporte de Ca^{2+} mitocondrial comienza con el paso de este a través de la MME gracias a los VDAC el cual, una vez alcanza el espacio intermembrana, atraviesa la MMI a través del MCU, todo ello favorecido por la gran diferencia de potencial entre la matriz mitocondrial y el citosol. En cuanto a su salida de la mitocondria, este puede ser transportado por los NCLX y por los mHCX. No obstante, en aquellas situaciones en las que se forma el PTPm, como en la apoptosis, el Ca^{2+} también es capaz de salir al citosol a través de él. A continuación, se describen brevemente cada uno de ellos.

2.5.1. El complejo uniportador de Ca^{2+} mitocondrial (MCU)

El complejo MCU está compuesto por diferentes dominios funcionales de los cuales el MCU es la subunidad central y formadora del poro. La subunidad MCU es una proteína de 351 residuos y 40 kDa localizada en la MMI cuya afinidad por Ca^{2+} es baja, con un $K_d \approx 0.01 \mu\text{M}$, por lo que sólo es posible que este complejo capte Ca^{2+} de microdominios, puesto que en condiciones fisiológicas la $[\text{Ca}^{2+}]_{\text{cit}} < 0.1 \mu\text{M}$. La captación de este catión está favorecida por el gran potencial de membrana que se genera en la cadena respiratoria por el bombeo de protones hacia el espacio intermembrana de la mitocondria. En cuanto a su inhibición, pueden emplearse rojo de rutenio, rutenio 360, gadolinio y lantánidos (Rizzuto et al. (2012), Foskett and Philipson (2015), Patron et al. (2019), Delierneux et al. (2020), Morris et al. (2021)).

En los humanos, la subunidad MCU está expresada a bajos niveles de forma ubicua a lo largo de todos los tejidos, si bien es cierto que su actividad varía entre ellos. No obstante, no lo hace de forma homogénea, pues se ha encontrado un patrón de expresión específico de tejido: ocho veces superior en el cerebelo y en el músculo esquelético respecto del corazón. Está compuesto por dos hélices transmembrana conectadas por un pequeño lazo el cual contiene varios residuos ácidos, lo que se conoce como motivo DIME, localizados sobre la hélice del TM2. En cuanto a sus extremos, todo apunta a que se localizan en la matriz mitocondrial, lo que supone que el lazo lo hace en el espacio intermembrana. Este es un oligómero cuya estequiometría todavía no ha sido resuelta (Foskett and Philipson (2015), Woods et al. (2019), Delierneux et al. (2020)).

Aparte de la subunidad MCU, el complejo MCU contiene numerosas subunidades que regulan la captación de Ca^{2+} : MCUB, EMRE, MICU1, MICU2, MICU3 y MCUR1. Cabe destacar que, a excepción de MCUB, la expresión de las subunidades reguladoras es superior a la del MCU (Patron et al. (2019)). A continuación, se describirán brevemente cada una de ellas:

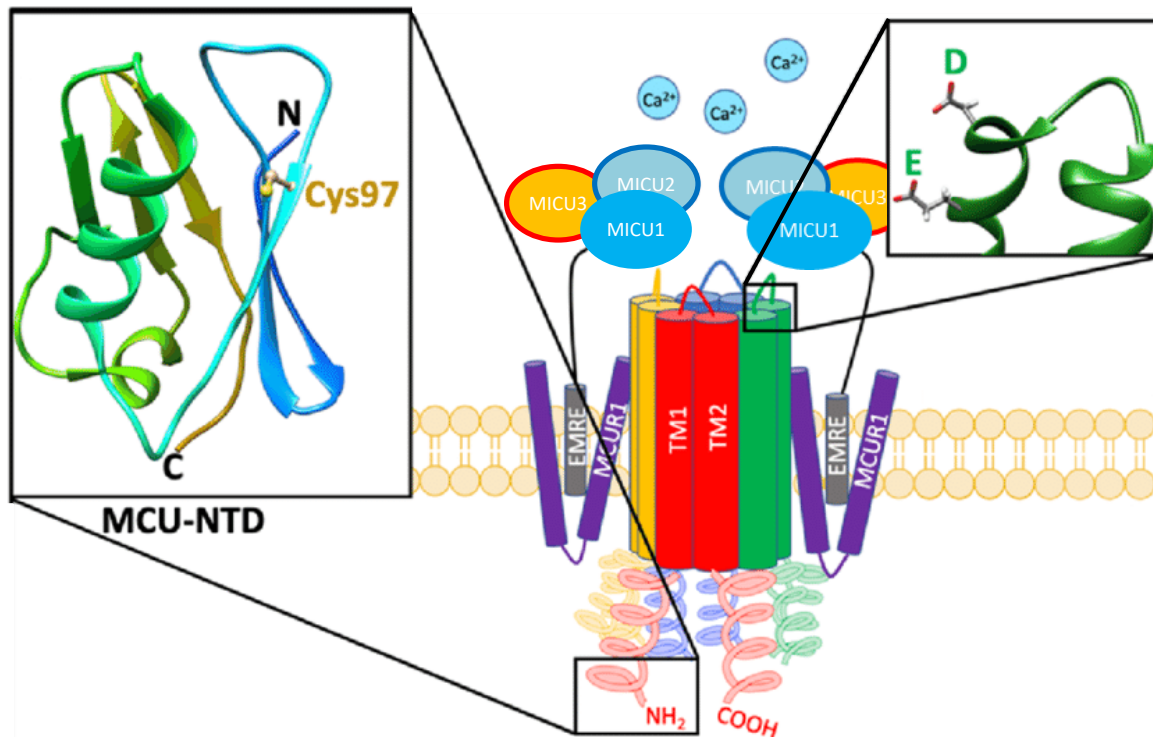


Figura 2.10: Complejo uniportador mitocondrial (MCU). El complejo MCU está localizado en la membrana mitocondrial interna. La subunidad principal de este complejo es, precisamente, MCU, que es la formadora del poro. Además, contiene diferentes subunidades reguladoras: MCUB, EMRE, MICU1, MICU2, MICU3 y MCUR1. En el recuadro de la izquierda se muestra la localización de la Cys97, que participa en la regulación positiva por ROS, y en el de la derecha se muestra el motivo DXXE, que contiene el aspartato sobre el cual actúa el rutenio 360 y bloquea MCU (modificado de Woods et al. (2019)).

- *MCUB*: es un parálogo de MCU con una identidad de secuencia entre ambos del 50 %, puede oligomerizar tanto consigo mismo, dando lugar a homooligómeros probablemente no funcionales, como con MCU con una estequiometría no conocida. Se ha encontrado que, aunque se expresa de forma ubicua, su mayor expresión se da en el tejido adiposo, sangre y bazo. En este sentido, MCUB ejerce un efecto negativo sobre MCU lo que conlleva una menor captación de Ca^{2+} . No obstante, el patrón de expresión observado para ambos es diferente, por lo que la expresión relativa entre ambas puede dar lugar a diferentes perfiles de permeabilidad mitocondrial por Ca^{2+} (Foskett and Philipson (2015), Patron et al. (2019)).
- *Regulador esencial de MCU: (EMRE, del inglés Essential MCU Regulator)* que, como su propio nombre indica, es el regulador esencial del complejo MCU. Es una proteína de 10 kDa con un único TM y un extremo C-terminal altamente ácido. En los mamíferos está expresado a lo largo de todos los tejidos. Una de las razones de su importancia en este complejo que hace de puente entre MCU y otros reguladores, como MICU1 y MICU2 (Foskett and Philipson (2015)).

- *MICU1*: subunidad del complejo MCU dimérica que tiene un papel dual en la regulación de la captación de Ca^{2+} . Cabe destacar que MICU1 no interacciona directamente con la subunidad MCU, sino que lo hace a través de EMRE. En cuanto al sistema de detección de Ca^{2+} , MICU1 contiene dos manos EF. Su efecto inhibitorio se observa para unas $[\text{Ca}^{2+}]_{\text{cit}} < 3 \mu\text{M}$, por lo que a concentraciones mayores esta se disocia de MCU y el poro se abre para permitir la entrada de Ca^{2+} . Por lo tanto, en condiciones basales, en las que $[\text{Ca}^{2+}]_{\text{cit}} < 0.1 \mu\text{M}$, MICU1 ejerce su efecto regulatorio sobre MCU. Se ha observado que mutaciones en esta subunidad en pacientes da lugar a desórdenes cerebrales y musculares a causa de una elevada concentración mitocondrial de Ca^{2+} ($[\text{Ca}^{2+}]_{\text{mit}}$) al ser incapaces de inhibir la captación de Ca^{2+} inducida por agonista (Foskett and Philipson (2015), Woods et al. (2019), Patron et al. (2019)).
- *MICU2*: subunidad reguladora la cual tiene un 27 % de identidad de secuencia con MICU1 y, al igual que esta, también contiene dos dominios de manos EF. Esta regula negativamente la captación de Ca^{2+} no de forma directa sobre la subunidad MCU, sino a través de la interacción con MICU1 mediante un puente disulfuro. Además, su perfil de expresión depende de la expresión de MICU1. Cabe destacar que la expresión de MICU2 es necesaria para el funcionamiento de MCU para poder inhibir la captación de Ca^{2+} a grandes concentraciones. En cuanto a su efecto sobre MCU favorece la captación de Ca^{2+} (Foskett and Philipson (2015), Woods et al. (2019), Patron et al. (2019)).
- *MICU3*: esta es una subunidad de 55 kDa con dos dominios manos EF. Tiene un 34 % de similaridad de secuencia con MICU1 y un 47 % con MICU2. La expresión de esta es elevada, sobretodo en el tejido cerebral. Se ha observado que, al igual que MICU2, forma un puente disulfuro con MICU1 para poder regular la actividad de MCU. En particular, y a diferencia de MICU2, MICU3 tiene un efecto positivo sobre la captación de Ca^{2+} (Patron et al. (2019)).
- *MCUR1*: subunidad reguladora de 40 kDa con dos hélices transmembrana localizada en la MMI. Sus extremos se dirigen al espacio intermembrana y, por ende, el lazo que las une está situado en la matriz mitocondrial. Esta subunidad reguladora puede interactuar directamente con MCU, pero no con MICU1, y actúa como regulador positivo. Está expresado ampliamente en todos los tejidos (Foskett and Philipson (2015), Woods et al. (2019)).

El complejo MCU también es regulado por cationes bivalentes como Mg^{2+} o el propio Ca^{2+} , de tal manera que se unen a una región del extremo N-terminal de MCU denominada *parche ácido regulador de MCU* (*MRAP*, del inglés, *MCU-regulating acidic patch*) y desestabiliza la estructura del oligómero MCU que forma el poro hacia monómeros (Woods et al. (2019)). También se ha encontrado que las ROS son capaces de regular directa y positivamente la actividad de MCU a través de una S-glutinationulación en su Cys-97, por ejemplo, en respuesta a señales proinflamatorias como el lipopolisacárido o la hipoxia derivada de ROS, y, además, esta regulación es independiente del resto de proteínas del complejo MCU. En consecuencia, se consigue una gran captación de Ca^{2+} y un aumento aun mayor de la producción de ROS (Delierneux et al. (2020)).

2.5.2. Canales aniónicos dependientes de voltaje (VDAC)

Los VDAC son las proteínas más abundantes en la MME, de las cuales se han descrito tres isoformas, VDAC1-3. De estas, sólo se ha resuelto la estructura de la primera, aunque se piensa que las tres comparten la misma estructura general: barril C-terminal de 19 hebras que forma un poro en la MME capaz de transportar metabolitos y una hélice N-terminal localizada dentro del poro. Las tres isoformas muestran una expresión elevada en el corazón, músculo esquelético y cerebro. Están implicadas en diferentes procesos fisiológicos mediados por la mitocondria como la regulación bioenergética, el flujo intracelular de energía o la apoptosis. Por un lado, controlan el flujo de Ca^{2+} y otros metabolitos (inferior a 5 kDa) a través de la MME, como por ejemplo la entrada de sustratos respiratorios como el ADP o el Pi y la salida de ATP (Fang and Maldonado (2018), Delierneux et al. (2020), Varughese, Buchanan, and Pitt (2021)).

Así, en relación al Ca^{2+} , estos canales permiten a este orgánulo actuar como tampón de Ca^{2+} , por ejemplo, en las zonas de contacto con la MP o con el RE. Por ejemplo, a través de la proteína GRP75 se establece una unión entre el VDAC1 y el IP3R que favorece este mecanismo. Por otro lado, también se ha descrito su relevancia en la regulación de la apoptosis. En este último caso, tras el estímulo apoptótico, VDAC1 oligomeriza e interactúa con proteínas proapoptóticas, regula la formación del PTPm y, como resultado, se libera citocromo c, se forma el apoptosoma y, en consecuencia, se activa la fase de ejecución de la apoptosis (Fang and Maldonado (2018), Delierneux et al. (2020), Varughese, Buchanan, and Pitt (2021)).

Existen diferentes mecanismos de regulación de la apertura y cierre de estos canales. Por un lado, pueden abrirse a través de la interacción con proteínas como la tubulina o la *hexoquinasa (HK)*. También pueden abrirse a través de modificaciones postraduccionales, como por ejemplo la fosforilación por la PKA o la GSK3 β (Fang and Maldonado (2018)).

2.5.3. Intercambiadores de Ca^{2+} mitocondriales

Como se ha visto anteriormente, la mitocondria juega un papel fundamental como mecanismo tampón de Ca^{2+} . No obstante, la concentración de este catión en la matriz mitocondrial es similar a la del citosol. No obstante, debido a su elevada diferencia de potencial, esta concentración podría llegar a $[\text{Ca}^{2+}]_{\text{mit}} = 0.1 \text{ M}$. Por lo tanto, es evidente que han de existir mecanismos que contrarresten la entrada de Ca^{2+} a la mitocondria, evitar la acumulación de este en su interior y mantener la concentración basal tan baja de su interior. En particular, se han descrito dos sistemas que expulsan Ca^{2+} de la mitocondria en contra del gradiente electroquímico: el NCLX y el mHCX.

El NCLX, codificado por el gen *SLC8B1*, expulsa un Ca^{2+} a cambio de introducir 3 Na^+ . Además, a diferencia de los otros miembros de la superfamilia de los intercambiadores $\text{Na}^+/\text{Ca}^{2+}$, el NCLX es el único capaz de transportar Li^+ . Se ha encontrado una mayor expresión de este en los tejidos excitables (Palty et al. (2010), Boyman et al. (2013), G.-C. Sánchez Julio César (2016)). Cabe destacar que se ha encontrado una relación entre la disminución en la expresión de NCLX y la capacidad metastásica de las células de cáncer de colon (Pathak et al. (2020)).

El otro sistema es el mHCX, a diferencia del NCLX, se expresa de forma mayoritaria en los tejidos no excitables y no tiene apenas relevancia en la extrusión de Ca^{2+} . Este ha sido identificado como la proteína LETM1. En cuanto a su funcionamiento, es capaz de expulsar un Ca^{2+} en contra del gradiente electroquímico gracias a la energía liberada por introducir 2 o 3 protones. Se ha encontrado que está relacionado con diferentes patologías tales como la resistencia a la insulina en la obesidad, la tumorigénesis y convulsiones. Además, se ha identificado a la delección de este gen como el causante del síndrome de Wolf-Hirschhorn (Tamai et al. (2008), D. Jiang, Zhao, and Clapham (2009), De Marchi et al. (2014), Natarajan et al. (2021)).

2.6. Entrada capacitativa de Ca^{2+}

Anteriormente, en el apartado “2.3.1.4 Canales de Ca^{2+} operados por depósitos (SOCC)”, se realizó una breve introducción de los SOCC junto con otros tipos de mecanismos de transporte de Ca^{2+} intracelular. No obstante, el mecanismo que regula la entrada a partir de estos canales, SOCE, es uno de los principales intereses de nuestro grupo de investigación y, por ende, se ha creído oportuno escribir este apartado para darle la importancia que merece. En primer lugar, se realizará un repaso histórico de su descubrimiento. A continuación, se resumirán las bases moleculares de SOCE para, posteriormente, presentar los diferentes protagonistas de este mecanismo de transporte y por último, se explicará brevemente cómo la mitocondria es capaz de regular la SOCE.

2.6.1. Descubrimiento de SOCE

Históricamente, el proceso desde que se planteó este mecanismo hasta que, tras dos décadas, se comenzaron a establecer sus bases moleculares puede suscitar cierta pasión -como ha sido el caso del autor-. Su descubrimiento tuvo lugar en los años 80, cuando Putney Jr (1986), basándose en estudios los estudios previos de como M. J. Berridge and Irvine (1984), propuso el término de *Entrada Capacitativa de Ca^{2+}* (CCE, del inglés *Capacitative Ca^{2+} Entry*) para intentar explicar cómo la activación de ciertos receptores de la MP, a través de la producción de IP₃, que son capaces de regular los depósitos intracelulares de Ca^{2+} , provocaban la entrada sostenida de Ca^{2+} al interior celular tras el vaciado de dichos depósitos. Erróneamente, propuso un mecanismo por el cual, tras la liberación y agotamiento de Ca^{2+} del RE provocado por la unión de IP₃ a su receptor en este orgánulo, se formaría una ruta directa entre el medio extracelular y el interior del RE (el condensador) sin que este Ca^{2+} pasara por el citosol, la cual permanecería abierta hasta conseguir el llenado de dichos depósitos (Putney Jr (1986)).

Tras el paso de los años y un sin fin de experimentos, desde electrofisiológicos acoplados a imagen de fluorescencia en los que se observó un retraso entre la estimulación del IP₃R y la correspondiente entrada de Ca^{2+} (Penner, Matthews, and Neher (1988)), ensayos con inhibidores de la SERCA (tapsigargina, CPA...) que mostraron un aumento de $[\text{Ca}^{2+}]_{\text{cit}}$ y cómo la propia salida pasiva a través de los canales *leak* era suficiente para inducir la “CCE” (Takemura et al. (1989)), hasta el descubrimiento de diferentes corrientes de Ca^{2+} inducidas por el agotamiento de Ca^{2+} de los depósitos mediante técnicas como *patch-clamp* -la corriente I_{CRAC} fue la primera descubierta, mientras que el término I_{SOC} hace mención al resto de corrientes que se descubrieron posteriormente- (Hoth and Penner (1992), Prakriya and Lewis (2015), Parekh and Putney Jr (2005)).

Por un lado, los experimentos electrofisiológicos anteriores fueron sumamente importantes para asentar las bases moleculares de la SOCE, pues gracias a ellos, por primera vez, se identificó de forma directa la SOCE, ya que hasta entonces se habían medido incrementos de Ca^{2+} en el citosol (Penner, Matthews, and Neher (1988) Hoth and Penner (1992)). En concreto, Hoth and Penner (1992), mediante experimentos de imagen de Ca^{2+} y *patch-clamp* realizados en mastocitos

de rata en condiciones de vaciado de Ca^{2+} de los depósitos mediante IP₃, con ionomicina o el agente quelante EGTA, registraron una corriente de entrada de Ca^{2+} pequeña y sostenida, a la que denominaron I_{CRAC} , y que no era activada por voltaje. En cuanto a la unión liberación-entrada de calcio, observaron que entre la liberación de Ca^{2+} inducida por IP₃ y la entrada de este desde el medio extracelular existe un retraso, lo que sugiere una desconexión entre ambos mecanismos que puso en contradicho la teoría de una recarga directa del RE desde el medio extracelular propuesta por Putney Jr (1986). Además, sugirieron que este sería el mecanismo por el cual las células no excitables podrían mantener elevada la $[\text{Ca}^{2+}]_{\text{cit}}$ y rellenar los depósitos de Ca^{2+} tras su vaciado (Parekh and Putney Jr (2005)).

Por otro lado, los experimentos realizados con inhibidores de la *Bomba ATPasa de Ca^{2+} Sarco/Endoplasmática (SERCA)*, como la tapsigargina o el CPA, arrojaron luz sobre el papel de corrientes de Ca^{2+} en la SOCE, pues permitieron estudiar la liberación de Ca^{2+} de los depósitos sin la necesidad de la acción del IP₃, a través de los canales *leak* que continuamente liberan Ca^{2+} al citosol y, al estar la SERCA inhibida, este Ca^{2+} que sale de forma pasiva no vuelve a entrar al RE de tal manera que el RE acaba por vaciarse (M. Berridge (2014)). En concreto, observaron que la entrada de Ca^{2+} desde el medio extracelular a través de la MP era inducida por el propio vaciamiento del RE y, además, que no existe efecto aditivo entre la acción de la tapsigargina y el IP₃ lo cual sugiere que han de actuar sobre la misma ruta de entrada de Ca^{2+} al interior de la célula y, además, que el Ca^{2+} no entra primero a los depósitos y posteriormente se libera al citosol, ya que en ese caso se esperaría un efecto aditivo y un mayor aumento de Ca^{2+} al aumentar IP₃ (Parekh and Putney Jr (2005)).

Tras estos y otros muchos experimentos, en los que se evidenció que la teoría propuesta por Putney Jr (1986) de una conexión directa RE-SOCC era errónea y que, en realidad, el Ca^{2+} sí que pasaba por el citosol, conllevó la revisión, desde la década de los 90, de las hipótesis originales sobre este mecanismo y, en consecuencia, se propusieron diferentes posibles modelos para intentar explicar sus bases moleculares, aunque sin éxito ya que ninguno de ellos se pudo demostrar, pues también eran erróneos: modelo del mensajero difusible, de fusión vesicular o de acoplamiento conformacional. No obstante, en la década de los 90, en concreto en 1993, se comenzó a instaurar un nuevo nombre para este mecanismo más acorde a la realidad de las bases moleculares de su funcionamiento, en concreto *Ruta de Entrada de Ca^{2+} Operada por Depósitos (SOCP y SOCE posteriormente)* (Parekh and Putney Jr (2005), Montero, Garcia-Sancho, and Alvarez (1993b), Montero, Garcia-Sancho, and Alvarez (1993a)); sin embargo, por aquellos tiempos algunos autores aun eran reacios a usarla (M. J. Berridge (1995)).

Otro hito importante en la descripción de las bases moleculares de este transporte tuvo lugar entre los años 2005 y 2006, pues fue cuando se descubrió a los “culpables” de la detección de la disminución de Ca^{2+} del RE y de la entrada de este desde el medio extracelular. En particular, en el año 2005 se encontró que *STIM1*, una proteína transmembrana del RE, es el sensor de Ca^{2+} y, en 2006, se descubrió que *ORAI1*, una proteína transmembrana de la MP, es activada tras el vaciado de los depósitos para permitir el paso de Ca^{2+} desde el medio extracelular (Liou et al. (2005), Feske et al. (2006), Prakriya et al. (2006)). Posteriormente, se han ido encontrando más participantes en

este mecanismo. Por un lado, en cuanto a sensores de Ca^{2+} en el RE se ha descrito también el papel de *STIM2* (Williams et al. (2001), Soboloff et al. (2006), Parvez et al. (2008)), otra isoforma de *STIM*. Por otro lado, se han descrito diferentes SOCC que varían en su selectividad a Ca^{2+} , desde las isoformas de *ORAI1*, *ORAI2* y *ORAI3* (Mercer et al. (2006)) hasta miembros de la familia de los *TRP*, tales como los *TRPC1-7* y los *TRPV4,6* (M. Berridge (2014)).

2.6.2. Bases moleculares de la SOCE

A pesar de haber tres isoformas de *ORAI* y dos de *STIM*, la forma “clásica” de activación de la SOCE es a través de la interacción *ORAI1* y *STIM1*, pues son suficientes para dar lugar al canal CRAC. Previo su formación, tiene lugar la activación de ciertos receptores de la MP, como ciertos TKR y GPCR (figura 2.11). En concreto, cuando estos últimos se activan, a su vez acaban por activar a la proteína G trimérica acoplada a ellos, de tal manera que o bien las subunidades $\beta\gamma$ de cualquier proteína G o la subunidad α de las proteínas G_q son capaces de desencadenar la activación de la Fosfolipasa $C\beta$ (*PLCB*), enzima capaz de generar IP₃ y DAG a partir de la hidrólisis del lípido de membrana fosfatidilinositol-4,5-bisfosfato (*IP2*). Este IP₃ activa los IP₃R de la membrana del RE para permitir la liberación de Ca^{2+} al citosol desde este orgánulo y dando lugar al vaciado del mismo. Este vaciado es detectado por sensores situados en la membrana del RE, en concreto *STIM1* o *STIM2*, los cuales poseen un dominio *mano-EF* que une Ca^{2+} .

Mientras que en condiciones basales dicho catión permanece unido a estos y permite que se distribuyan uniformemente por la membrana y se muevan libremente a través de ella, cuando se libera al citosol y baja la $[\text{Ca}^{2+}]_{\text{RE}}$ este catión deja de unirse a los sensores y, en consecuencia, los *STIM* reducen su movilidad, oligomerizan y forman unas estructuras conocidas como *puncta*. Estas estructuras “acercan” la membrana del RE a la MP y sirven como nexo para que los sensores interactúen con los SOCC, los cuales oligomerizan y forman canales que permiten la entrada de Ca^{2+} al citosol desde el medio extracelular. Cabe destacar que existen diferentes SOC con diferente selectividad por este catión, desde las isoformas de *ORAI* hasta algunos miembros de la familia de los *TRP* (M. Berridge (2014)). Tras la entrada de Ca^{2+} desde el medio extracelular a través de los SOC, parte de este entra al RE a través de bombas como la *SERCA* con el fin de alcanzar los niveles basales de $[\text{Ca}^{2+}]_{\text{RE}}$. Así, tanto el incremento de $[\text{Ca}^{2+}]_{\text{cit}}$ como de $[\text{Ca}^{2+}]_{\text{RE}}$ hasta los niveles basales da lugar a la desoligomerización de los sensores y, en consecuencia, a detener este mecanismo de entrada (Parekh and Putney Jr (2005)) (figura 2.11).

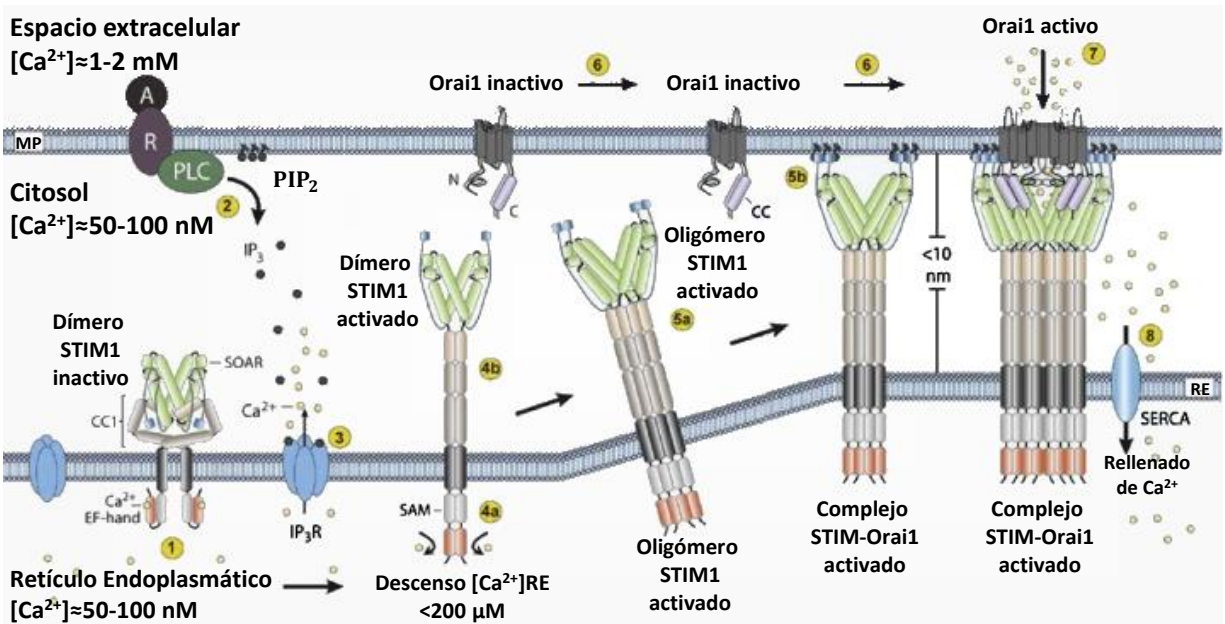


Figura 2.11: Activación de SOCE a través de SIMT1 y ORAI1. 1) En condiciones basales, cuando los depósitos de Ca^{2+} aún no se han vaciado, tanto los sensores de STIM1-2 como el SOCC ORAI1 forman homodímeros y se distribuyen a lo largo de las membranas del RE y plasmática, respectivamente. 2) No obstante, tras la estimulación de determinados agonistas sobre ciertos GPCR o TKR, se activa la PLC la cual hidroliza el fosfolípido de membrana PIP₂ y, como resultado, se generan dos productos de reacción: DAG e IP₃. 3) El IP₃ formado difunde por el citosol hacia el RE en cuya membrana se encuentra el IP₃R, cuya activación por esta molécula conlleva su apertura y liberación de Ca^{2+} hacia el citosol. 4a) En esta misma membrana también se localizan los sensores de Ca^{2+} STIM1-2 los cuales detectan la bajada de $[\text{Ca}^{2+}]_{\text{RE}}$ gracias a sus dominios EF, de tal manera que cuando $[\text{Ca}^{2+}]_{\text{RE}}$ disminuye el Ca^{2+} se disocia de estos dominios. En consecuencia, STIM dimeriza a través de la interacción entre los dominios EF-SAM en el lumen del RE. 4b) Esta oligomerización induce un cambio conformacional de STIM, de tal manera que adquiere una conformación extendida de sus dominios de hebra arrollada. Este cambio conformacional expone sus dominios SOAR y ricos en lisina. 5a) Los dímeros de STIM1 oligomerizan también a través de los dominios de hebra arrollada y esto les permite translocarse a las regiones de contacto entre la MP y el RE. 5b) Estos oligómeros se unen al PIP₂ de la cara interna de la MP a través de sus dominios ricos en lisina, lo que acaba por estabilizar los sitios de contacto entre el RE y la MP. 6) La estabilización de estos sitios de contacto favorece la acumulación de oligómeros de STIM1 y ORAI1 en estas regiones. 7) Así, a través de los dominios SOAR de los oligómeros de STIM1 se activan los canales ORAI1, por lo que se induce su apertura y el transporte de Ca^{2+} desde el medio extracelular al citosol. 8) Finalmente, gracias a la SERCA este Ca^{2+} entra en el RE en contra del gradiente electroquímico (modificado de Bhardwaj, Hediger, and Demaurex (2016)).

2.6.3. ORAI

ORAI, bautizado así en honor al *guardián de las puertas del cielo* de la mitología griega, es un grupo de genes que codifican un grupo de SOCC. En particular, en mamíferos se han descrito hasta un total de tres genes homólogos de ORAI: *ORAI1*, *ORAI2* y *ORAI3*. Estos codifican para las proteínas ORAI1-3 las cuales presentan ciertas diferencias entre ellas, tanto a nivel funcional como estructural (figura 2.12b). A nivel funcional, las tres isoformas presentan diferencias tanto en la eficacia para introducir Ca^{2+} a la célula como en su selectividad. Por un lado, el incremento de SOCE es mayor a través de ORAI1, seguido de ORAI2 y, en último lugar, ORAI3 (Mercer et al. (2006)). Además, aunque las tres isoformas presentan una gran selectividad por Ca^{2+} respecto del Na^+ , se ha observado que la permeabilidad a Na^+ es mayor para ORAI3 (Lis et al. (2007)).

Por otro lado, estas isoformas también muestran un patrón de inactivación diferente. Por ejemplo, a altas concentraciones $\sim 50 \mu M$ del antagonistas de la SOCE 2-APB se consigue una inhibición total del flujo de Ca^{2+} a través de ORAI1 y parcial para ORAI2. Sin embargo, el efecto observado en ORAI3 es el contrario; es decir, que este canal se activa (Lis et al. (2007), S. L. Zhang et al. (2008)). Además, también existen dos mecanismos de inactivación fisiológica denominados *rápida* o *lenta dependiente de Ca^{2+}* que también muestran diferencias entre isoformas. Por un lado, la inactivación lenta por Ca^{2+} sólo se observa para la isoforma ORAI1. Por otro lado, la inactivación rápida es mayor para ORAI3, seguido por ORAI2 y, en último lugar, ORAI1 (Lis et al. (2007), K. P. Lee et al. (2009)).

ORAI1, también conocido como el canal de *liberación de Ca^{2+} activada por Ca^{2+}* (CRAC), presenta una conductancia muy baja. Además, se han descrito dos isoformas de este canal: ORAI1 α -larga- y β -corta- (Sallinger, Berlansky, and Frischauf (2020)). Se ha descrito una mutación de pérdida de función en este gen, *-orai1 R91W-*, que es la causante de la inmunodeficiencia combinada severa. Como consecuencia de esta, los portadores de dicha mutación tienen una SOCE prácticamente inexistente. Cabe destacar que fue gracias a su implicación en esta enfermedad que se descubrió el papel de ORAI1 en SOCE y, además, de que fuera el primero que se asoció a dicho mecanismo (Feske et al. (2006)).

A nivel estructural, todas las isoformas están formadas por cuatro TM (TM1-TM4) altamente conservados unidos entre sí por dos lazos intracelulares (1 y 3) y uno extracelular (lazo 2); en cuanto a sus extremos C y N-terminal, ambos se localizan en la cara citoplasmática (figuras 2.12a y 2.12b). Respecto a cuánto de similares son las tres isoformas, de forma general ORAI1 presenta una identidad/homología del 62.99/60.3 % con ORAI2 y del 59.77/63.2 % con ORAI3, mientras que entre ORAI2 y ORAI3 es del 64.02/66.4 %. Respecto de los TM, la homología aumenta por encima del 90 %: 92.5 % para ORAI1-ORAI2 y un 93.8 % para ORAI1-ORAI3 y ORAI2-ORAI3 (Derler et al. (2012), Hasna et al. (2015)). A continuación, se van a explicar los diferentes dominios de ORAI, para lo que se usa como referencia ORAI1 y se indicarán aquellos casos en los que alguna de las otras dos isoformas difieran de esta. Así, desde el extremo N-terminal, pueden distinguirse diferentes regiones (figura 2.12) (Hasna et al. (2015), Fahrner et al. (2018), Sallinger, Berlansky, and Frischauf (2020)):

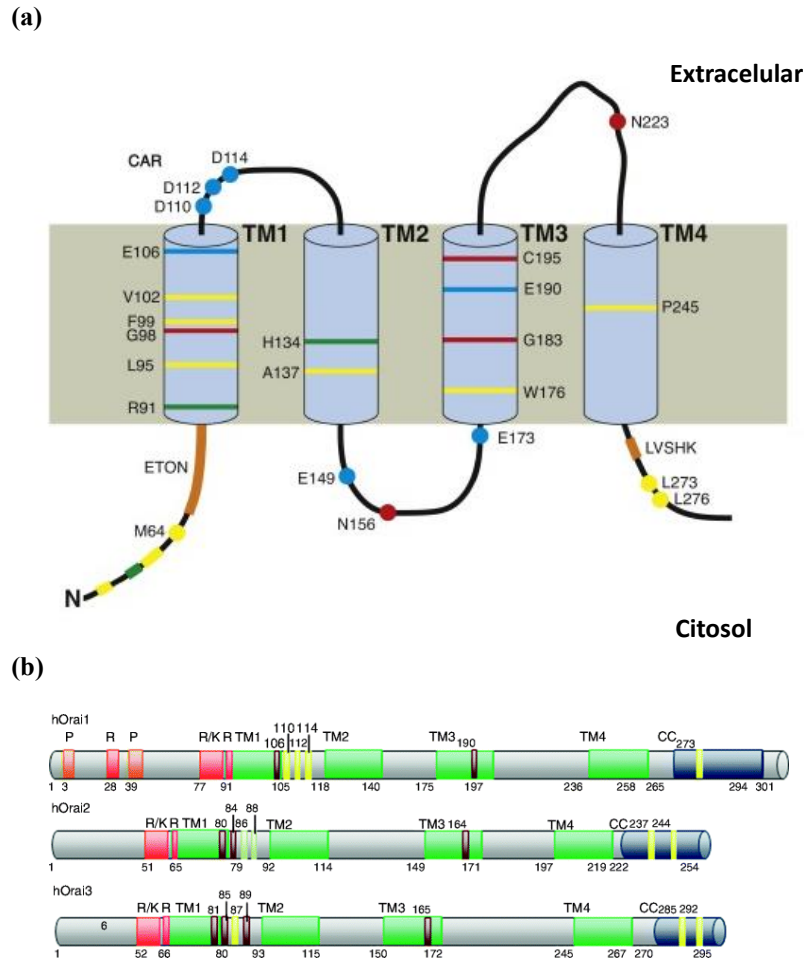


Figura 2.12: Representación de las tres isoformas de ORAI humano. a) Topología de las proteínas ORAI humanas (hORAI), usando la isoforma ORAI1 como ejemplo. Las tres isoformas comparten una topología común, que consta de cuatro dominios transmembrana (TM) conectados por dos lazos extracelulares (lazos 1 y 3) y uno intracelular (lazo 2), cuyo dominio N-terminal (en el que se encuentra la región ETON que participa en la unión STIM1-ORAI1, señalado en naranja) y C-terminal están localizados en la cara citosólica. Además, existen una serie de residuos importantes en su función y/o regulación, los cuales se indican en la figura en amarillo (apolares), verde (básicos), azul (ácidos) y rojo (neutros/polares). b) Predicción de los dominios de hORAI1-3. Abreviaturas: P, región rica en prolina; R, región rica en arginina; R/K, región rica en arginina/lisina; TM, dominio transmembrana; CC, dominio de hélices arrolladas o *coiled-coil* (Modificado de Sallinger, Berlansky, y Frischauf (2020), y Derler et al. (2012)).

- **Dominio N-terminal:** en este dominio se distinguen la región rica en prolina y argininas, la región de modificación post-transcripcional (fosforilación), y el *dominio N-terminal transmembrana extendido (ETON)* (Bergsmann et al. (2011), Hasna et al. (2015)). La *región rica en prolina y argininas*, que sólo está presente en ORAI1, participa en la reactivación tras la inactivación rápida dependiente de Ca^{2+} . A diferencia de ORAI2 y ORAI3, ORAI1 presenta una región de fosforilación por la PKC en los residuos Ser-27 y Ser-30 que puede regular negativamente SOCE y CRAC (Hasna et al. (2015)). En cuanto a ETON, en ORAI1 esta región contiene una zona de unión a Ca^{2+} – CaM y juega un papel importante tanto en la unión al *dominio de activación de CRAC (CAD)* de STIM1 como en la inactivación rápida dependiente de Ca^{2+} (Bonhenry et al. (2019), Mullins, Yen, and Lewis (2016)).

- *TM1*: esta región transmembrana, que es la más conservada entre las diferentes isoformas, es la encargada de formar el poro y el filtro de selectividad iónica, con un grado de conservación de alrededor del 81 % en las tres isoformas (Hasna et al. (2015)). En ella se distinguen diferentes residuos con diferentes funciones y cuyas mutaciones pueden dar lugar a desde formas inactivas hasta otras constitutivamente activadas independientemente de STIM1 y no selectiva a Ca^{2+} , pero selectiva en presencia de STIM1. (Fahrner et al. (2018), Sallinger, Berlansky, and Frischauf (2020)). Un residuo importante es R91, pues participa en la formación de una capa de agua y, además, actúa como puerta del canal. La sustitución R91W está implicada en la enfermedad de SCID, cuyos pacientes muestran una SOCE prácticamente nula debido al bloqueo del poro. Los residuos L95, F99 y V102 participan en la hidrofobicidad de la puerta del canal y, por ende, forman la barrera iónica en el estado cerrado que impide el paso de los iones. Mutaciones en alguno de estos tres residuos pueden estar implicadas en la *miopatía de agregado tubular (TAM)* y podrían dar lugar a una forma constitutivamente activa). En cuanto a G98, este residuo actúa como bisagra en la apertura del canal, de tal manera que algunas mutaciones pueden dar lugar a un ORAI1 constitutivamente activo, como las implicadas en la TAM -G98C/D/P/S-, y otras favorecen su inactivación, como G98R la cual está implicada en enfermedades autoinmunes. Por último, el residuo E106 (uno por cada monómero) da lugar al filtro de selectividad iónica al generar una región hidrofóbica y negativa que atrae a los cationes Ca^{2+} al canal -en ORAI3 este residuo se corresponde con E81 (Hasna et al. (2015))- La evidencia existente indica que la mutación E106Q da lugar a una forma inactiva del ORAI1 en TAM.
- *Lazo 1*: se ha descrito una región de acumulación de Ca^{2+} (CAR), constituida por tres residuos de aspartato no consecutivos que, debido a sus cargas negativas, atraen a este catión hacia el poro y facilita la permeación de este catión (Fahrner et al. (2018), Sallinger, Berlansky, and Frischauf (2020)).
- *TM2-TM4*: de forma general, se puede decir que el papel de los TM2-TM4 es estabilizador, pues crean una segunda capa que rodea el poro formado por los TM1 de cada subunidad. Sin embargo, estas también podrían ejercer diferentes efectos sobre SOCE. De hecho, la mayoría de las mutaciones que dan lugar a una forma de ORAI1 constitutivamente activa se localizan entre TM2 y TM4, lo que sugiere que el cierre del canal está estabilizado por interacciones múltiples entre los TM que cesan cuando STIM1 se une a este para permitir la apertura del canal. Es más, se han encontrado ciertas mutaciones de ganancia de función en TM2, tales como H134S/A/C/T/V/Q/E/M o A137V, que dan lugar a un ORAI1 constitutivamente activo independiente de STIM1, de tal manera que se crean puentes de hidrógeno entre S90 y R91 y se disminuye la hidrofobicidad de la puerta del canal al generar una cadena de moléculas de agua a su través. En concreto, la mutación A137V está relacionada con el cáncer de colon (Network and others (2012), Frischauf et al. (2017)).

- *Lazo 2*: se ha descrito que el lazo 2 de ORAI1 tiene un efecto inhibitorio a través de la región ¹⁵¹VSNV¹⁵⁴ que participa en la inactivación rápida del canal dependiente de Ca²⁺. Además, la evidencia parece indicar que podría ser responsable de las diferencias de comportamiento entre las isoformas ORAI1 y ORAI3 debido a la interacción entre el Y80 del extremo N-terminal y el N156 del lazo 2, de tal manera que el mantenimiento de la SOCE podría deberse a la facilidad con la que el extremo N-terminal se comunica con la región del lazo 2 y que será específica de cada isoforma (Fahrner et al. (2018)). Además, Fahrner et al. (2018) sugieren que en ORAI1, este lazo podría jugar un papel inhibitorio en su activación por STIM1 al dificultar la interacción entre ambos.
- *TM3*: este TM incluye la C195 que juega un papel clave en la regulación redox de ORAI1, lo que también explicaría la insensibilidad a la inhibición por H₂O₂ de ORAI3, pues no tiene un residuo correspondiente a C195 de ORAI1, sino que está sustituido por una glicina (G170) (Hasna et al. (2015), Bhardwaj, Hediger, and Demaurex (2016)). Este residuo puede oxidarse, por ejemplo, por el aumento de ROS consecuencia de la sobrecarga de Ca²⁺ mitocondrial (Ben-Kasus Nissim et al. (2017)). Además, algunas mutaciones en este dominio pueden afectar a la apertura y selectividad del canal, pero con diferentes resultados: W176C o E190Q dan lugar a una forma constitutivamente activa, mientras que L174D o G183A/D a una inactiva, esta última implicada en el cáncer, en concreto en glioblastomas. Además, E190Q también podría tener un papel en la sensibilidad a cambios de pH en ausencia de Ca²⁺.
- *Lazo 3*: esta región, que une TM3 y TM4 por la cara extracelular, es una de las menos conservadas entre isoformas, lo que se ve reflejado en una mayor longitud de esta en ORAI3 (72 residuos) y menor en ORAI1 (38 residuos), entre los que hay un 21 % de identidad de secuencia (Hasna et al. (2015)). Diferentes estudios proponen a esta región como una región muy flexible capaz de interactuar con residuos del lazo 1 y, por ende, de modular la permeación. A diferencia de ORAI2 y ORAI3, ORAI1 presenta una zona de glicosilación en N223, por lo que diferentes estados de glicosilación pueden conllevar diferencias en la señalización de Ca²⁺ mediada por la SOCE y, en consecuencia, tener un papel clave en diferentes patologías como el cáncer.
- *TM4, región Extendida de TM4 (M4ext) y región C-terminal*: con la apertura del canal, el dominio TM4 y su región extendida se enderezan debido al cese de la interacción entre el residuo P245 y el motivo SHK del C-terminal (residuos 263-265), lo que expone las zonas de unión con STIM1. De esta manera, la hélice TM4 se dobla en la posición P245 y se hace posible la interacción entre M4ext con el dominio CAD de STIM1 a través de la formación de las *coiled-coils*¹, estructuras clave en la estabilización del estado cerrado del canal y en la interacción con otras subunidades de ORAI. Además, otra característica importante

¹En ocasiones, las hélices- α de una proteína se enlazan entre ellas y dan lugar a una estructura muy estable y fuerte conocida como *hélices arrolladas* o *coiled-coil*. En concreto, para que esta estructura se dé, ha de haber dos o tres hélices- α en la que la mayoría de sus residuos presenten cadenas laterales hidrofóbicas a un lado permitiendo así la interacción hidrofóbica entre dichas cadenas.

es la flexibilidad en la unión entre TM4 y el extremo C-terminal, debida a la región de unión conocida como *nexus* -LVSHK-, imprescindible para el acoplamiento ORAI1-STIM1. Además, el extremo C-terminal de las isoformas ORAI2 y ORAI3 también participa en la inactivación rápida dependiente de Ca^{2+} que, como se explicó anteriormente, es mayor para ORAI3 que para ORAI2 y esta, a su vez, superior a ORAI1, lo que posiblemente pueda explicarse por la presencia de tres residuos de glutamato en este extremo C-terminal en las isoformas ORAI2 y ORAI3 de los cuales uno no está en ORAI1 (Lis et al. (2007), K. P. Lee et al. (2009)). En todos estos procesos, formación de los *coiled-coil*, unión STIM-ORAI y activación del canal, los residuos L273 y L276 juegan un papel clave de tal manera que mutaciones como L273S/D o L276D dan lugar a formas inactivas de ORAI1.

En la figura 2.12b se han mostrado la estructura secundaria y terciaria de ORAI1, y se ha destacado ciertas diferencias respecto de las otras dos isoformas ORAI2 y ORAI3. No obstante, aunque a nivel electrofisiológico sus características son muy similares, pues tienen una alta selectividad por Ca^{2+} respecto de los cationes monovalentes -los residuos del poro de selectividad son idénticos-, sí que se diferencian en cuanto a su inactivación rápida dependiente de Ca^{2+} (Bonhenry et al. (2019)). Además, se ha descrito que las tres isoformas de ORAI presentan estructura cuaternaria y pueden formar tanto homómeros como heterómeros. En particular, se piensa que los homómeros ORAI1 podrían ser hexámeros en los que el TM1 de cada monómero se sitúe en la cara interna de esta estructura para formar el poro y al resto de TM alrededor de dicho poro, si bien es cierto que también se baraja la posibilidad de que formen tetrámeros (Hao Dong et al. (2019), Sallinger, Berlansky, and Frischauf (2020)).

Respecto a STIM1, que podría formar complejos octaméricos, se une, a través de su dominio DAC/SOAR, al extremo C-terminal de ORAI1 y, posteriormente, al N-terminal, aunque todavía no está del todo claro el mecanismo. Actualmente, hay dos modelos que tratan de explicar diferentes estadios de la activación de CRAC (Fahrner et al. (2018), Butorac, Krizova, and Derler (2020), Niu et al. (2020), Sallinger, Berlansky, and Frischauf (2020)): el *Modelo dimérico*, que propone que el dímero de STIM1 se une a dos ORAI1 a través de sus regiones M4ext, y el *monomérico*, el cual sugiere que el dímero de STIM1 se une a un único ORAI1 por su extremo C-terminal. Este último modelo podría explicar cómo el dominio CAD/SOAR del dímero que quedaría libre sería el causante de la unión cruzada con otros ORAI1 que da lugar a clústers.

Independientemente del modelo considerado, y aunque todavía no se conocen bien los mecanismos de apertura de ORAI1, se han propuesto varios modelos en los que la interacción entre STIM1 y el extremo C-terminal de ORAI da lugar a un cambio conformacional en dicho extremo que conlleva la torsión del TM1 -forma el poro del ORAI- y, en consecuencia, permite la permeación selectiva de Ca^{2+} (Yeung, Yamashita, and Prakriya (2017)). Estos modelos son el *modelo girar para abrir* y el *modelo del reclutamiento aniónico*, los cuales abren la posibilidad a la existencia de dos compuertas, una en la región básica y otra en la región hidrofóbica (Frischauf et al. (2017), Yeung, Yamashita, and Prakriya (2017), X. Hou, Burstein, and Long (2018), Hao Dong et al. (2019), Xiaofen Liu et al. (2019)):

- *Modelo girar para abrir*: tras la unión de STIM1, las cadenas laterales hidrofóbicas de los residuos C102 y F99, que forman la barrera hidrofóbica que impide el paso de iones en el estado cerrado del canal, rotan unos 20° hacia afuera (Yeung, Yamashita, and Prakriya (2017), Hao Dong et al. (2019)). Además, la interacción entre TM1 y TM3 a través de las interacciones R83-E149 y K85-E173 podrían ser clave en la activación de la apertura de estos canales mediada por STIM1 así como de la activación gradual de estos (Hao Dong et al. (2019)).
- *Modelo del reclutamiento aniónico*: tras la activación de ORAI1 por STIM1, tiene lugar un ligero ensanchamiento del poro en la región básica (Frischauf et al. (2017), X. Hou, Burstein, and Long (2018)). Además, podría existir una ruta de transducción conformacional desde la hélice TM4 periférica a través de TM3 hacia la sección básica más interna de TM1. Por un lado, en el estado cerrado, TM4 actuaría como pestillo para limitar la apertura del poro por la cara citosólica del canal al estabilizar la conformación en la que la sección básica de TM1 está en la cara interna del poro y, debido a la repulsión de sus cargas positivas y al taponamiento aniónico, el flujo de Ca^{2+} se bloquea. Por otro lado, tras la activación de ORAI1 a través de STIM1, se pasaría a la conformación abierta del canal. Así, un cambio conformacional en la hélice periférica TM4, en concreto un enderezamiento de estas, se transduce al interior de la hélice TM1, lo que provoca el giro de la sección básica de TM1 hacia la cara citosólica que favorece el agregado de aniones, lo que disminuye la repulsión entre las cargas positivas, aumenta el gradiente de potencial a través de la membrana aumenta y, en consecuencia, se facilita la permeación de Ca^{2+} (figura 2.13) (Xiaofen Liu et al. (2019)).

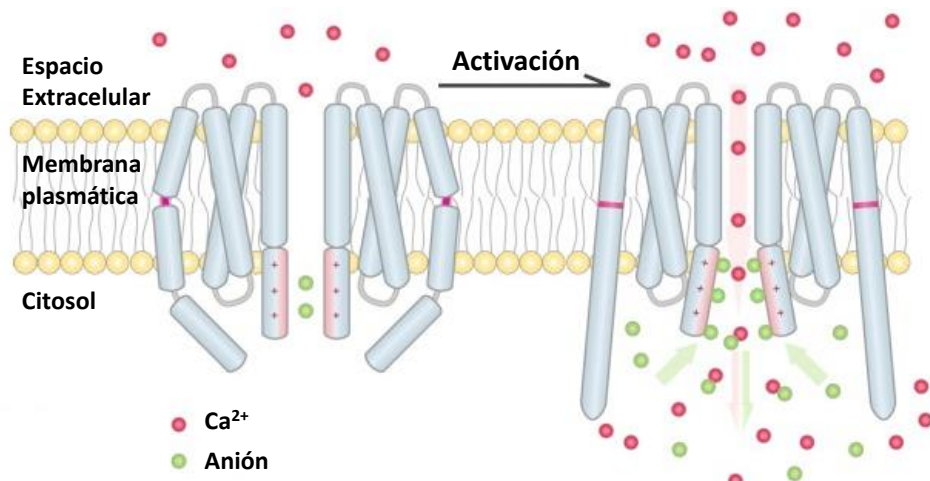


Figura 2.13: Modelo del reclutamiento aniónico de la apertura de ORAI y la permeabilidad de Ca^{2+} . A pesar de que se piensa que el homómero de ORAI puede ser un tetrámero o un hexámero, se muestran únicamente dos subunidades. En el estado cerrado, la hélice del TM4 limita la apertura del poro en la cara citosólica, mientras que en el estado activado la oscilación de la hélice induce un giro hacia fuera del poro en la cara citosólica que permite la acumulación de aniones (esferas verdes). En consecuencia, esta acumulación de aniones favorecen la disminución de la repulsión entre las cargas positivas e incrementa el gradiente de potencial a través de la membrana, lo que permite la entrada de Ca^{2+} (esferas rojas) (Modificado de X. Liu et al. (2019)).

No obstante, el mecanismo de entrada de Ca^{2+} dependiente de depósitos no es el único mecanismo de entrada selectiva de este catión en los que participan STIM y ORAI, pues también se ha descrito la implicación de ORAI en la entrada de Ca^{2+} desde el medio extracelular mediante otros mecanismos independientes de depósitos, regulados o no por STIM1:

- *Entrada de Ca^{2+} independiente de depósitos regulado por araquidonato (ARC):* se ha descrito la formación de pentámeros ORAI1/ORAI3 en relación 3:2, conocidos como canales ARC. Estos son altamente selectivos a Ca^{2+} , cuya activación es sensible al ácido araquidónico e independiente de depósitos. Además, las corrientes ARC, a diferencia de las CRAC, no muestran inactivación rápida. En estos canales, la presencia de dos ORAI3, y en especial su dominio N-terminal de ORAI3, es necesario no solo para que este sea sensible a la activación a bajas concentraciones de ácido araquidónico, alrededor de 2-8 μM , sino también para que sea insensible a la inactivación por 2-APB. Además, estos canales ARC están regulados por el STIM1 de la MP, que suponen entre el 10 y el 20% del total de STIM1 (Shuttleworth, Thompson, and Mignen (2007), Mignen, Thompson, and Shuttleworth (2007), Hasna et al. (2015), Mignen, Thompson, and Shuttleworth (2008), Mignen, Thompson, and Shuttleworth (2009)). Cabe destacar que el ácido araquidónico, por un lado, puede inducir el aumento de $[\text{Ca}^{2+}]_{\text{cit}}$ a través del aumento de la entrada de Ca^{2+} y, a su vez, inhibir SOCE inducida por ATP o taspigargina (Alessandra Fiorio Pla et al. (2008)).

El SARAF localizado en la MP puede regular de forma negativa la entrada de Ca^{2+} a través de ORAI1 independiente de depósitos y dependiente de ARC, pues está involucrado en la formación de los canales de ARC, formados por ORAI1:ORAI3 en relación 3:2 que, además, requieren del STIM1 localizado en la MP. Además, cuando la expresión de STIM1 está disminuida, SARAF interactúa con ORAI1 a través del extremo C-terminal de ORAI1 de forma transitoria y lo activa tras la estimulación del canal con determinados agonistas y este alcanza un máximo de 30 segundos de activación tras la estimulación (Lopez et al. (2016)). De hecho, autores como Galeano-Otero et al. (2021) han encontrado que SARAF y ORAI1 contribuyen a la activación de las células endoteliales y a la angiogénesis a través de una interacción funcional entre ellos por la que, en respuesta a VEGF, aumentan $[\text{Ca}^{2+}]_{\text{cit}}$.

- *Entrada de Ca^{2+} ORAI1-SPCA2 independiente de STIM:* la interacción entre el extremo N-terminal de la SPCA2 con el extremo C-terminal de ORAI1 activa la entrada de Ca^{2+} independientemente de depósitos y de los sensores STIM. Además, esta señalización constitutiva de Ca^{2+} puede tener implicaciones en la tumorigénesis (M. Feng et al. (2010)).

Existen diferentes patologías derivadas de mutaciones o expresiones anómalas en los genes *ORAI*, como SCID, relacionada con la mutación R91W, o la TAM, cuyo origen puede deberse a mutaciones en los TM de *ORAI1* -S97C, V107M, T184M...- (Sallinger, Berlansky, and Frischauf (2020)). Además, alteraciones en la expresión de STIM y ORAI están involucradas en enfermedades cardiovasculares, respiratorias y cáncer, lo que abre la posibilidad a posibles tratamientos terapéuticos dirigidos contra estas (X. Zhang et al. (2020)). En particular, se ha

descrito un remodelado en la expresión de los canales ORAI estrechamente relacionado con algunas de las características del cáncer: mantenimiento de la actividad proliferativa; resistencia a la muerte; capacidad metastática y migratoria, y angiogénesis tumoral o el escape inmune (figura 3.3) (Chalmers and Monteith (2018), Sakhno et al. (2019)). Tan es así que la inhibición de ORAI puede, por ejemplo, disminuir el crecimiento del cáncer de colon (Xie et al. (2016)).

2.6.4. STIM

Las *moléculas de interacción estromal* (STIM, del inglés *STromal Interaction Molecules*) son proteínas que, junto con ORAI, forman el sistema de canales CRAC. Su papel no se reduce a esta vía de transporte de Ca^{2+} dependiente del vaciado de los depósitos, pues también participa en otros mecanismos de entrada de este catión independiente del vaciado de los depósitos como la entrada de Ca^{2+} activada por ácido araquidónico (Grabmayr, Romanin, and Fahrner (2021), Shuttleworth, Thompson, and Mignen (2007), Mignen, Thompson, and Shuttleworth (2007)).

Existen dos isoformas de STIM: STIM1 y STIM2. Además, cada una de ellas presenta diferentes variantes de *splicing*. Por un lado, STIM1 presenta 22 variantes, 3 de los cuales son ncRNA. Por otro lado, de STIM2 sólo se han descrito 3: STIM2.1 (STIM2 β), STIM2.2 (STIM2 α , que es la convencional) y STIM2.3. En la figura 2.15 sólo se muestran 3 de estas 19 posibles proteínas de STIM1: STIM1; STIM1 Long (STIM1L), y STIM1A. La evidencia revela que las diferentes isoformas y variantes de *splicing* de STIM están interconectados a nivel funcional y pueden dar lugar a muy diferentes patrones de modulación de procesos como SOCE. Por ello, habría que considerarlas como un conjunto en lugar de como elementos individuales (Miederer et al. (2015), Rana et al. (2015), Berna-Erro et al. (2017), Grabmayr, Romanin, and Fahrner (2021)).

El STIM localizado en la membrana del RE actúa, gracias a su dominio mano EF de su extremo N-terminal de la cara luminal, como un sensor de Ca^{2+} . De esta manera, cuando $[\text{Ca}^{2+}]_{\text{RE}}$ disminuye este catión se disocia de STIM. En consecuencia, STIM sufre una serie de cambios conformacionales que comienzan por su extremo N-terminal y se propaga hacia el C-terminal y, en consecuencia, esta proteína se activa (Grabmayr, Romanin, and Fahrner (2021)). Esta serie de cambios dan lugar a una forma más extendida de STIM capaz de oligomerizar y de acercarse a la MP e interactuar con ORAI (Derler, Jardin, and Romanin (2016), Grabmayr, Romanin, and Fahrner (2021)). Este acoplamiento, que puede llevarse a cabo tanto a través del extremo C-terminal como del N-terminal de STIM1 y del lazo 2 de ORAI, provoca la apertura de este canal (figura 2.14). Hasta la fecha, a pesar de que no se ha resuelto la estructura completa de las proteínas STIM, sí que ha sido posible resolver la estructura de los dominios N-terminal y C-terminal (Zheng et al. (2011), Stathopoulos et al. (2008), Grabmayr, Romanin, and Fahrner (2021)). No obstante, los modelos predictivos, como los modelos de dinámica molecular, pueden arrojar luz sobre estas proteínas. Por un lado, las predicciones apuntan que el extremo N-terminal de unión a Ca^{2+} estaría unido al dominio TM. Por otro lado, el extremo C-terminal podría pasar de una conformación apretada en el estado quiescente a una conformación más extendida en el activo (Y. Zhou et al. (2015), Derler, Jardin, and Romanin (2016), Tiffner et al. (2021)) (figura 2.14).

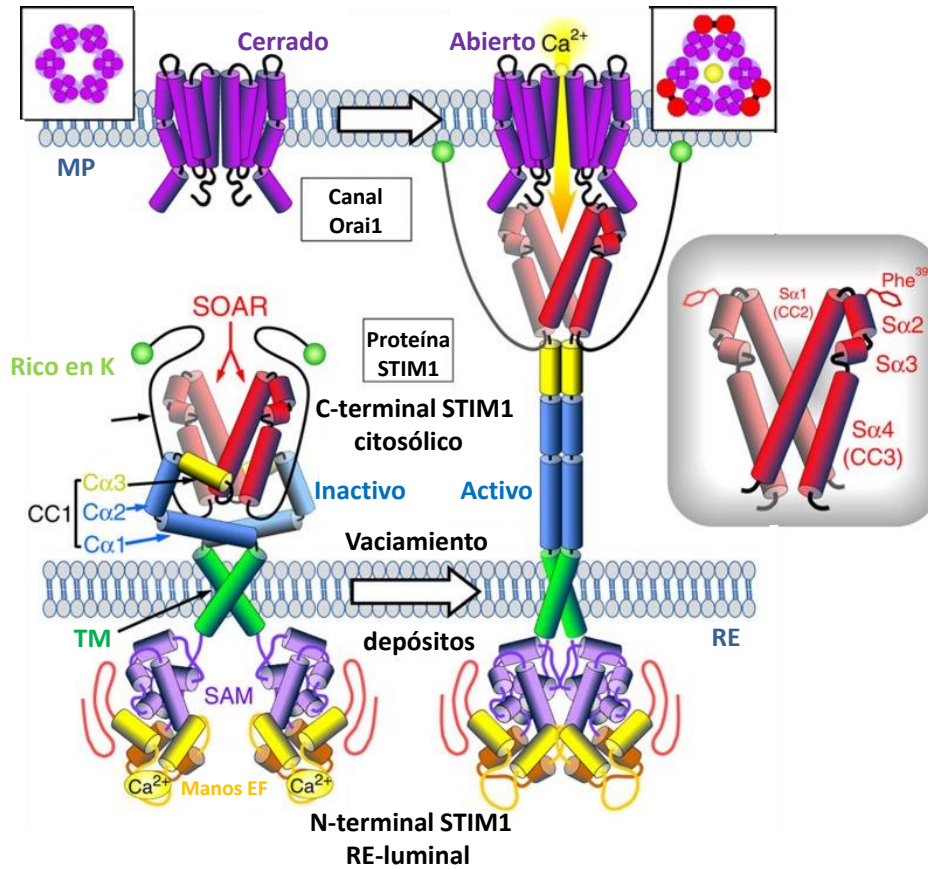


Figura 2.14: Interacción de STIM1 con ORAI1. A la izquierda se muestra la conformación inactiva o en reposo de ORAI1 y del dímero STIM1 y en la derecha la activa en la que los dímeros de STIM1 oligomerizan e interaccionan con ORAI1 para inducir la apertura del canal. En su extremo N-terminal se localiza el complejo de mano EF-motivo α estéril (SAM), que es el responsable de que esta proteína sea un sensor de Ca^{2+} . En cuanto al dominio C-terminal, de longitud superior al del extremo N-terminal, se localiza en la cara citoplasmática y contiene la región activadora de STIM-ORAI (SOAR) y una región flexible que contiene una región rica en lisina. En cuanto al dominio SOAR, contiene cuatro hélices que interactúan entre sí a través de interacciones hidrofóbicas y puentes de hidrógeno: las dos hélices más alargadas son $\text{S}\alpha 1$ (o CC2) y $\text{S}\alpha 4$ (CC3) las cuales flanquean a $\text{S}\alpha 2$ y $\text{S}\alpha 3$. Previo al vaciado de los depósitos, los dímeros de STIM1 presentan una conformación plegada e inactiva (izquierda) en la que los dominios SAM del dímero están separados y las SOAR de cada unidad están ocluidas en un complejo plegado de SOAR, CC1 y el extremo C-terminal flexible. Tras el vaciado, el Ca^{2+} se disocia del lado luminal de los sensores STIM1 y los dominios N-terminales se asocian y dan lugar a una serie de reordenamientos de los dominios TM de tal manera que transducen el cambio conformacional al extremo C-terminal. En concreto, el ángulo entre las hélices TM se modifica y la estructura abrazadera inhibitoria intramolecular entre $\text{CC}1\alpha 1$ y $\text{CC}3$ (de SOAR) se disocia y dicho extremo C-terminal experimenta un despliegue y alargamiento. En consecuencia, sufren un cambio conformacional y pasan de la conformación plegada e inactiva (izquierda) a una extendida (centro) y activa. Este cambio conformacional en el que el extremo C-terminal se extiende permite que la región rica en lisinas de STIM1 se una a la MP localizada en las uniones RE-MP, EB1. Además, al exponer las SOAR se induce la oligomerización (superior a dímeros) de STIM1 en el que los dominios CC3 tienen un papel fundamental. Esta oligomerización redistribuye a esta proteína en grupos discretos de la misma y da lugar a la formación de unas estructuras conocidas como *punctas* cercanas a la MP, lo que facilita la unión al extremo C-terminal de ORAI1 a través de los dominios $\text{CC}1\alpha 3$ – $\text{CC}2$ de STIM1 en las uniones RE-PM. Cabe destacar que $\text{S}\alpha 2$ incluye un residuo, F394 (Phe394), fundamental para la unión y apertura de los canales ORAI1 a través de STIM1 cuya sustitución en otras isoformas explica las diferentes de eficacia de apertura entre ellas. Esta interacción induce la apertura del canal ORAI1 la cual permite la entrada de Ca^{2+} desde el medio extracelular (Modificado de Zhou et al. (2015)).

STIM es una proteína transmembrana tipo I de paso único dimérica que, principalmente, está presente en la membrana del RE, si bien es cierto que también se localiza en depósitos ácidos y en la MP (Grabmayr, Romanin, and Fahrner (2021)). Su extremo N-terminal se sitúa en el lumen del RE, donde se localiza el sitio de unión a Ca^{2+} , llamado *motivo- α Esteril EF (SAM)*. En cuanto a su extremo C-terminal, da a la cara citosólica y presenta tres segmentos *hélices arrolladas* o *coiled-coil* (CC1-3) seguidos de una región flexible (figura 2.15)(Hooper et al. (2013), Fahrner et al. (2017), Novello et al. (2020), Grabmayr, Romanin, and Fahrner (2021), Tiffner et al. (2021), Berlansky et al. (2021)).

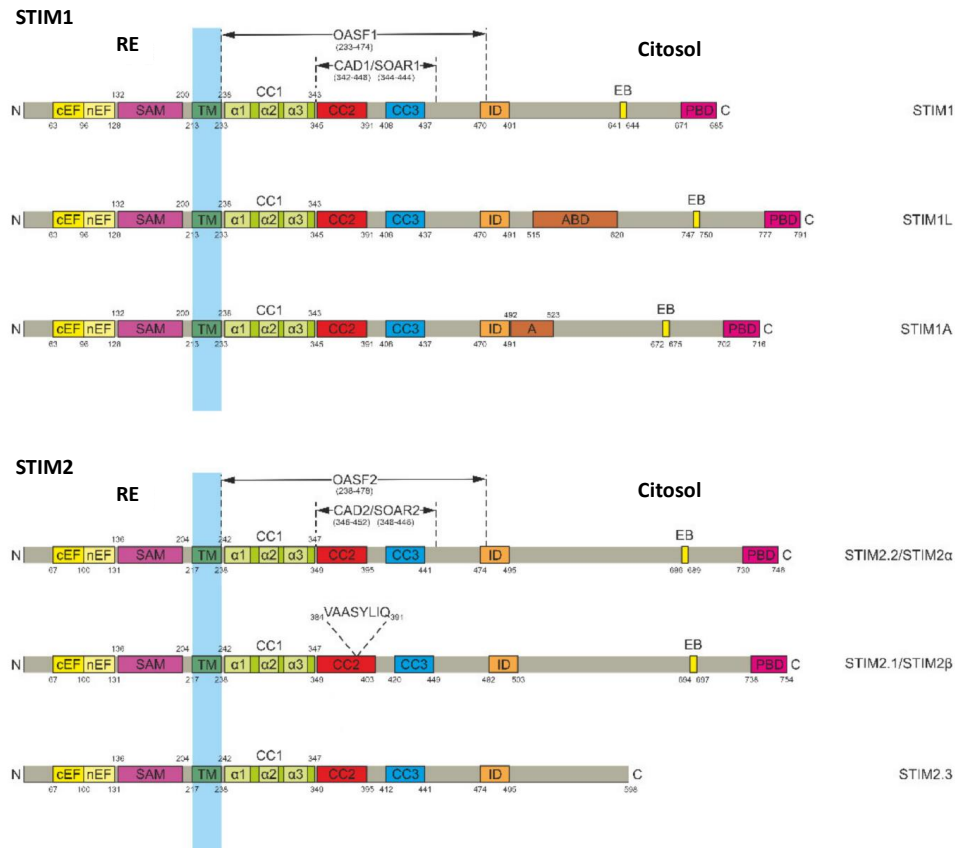


Figura 2.15: Estructura y dominios de las proteínas STIM. En la figura se muestra la estructura primaria y dominios más importantes de tres de las isoformas más estudiadas de STIM1 (superior) y de STIM2 (inferior). El dominio N-terminal, localizado en el lumen del RE, incluye al péptido señal, SP, al dominio mano EF canónico (cEF) y no canónico (nEF), y al motivo α estéril (SAM). El extremo C-terminal, que es el dominio citoplasmático, incluye el fragmento pequeño activador de ORAI (OASF) constituido por tres dominios *coiled coil* (CC1-3), el primero de ellos, a su vez, dividido en $\alpha 1-3$. Además, al fragmento constituido únicamente por CC2 y CC3 se le conoce como dominio activador de CRAC (CAD) o como región activadora STIM-ORAI (SOAR). Este extremo C-terminal también incluye el dominio de inactivación (ID), el dominio de unión al extremo de los microtúbulos (EB) y el dominio rico en lisinas o polibásico (PBD). En cuanto a las diferencias entre las diferentes isoformas, STIM1L y STIM1A contienen una inserción, debido a *splicing* alternativo, en el extremo C-terminal posterior al dominio ID: dominio de unión a actina (ABD) y dominio A, respectivamente. Cabe destacar que a la isoforma de STIM2.2 se le ha eliminado los 87 residuos del péptido señal en la representación. La similitud entre STIM1 y STIM2.2 es elevada y, por ello, sus dominios funcionales son equivalentes. No obstante, las isoformas STIM2.1 y STIM2.3 son variantes de *splicing*: en STIM2.1 se ha insertado un fragmento de 8 residuos (VAASYLIQ) dentro del dominio CC2, mientras que en STIM2.3 se ha eliminado el extremo final de 148 residuos (modificado de Grabmayr, Romanin, and Fahrner (2021)).

A continuación, se desarrollan en más detalle las diferentes regiones y dominios de estas proteínas, comenzando desde el extremo N-terminal (figura 2.15). Además, se explicarán las diferencias entre las dos isoformas más convencionales, STIM1 y STIM2.2, y las implicaciones funcionales que estas tienen que, probablemente, prevengan la sobrecarga de Ca^{2+} mediada por STIM2.2 (Hooper et al. (2013), Fahrner et al. (2017), Novello et al. (2020), Grabmayr, Romanin, and Fahrner (2021), Tiffner et al. (2021), Berlansky et al. (2021)).

Por un lado, en el *extremo N-terminal*, localizado en la cara luminal del RE, se localiza el *complejo mano EF-motivo α estéril (EF-SAM)*: este dominio es el que le confiere a STIM su papel como sensor de Ca^{2+} . En concreto, es capaz de detectar cambios en la concentración de este catión dentro del lumen del RE en un rango de 100 a 400 μM en función de la isoforma. En particular, para conseguir una activación semimáxima (EC50) de STIM1 el límite es 200 μM , mientras que para STIM2.2 es de 400 μM . Este dominio, a su vez, está formado por dos dominios la *mano EF canónica (cEF)*, que constan de dos motivos *hélice-lazo-hélice* (estructura consenso de las manos EF), y la *mano EF no canónica (nEF)*, de tal manera que el primero es el dominio de unión a Ca^{2+} y el segundo proporciona estabilidad a esta unión (Grabmayr, Romanin, and Fahrner (2021)). Así, mientras el Ca^{2+} esté unido a STIM los dominios EF-SAM permanecerán separados unos de otros. Por el contrario, cuando los niveles de Ca^{2+} disminuyen entonces este se disocia de STIM y se produce un entrecruzamiento entre los dominios N-terminales de STIM que da lugar a su oligomerización (Y. Zhou et al. (2015), Grabmayr, Romanin, and Fahrner (2021)).

Tras su oligomerización, la interacción entre los dominios EF-SAM induce la formación de hebras arrolladas, lo que genera los reordenamientos necesarios de los dominios TM para la formación de las hebras arrolladas y propagar la señal al extremo C-terminal en la cara citosólica. Esta oligomerización redistribuye a STIM en grupos discretos localizados en la membrana del RE y da lugar a la formación de unas estructuras conocidas como *punctas* cercanas a la MP las cuales facilitan la unión al extremo C-terminal de ORAI1 en las uniones RE-PM. Además, estas uniones entre ambas membranas están favorecidas por una serie de proteínas que interactúan con STIM1 como ESut1, que interactúa con el PIP2, EB1, junctate o STIMATE, esta última capaz de favorecer la formación de las *punctas* en las zonas de unión RE-PM (Derler, Jardin, and Romanin (2016)).

Así, la interacción entre STIM1 y ORAI1 induce la apertura del canal ORAI1 la cual permite la entrada de Ca^{2+} desde el medio extracelular. Estos oligómeros de STIM se unen al PIP2 de la cara interna de la MP a través de sus dominios ricos en lisina, lo que acaba por estabilizar los sitios de contacto entre el RE y la MP, lo cual favorece la acumulación de oligómeros de STIM y ORAI en estas regiones (Bhardwaj, Hediger, and Demaurex (2016)). En este dominio es donde radica una de las diferencias fundamentales entre las isoformas STIM1 y STIM2.2, su afinidad por el Ca^{2+} , pues es inferior para STIM2.2 (EC50 de 400 μM) respecto de STIM1 (200 μM) y, en consecuencia, es más sensible a cambios en la $[\text{Ca}^{2+}]_{\text{RE}}$. Esto es resultado de los diferentes grados de estabilidad de los dominios EF-SAM que pueden presentar los diferentes homólogos de STIM. En consecuencia, estas proteínas se pre-localizan en las zonas de contacto de RE con la MP (RE-MP) cuando los depósitos de Ca^{2+} aun están en niveles basales y son capaces de ajustar la velocidad de entrada de

este catión con el fin de regular y mantener la homeostasis del calcio intracelular. Por un lado, ante un vaciado pequeño de los depósitos, STIM2.2 es capaz de activar una SOCE pequeña y mantenida, lo que, a su vez, impide la activación de STIM1. Por otro lado, cuando el vaciado de los depósitos es pronunciado, STIM1 se activa y da lugar a una muy diferente a la que genera STIM2.2, es decir, más grande y transitoria, pues ahora interesa conseguir un llenado de los depósitos lo más rápido posible para alcanzar nuevamente los niveles basales (Grabmayr, Romanin, and Fahrner (2021)).

En cuanto al *dominio transmembrana*, este se encarga de propagar los cambios conformacionales que tienen lugar en el extremo N-terminal hacia el C-terminal para dar lugar a la activación de los SOCC. En este *extremo C-terminal*, cuya longitud es superior al extremo N-terminal, se distinguen diferentes partes (Grabmayr, Romanin, and Fahrner (2021)):

- *Fragmento pequeño activador de ORAI (OASF)*: este fragmento abarca los tres dominios de hebra arrollada (CC): CC1 y CC2-CC3. El primero se divide en tres hélices α ($\alpha 1 - 3$) dispuestas de forma antiparalela y unidas entre sí por dos lazos flexibles (Rathner et al. (2021)). En cuanto a los otros dos, forman el *dominio activador de CRAC (CAD)*, también llamado *región activadora de STIM-ORAI (SOAR)*, que participa en la activación de los canales ORAI. Este dominio se puede subdividir en cuatro regiones helicoidales, $S\alpha 1 - 4$, tal que $S\alpha 1$ se corresponde con CC2 y $S\alpha 4$ con CC3, las cuales flanquean a $S\alpha 2$ y $S\alpha 3$. Además, la hélice apical $S\alpha 2$ incluye un residuo, F394, fundamental para la unión y apertura de los canales ORAI1 a través de STIM1 (Y. Zhou et al. (2015)). A pesar de que dominios CAD/SOAR de STIM1 y STIM2.2 están muy conservados, existe alguna sustitución que puede dar lugar a diferencias entre estos homólogos, como la sustitución del residuo F394 de STIM1 por el L485 de STIM2.2 o del residuo G379 de STIM1 por el E470 de STIM2.2. En primer lugar, la sustitución del residuo F394 de STIM1 por el L485 de STIM2.2 en $S\alpha 2$ explica la menor eficacia de apertura de ORAI1 por STIM2.2, por lo que este residuo podría ser clave en el acoplamiento STIM1-ORAI a través de la unión entre el F394 de STIM1 con la extensión de la hélice TM4 del extremo C-terminal de ORAI (Baraniak Jr et al. (2020)). En segundo lugar, sustitución del residuo G379 de STIM1 por el E470 de STIM2.2 es capaz de explicar el acoplamiento más débil entre STIM2.2 y ORAI1 en comparación con STIM1. También se ha encontrado que $S\alpha 3$ tiene un papel fundamental en la activación de ORAI y en la transmisión de la señal desde STIM hasta ORAI a través de los residuos L402 de STIM1 o L493 de STIM2.2² y A166 de ORAI1, conocido como *interfaz de apertura STIM1-ORAI (SOGI)* (Grabmayr, Romanin, and Fahrner (2021)).

En el estado inactivo de STIM (depósitos de Ca^{2+} llenos) el extremo C-terminal presenta una conformación sumamente plegada estabilizada por la interacción entre CC1 y CAD/SOAR, con un papel fundamental de $CC1\alpha 1$ y CC3, la cual recibe el nombre de *abrazadera de hebra arrollada*. Además, $CC1\alpha 2$ también podría estar implicado en la formación de esta estructura en abrazadera por interacciones con $CC1\alpha 1$. Tras el vaciado de los depósitos y la

²La secuencia de STIM2 se suele mostrar de dos formas diferentes en función de si se tienen o no en cuenta los 87 aminoácidos adicionales del péptido señal. En este caso, sí se han tenido en cuenta.

activación de STIM1, CC1 y CC3 también participan en el proceso de oligomerización, pues CC1 α 1 homomeriza dentro del dímero STIM1/STIM1 a la par que se pierde la estructura de abrazadera y, a continuación, a través de CC3 los dímeros de STIM1 oligomerizan y da lugar a los clústers de STIM1 (Derler, Jardin, and Romanin (2016), Rathner et al. (2021)). Por lo tanto, el control de la activación de los canales CRAC podría deberse a un equilibrio entre las interacciones interhelicoidales de CC1 y la interacción entre CC1 y el dominio CAD/SOAR de CC3 (Rathner et al. (2021)). En estos dominios pueden tener lugar diferentes mutaciones con diferentes efectos. Por ejemplo, la mutación R304W, causante del síndrome de Stormoken, a pesar de estar localizada en CC1 α 2 también parece tener un papel en la formación o, más bien, en la no formación de la abrazadera de hebra arrollada, lo que da lugar a un estado activo de STIM1, lo que podría explicarse por un aumento de la homomerización de CC1 α 1 dentro de los dímeros STIM1 y por un aumento de la rigidez en el lazo entre CC1 α 2 y CC1 α 3.

- *Dominio de inactivación (ID o ID-STIM)*: es una región cargada negativamente que participa en la inactivación dependiente de Ca²⁺ del canal CRAC, en la que los residuos D476, D478 y D479 tiene un papel fundamental. Además, es posible que el W76 de ORAI1 esté asociado con ID-STIM y que actúen juntos para llevar a los canales CRAC desde un estado inactivo residual hacia una completa inactivación (Mullins and Lewis (2016)).
- *Dominio de unión al extremo de los microtúbulos (EB)*: este dominio hace posible la unión de STIM al extremo de los microtúbulos en crecimiento a través de las *proteínas de unión al extremo de los microtubulos (EB)*. En concreto, STIM1 es capaz de moverse rápidamente mediante EB1, proteína que no parece tener efecto sobre la SOCE. No obstante, la unión de EB1 sí que podría disminuir la capacidad de STIM1 para migrar a las regiones RE-MP para interactuar con ORAI1. Además, el vaciado de los depósitos podría suponer la liberación de STIM de los microtúbulos.
- *Dominio de unión a actina (ABD)*: este dominio sólo está presente en STIM1L. En concreto, esta es una variante alternativa de *splicing* del exón 11 que da lugar a la incorporación de 106 aminoácidos en el extremo C-terminal que actúan como un dominio de unión a actina y que permite la formación de clústers de STIM1L y la interacción con los canales ORAI1 (Grabmayr, Romanin, and Fahrner (2021)).
- *Dominio A (A)*: este dominio sólo está presente en la isoforma STIM1A (Grabmayr, Romanin, and Fahrner (2021)).
- *Dominio polibásico (PBD) o rico en lisina*: este dominio permite a STIM interactuar con los fosfolípidos cargados negativamente de la MP y podría ayudar en la formación previa de grupos de STIM1.

En el RE, STIM1 podría estar formando dímeros en condiciones basales con el fin de detectar disminuciones en la $[Ca^{2+}]$. Así, tras la disminución de este, sufriría un cambio conformacional que le permitiría oligomerizar y, en consecuencia, dirigirse hacia uniones RE-MP. Una vez en estas regiones, sus dominios CAD/SOAR se unen al canal ORAI1. De esta manera, se produce un cambio conformacional del complejo que forma el canal y este se abre para permitir la entrada de Ca^{2+} desde el medio extracelular. Posteriormente, tras el llenado de los depósitos, STIM1 deja de interactuar con ORAI1 y se inactiva la entrada de Ca^{2+} . En consecuencia, tanto ORAI1 como STIM1 regresan a su situación en reposo (Tiffner et al. (2021)). En cuanto a STIM2.2, se ha observado que forma heterodímeros con STIM1, lo que aumenta el reclutamiento de este último hacia las uniones RE-MP y facilita su activación.

Como se ha dicho anteriormente, los diferentes isoformas y variantes de *splicing* de STIM y cómo se relacionan entre sí puede dar lugar a diferentes formas de modular procesos como la SOCE. Por ello, es importante tener una visión global a la hora de estudiarlos, es decir, analizarlos como un conjunto más que en elementos individuales (Berna-Erro et al. (2017), Grabmayr, Romanin, and Fahrner (2021)). Dicho esto, se van a dar unas ligeras pinceladas sobre la importancia de alguno de las isoformas y variantes de *splicing* con el fin de tener una visión global de ellos y de poder integrar el comportamiento individual de cada uno de ellos en un comportamiento global de todos ellos (Grabmayr, Romanin, and Fahrner (2021)):

- A diferencia de STIM1, el papel de STIM2.2 podría tener un papel complementario, pues no se han encontrado, hasta la fecha, enfermedades humanas debidas a mutaciones ni de ganancia ni de pérdida de función de esta isoforma. No obstante, está empezando a cobrar importancia tanto en el sistema nervioso como en el inmune, así como en enfermedades relacionadas con el envejecimiento y neurodegenerativas como el Alzheimer o el Huntington (Berna-Erro et al. (2017), Grabmayr, Romanin, and Fahrner (2021)).
- De forma general, se ha observado que STIM1 se puede unirse a TRPC1, TRPC4 y TRPC5 también. También se ha observado que la inhibición de STIM1 inhibe la activación de TRPC5 por receptor, aunque no por La^{3+} , lo que indica que STIM1 es necesario para la activación de los TRPC por agonista, pero no lo es para que estos ejerzan su función (J. P. Yuan et al. (2007)). STIM1 también es capaz de regular TRPC3 y TRPC6, pues, aunque no interactúa directamente con ellos, sí facilita la formación de los heterodímeros TRPC3/TRPC1 y TRPC6/TRPC4 (J. P. Yuan et al. (2007)).
- STIM1L, a diferencia de STIM1, posee el dominio ABD capaz de unirse a actina y que le permite formar agregados de STIM1L e interactuar con los canales ORAI1. En cuanto a su expresión, esta se limita al cerebro y corazón de ratones, cardiomiocitos de ratas neonatas o células del músculo esquelético humanas. Esta isoforma se relaciona con una rápida activación de la SOCE en las células del músculo esquelético, inferior a un segundo, al comparación con otros tipos celulares en los que puede llevar minutos. Esta isoforma se asocia con el *receptor de potencial transitorio canónico 1 (TRPC1)*, TRPC3, TRPC4 y TRPC6, mostrando preferencia por TRPC1 frente a ORAI1.

- **STIM1A:** a diferencia de otras isoformas, esta contiene un exón adicional, A, entre los exones 10 y 11, que le aporta 31 residuos adicionales al extremo C-terminal junto a ID-STIM. Aunque se expresa en muchos tejidos, la mayor cantidad se encuentra en los testículos, el corazón, el riñón y los astrocitos. Esta isoforma reduce SOCE y la corriente CRAC en los astrocitos. Este exón adicional podría impedir la estabilización de la interacción entre ETON de ORAI1 y la región CAD/SOAR de STIM1A.
- **STIM2.1:** esta isoforma contiene una inserción de 8 residuos (VAASYLIQ) en el dominio CC2 de CAD/SOAR codificado por un exón adicional que, como consecuencia, dificulta la interacción con ORAI1. La expresión de esta isoforma es ubicua y dependiente del tipo celular (a diferencia de STIM2.2). Por ejemplo, se han encontrado altos niveles de STIM2.1 en células T vírgenes y su expresión es similar a la de STIM2.2. A diferencia de STIM2.2, que promueve la SOCE, STIM2.1 muestra un efecto inhibitorio, lo que podría deberse a la heterodimerización de STIM2.1 tanto con STIM2.2 o STIM1 y dar lugar a la ausencia de entrecruzamiento del canal ORAI1 (el entrecruzamiento aumentaría la eficacia del canal). También se ha observado que el acoplamiento el extremo C-terminal de ORAI1 está impedida lo cual previene la activación de ORAI1. Se ha encontrado que esta isoforma está relacionada con el fallo cardíaco por una disminución en la expresión de STIM2.1, lo que conlleva a la estimulación de SOCE y, posiblemente, a la hipertrofia cardíaca. También se ha encontrado relación con la miogénesis a través del arresto del ciclo celular, ya que inhibe SOCE y bloquea la proliferación celular y, en consecuencia, la célula pasa de un estado proliferativo a la diferenciación. En cuanto a STIM2.2, que activa SOCE, parece ejercer un efecto complementario en el control de la homeostasis del Ca^{2+} intracelular, pues no se conocen enfermedades asociadas a mutaciones de ganancia o pérdida de función de esta.
- **STIM2.3:** en la actualidad, la función de esta isoforma no se conoce. Esta se genera por un exón 13 alternativo. La expresión de esta isoforma parece ser limitada, pues, por ejemplo, en los linfocitos no se detecta su ARNm pero sí el de las otras dos isoformas.

2.6.5. MS4A12

La evidencia existente ha demostrado que los miembros de la familia MS4A (subfamilia A de la familia con 4 dominios TM), compuesta por 18 miembros, muestran un patrón de expresión específico de tejido. Además, el CD20 (MS4A1), expresado de forma específica en las células B, no es solo un marcador del linfoma de células B, sino también una interesante diana terapéutica del tratamiento con mAb (Reff et al. (1994)). En cuanto a su relación con la homeostasis del Ca^{2+} intracelular, se ha observado la activación de la entrada de Ca^{2+} a la célula tras la unión de mAb anti-CD20 (Bubien et al. (1993)). Por ello, Koslowski et al. (2008), con el fin de encontrar genes expresados de forma concreta en la membrana de las células epiteliales de colon y que fueran estructuralmente similares a la proteína CD20, emplearon métodos de minería de datos sobre el genoma completo. Como resultado, encontraron que la proteína MS4A12 cumplía dichos criterios,

pues de los 27 tejidos que estudiaron sólo la mucosa de colon normal mostró una expresión por encima del límite de detección. En cuanto a su localización celular, en condiciones normales la proteína se localiza en la región apical de las células epiteliales de colon (J. W. Han et al. (2021)). Se ha descrito que esta proteína participa en procesos de diferenciación, proliferación y regulación del ciclo celular (L. He, Deng, and others (2017)).

Esta proteína y los mecanismos moleculares subyacentes en su participación en la entrada de Ca^{2+} a la célula y, en concreto, en SOCE apenas han sido estudiados. No obstante, el grupo de J. W. Han et al. (2021) ha encontrado, recientemente, resultados interesantes respecto a su papel en SOCE a través de la interacción con las proteínas ORAI1 y STIM1 y la formación de heterodímeros con ORAI1, su localización celular e, incluso, de su expresión diferencial respecto del tejido de cáncer de colon. En primer lugar, a través de estudios de fluorescencia (fusión de GCaMP con MS4A12) encontraron que puede formar tanto homodímeros como heterodímeros con ORAI1 en la MP. Además, también encontraron que esta interacción aumenta tras el vaciado de los depósitos inducido por tapsigargina. No obstante, la coexpresión de MS4A12 junto con STIM1 inhibió la formación de homodímeros de MS4A12 tras el uso de tapsigargina los cuales se recuperan tras la actuación de ARNs frente a STIM1. Además, la expresión de ORAI1 disminuye la interacción entre STIM1 y MS4A12, por lo que la unión de STIM1 a esta está inhibida de forma competitiva por ORAI1 (J. W. Han et al. (2021)). Cabe destacar que este tipo de interacciones también se han observado con otros canales no canónicos de SOCE como los TRPC (J. P. Yuan et al. (2007)). Por lo tanto, parece evidente que MS4A12 forma un complejo con ORAI1 implicado en SOCE la cual puede ser inhibida por la sobreexpresión de STIM1.

2.6.6. Moduladores de SOCE

Se han descrito diferentes proteínas que son capaces de regular la actividad de diferentes miembros de SOCE y que pueden jugar un papel clave en este mecanismo de transporte. A continuación, se expondrán brevemente algunas de ellas (figura 2.16).

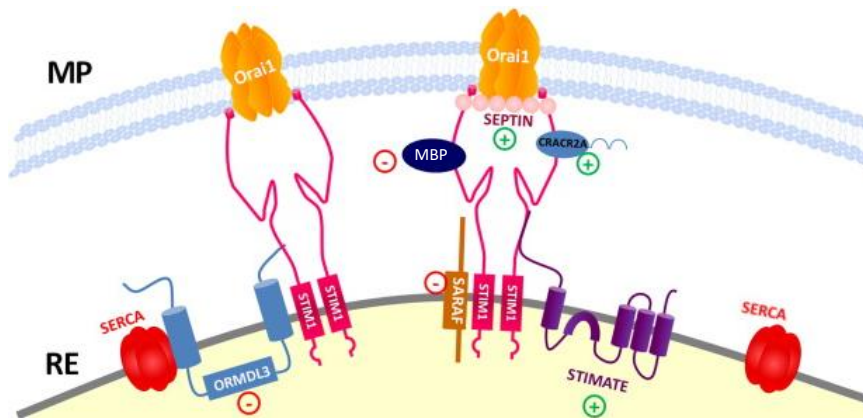


Figura 2.16: Moduladores de SOCE. Se representan algunos de los reguladores de SOCE mejor conocidos. El signo indica si estos son reguladores positivos o negativos de SOCE (modificado de López et al. (2016)).

2.6.6.1. CRACR2A

En 2007, Várnai et al. (2007) describieron que la interacción ORAI1-STIM1 requiere un espacio entre las dos membranas (RE y MP) superior al que posee STIM1 en su cara citosólica, lo que indica que ORAI1 debía ser parte de un complejo macromolecular con una protuberancia de entre 11 y 14 nm en el citoplasma frente a los 5-6 nm del dominio citoplasmático de STIM1. En consecuencia, debía existir algún componente desconocido en el complejo STIM1-ORAI1 que regulase su interacción. Así, Srikanth et al. (2010) encontraron que la proteína de unión a Ca^{2+} CRACR2A (regulador 2A de CRAC) es un regulador de la interacción entre ORAI1 y STIM1 que interactúa con las regiones citoplasmáticas tanto de ORAI1 como de STIM1, lo que da lugar a un complejo ternario. Esta proteína, codificada por el gen *CRACR2A* del cromosoma 12 y localizada predominantemente en el citosol, contiene dos dominios manos EF en su extremo N-terminal y, en su extremo C-terminal, un dominio de *hélices arrolladas* o *coiled-coil* y un dominio rico en leucina altamente conservado en los vertebrados. Mediante estudios con siARN y mutagénesis del dominio EF de CRACR2A se ha demostrado que esta proteína es clave en la formación del complejo STIM1-ORAI1 tras el vaciado de los depósitos. Por lo tanto, en cuanto a su interacción con ORAI1 y STIM1, por un lado, CRACR2A interactúa físicamente con residuos cargados 64-93 positivamente del extremo N-terminal de ORAI1 en una forma sensible a Ca^{2+} . Por otro lado, interactúa con la región *coiled coil* y la región rica en prolina/lisina de STIM1, pero no con la región rica en setina/treonina, la cual solapa con el dominio de multimerización de STIM1 y SORA. Además, este mecanismo de unión entre CRACR2A parece estar conservado en las otras dos isoformas de ORAI, ORAI2 y ORAI3. Por otro lado, CRACR2A interactúa con los dominios *coiled-coil* y secuencias ricas en prolina y lisina de STIM1 (Srikanth et al. (2010)).

Se ha observado una relación entre la actividad física y la expresión de este gen, pues a mayor actividad física menor metilación de su región promotora y mayor expresión de este. Es más, la proteína que codifica está involucrada en la respuesta inmune y la desgranulación de los neutrófilos, dos procesos relacionados directamente con la actividad física (Dai et al. (2019a), Dai et al. (2019b)). Por ejemplo, en las células T, en las que SOCE es un mecanismo de transporte de Ca^{2+} sumamente importante, hay una expresión elevada de CRACR2A el cual participa directamente en la regulación de este mecanismo de transporte (Srikanth et al. (2010), Dai et al. (2019a)). En otros tipos celulares, como por ejemplo las células endoteliales, se ha descrito que en experimentos de depleción de CRACR2A mediante ARN de interferencia diseñado específicamente para esta proteína no se observa efecto alguno sobre los canales CRAC, pero sí una disminución de una proteína con, aproximadamente, el doble de masa que esta. Es más, en estas mismas células se ha observado la presencia de una variante larga de esta proteína, CRACR2A-L, cuyo dominio N-terminal es idéntico a CRACR2A y su extremo C-terminal le convierte en un miembro de la superfamilia de las GTPasa Rab, de tal manera que este extremo C-terminal impide que participe en SOCE y, además podría participar en la formación del tubo endotelial, es decir, en la angiogénesis (Wilson et al. (2015)). No obstante, a pesar de que en las células endoteliales sólo se expresa CRACR2A-L, la regulación de la forma corta de *splicing* alternativo podría dar lugar a la coexpresión de ambas en ciertos contextos, de tal manera que CRACR2A regule positivamente

los canales CRAC, de tal manera que el aumento de $[Ca^{2+}]_{cit}$ module, a su vez, la actividad de la forma larga CRACR2A-L (Wilson et al. (2015)).

Srikanth et al. (2010) propone que en condiciones basales CRACR2A se encuentra en el citoplasma y tanto ORAI1 como STIM1 están en las membranas plasmática y del RE, respectivamente, sin formar complejos ni oligomerizar con ellos mismos. No obstante, tras el vaciado de los depósitos, STIM1 y ORAI1 se reagrupan hacia zonas de contacto entre las membranas del ER y la MP, proceso en el que CRACR2A podría participar ya sea favoreciendo la oligomerización de STIM1, la de ORAI1, la de ambos, o que, simplemente, se una de forma pasiva a los clústers de STIM1 y ORAI1. Lo que sí que parece estar más claro es que esta proteína estabiliza el complejo ORAI1-STIM1. De esta forma, se genera un complejo ternario que permanece unido hasta que aumenta $[Ca^{2+}]_{cit}$, momento en el que se disocia el CRACR2A de dicho complejo tras la unión de este Ca^{2+} a las manos EF del CRACR2A. Además, también se ha observado que la unión de CaM al extremo N-terminal de ORAI1 da lugar a la inactivación de los canales CRAC y, curiosamente, estos sitios de unión a CaM tanto de STIM1 como de ORAI1 se solapan con las regiones de unión a CRACR2A. Es más, debido a que CRACR2A y CaM tienen una afinidad similar por Ca^{2+} , a bajas $[Ca^{2+}]_{cit}$ CRACR2A estaría activando ORAI1 mientras que al aumentar más esta concentración CaM ejercería un efecto inactivador dependiente de Ca^{2+} de ORAI1. Por lo tanto, el papel del CRACR2A como modulador del complejo STIM1-ORAI1 en la activación de SOCE podría ser dependiente de Ca^{2+} , si bien es cierto que aun es necesaria una mayor investigación en cuanto a su papel modulador de estos canales. Cabe destacar que estudios de sobreexpresión de esta proteína en células Jurkat y HeLa aumenta SOCE inducida por tapsigargina en aquellas células que también sobreexpresan ORAI1 y STIM1, pero no únicamente ORAI1, por lo que no actúa como sensor de Ca^{2+} capaz de activar SOCE por sí solo (Srikanth et al. (2010)).

En conclusión, el papel de CRACR2A a través de su unión a ORAI1 y STIM1 es polifacético. Por un lado, podría estabilizar la interacción entre ambos en áreas de contacto entre la MP y el RE, por lo que juega un papel clave en la formación del complejo STIM1-ORAI1 tras el vaciado de los depósitos. Por otro lado, a diferencia de STIM1, CRACR2A no es suficiente para inducir grandes corrientes CRAC cuando se co-expresa con ORAI1, lo que indica que no es un sustituto de STIM1. Sin embargo, sí actúa como sensor de Ca^{2+} gracias a sus dominios mano EF y, por ende, también participa en SOCE ya que, ante el aumento de $[Ca^{2+}]_{cit}$, esta se disocia tanto de ORAI1 y STIM1 (Srikanth et al. (2010), Dai et al. (2019a)). Sin embargo, Srikanth et al. (2010) propone que el papel de CRACR2A podría ser relevante a unos niveles fisiológicos de expresión de ORAI1 y STIM1, pero no cuando estos están sobreexpresados.

2.6.6.2. Septinas

Son GTPasas que forman complejos filamentosos entre sí que dan lugar a estructuras más complejas. Existen cuatro subgrupos en función de su homología de secuencia: SEPT2 (1, 2, 4, 5), SEPT6 (6, 8, 10, 11, 14), SEPT7 y SEPT3 (3, 9, 12) (Neubauer and Zieger (2017)). Los miembros de cada uno de los grupos ocupan diferentes posiciones en el filamento, pues la estructura hexamérica del oligómero que forman SETP7, SEPT6 y SEPT2 muestra que los miembros del subgrupo SEPT2 ocupan el lugar central mientras que los del subgrupo SEPT7 ocupan las posiciones terminales. En humanos, las septinas se ensamblan, por lo general, en heterphexámeros o heterooctámeros, donde estos últimos contienen miembros del subgrupo SEPT3 (Neubauer and Zieger (2017)). Cabe destacar que los miembros de un mismo subgrupo, en los heterómeros, pueden sustituirse entre ellos en la misma posición. En función del tipo celular pueden tener diferentes funciones. Por ejemplo, la disminución de SEPT7 en fibroblastos bloquea la división celular, pero no en linfocitos. Además, los linfocitos T, si están en contacto con otras células no requieren de septinas para la división celular, pero sí en el caso de que estén aisladas. Otro tipo de funciones en las que participan estas proteínas son la dinámica de la actina, la regulación de los microtúbulos, el tráfico a través de la membrana, el transporte vesicular, la exocitosis, ensamblado de plataformas de proteínas adaptadoras, degradación proteica o estabilidad mecánica (Neubauer and Zieger (2017)). También se les ha asociado con diferentes enfermedades humanas, como neurodegenerativas o cáncer. Por ejemplo, se ha encontrado que la desregulación de las septinas pueden tener un papel relevante en el proceso tumoral. Por ejemplo, la SEPT9 favorece la transcripción mediada por HIF-1a en las células tumorales (Neubauer and Zieger (2017)).

En cuanto a la homeostasis del Ca^{2+} intracelular, se ha observado, en células Jurkat, que la pérdida de SEPT2, SEPT4 y SEPT5, miembros del subgrupo SEPT2, conlleva la reducción de SOCE, por lo que son reguladores positivos de esta. Esto mismo se observa en *Drosophila* tras la delección de dSEPT1 y dSEPT4 en neuronas. Además, los miembros del subgrupo SEPT2 actúan como factores de nucleación para la formación de los complejos filamentosos de septinas. Además, también es imprescindible un miembro del subgrupo SEPT6 para que el complejo sea funcional. En neuronas de *Drosophila*, la reducción de dSEPT7 (único miembro del subgrupo SEPT7, al igual que en humanos) parece no alterar SOCE, pero sí cuando IP3R está mutado o reducido, o dSTIM1 está reducción mejora SOCE sin la necesidad del vaciado total de los depósitos actuando como regulador negativo de dORAI. Por lo tanto, podría favorecer una conformación de dORAI que permita la entrada de Ca^{2+} independiente de los depósitos. Cabe preguntarse cómo es posible que un complejo formado por SETP7, SEPT6 y SEPT2, pertenecientes a dos grupos con efectos contrarios sobre SOCE, lo que podría deberse al orden en que se ensamblan las diferentes septinas en la formación de los complejos filamentosos y la estructura a la que da lugar (Deb and Hasan (2016)).

La función de los complejos filamentosos de septinas, más allá de actuar como reguladores de SOCE, sería la de favorecer la formación de dominios lipídicos en la MP, de tal manera que podrían ayudar a secuestrar ORAI y no permitir la interacción STIM-ORAI. Esto se debe a que son capaces de crear barreras de difusión a través de su interacción con el PIP2 de la MP. Sin embargo, la

pérdida parcial de dSETP7, respecto de la expresión de dSEPT1/4 (subgrupo SEPTN2) y dSEPT2/5 (subgrupo SEPTN6) da lugar a filamentos más cortos que favorecen la reorganización de la MP y permite la interacción STIM-ORAI. Es más, la pérdida parcial de dSEPT7 da lugar a un ORAI constitutivamente activo y aumenta la oligomerización de ORAI. En el caso de mamíferos, se ha visto que la eliminación de SEPT2/4/5 da lugar a una oligomerización aberrante de ORAI y, en consecuencia, a la reducción del acoplamiento STIM/ORAI (Deb and Hasan (2016)).

Las septinas se localizan cerca de la MP y de la membrana del RE y regulan el reclutamiento de STIM hacia las zonas de contacto RE-MP. En las neuronas de *Drosophila* se ha observado que la reducción de dSEPT7 favorece el acercamiento de dSTIM1 a la MP. Además, en mamíferos, la delección de los miembros de SEPT2 conlleva un retraso en el reclutamiento de STIM a las regiones de contacto de RE-MP tras el vaciado de los depósitos (Deb and Hasan (2016)).

2.6.6.3. ORMDL3

La isoforma 3 de la proteína reguladora de la biosíntesis de esfingolípidos similar al orosomucoide (ORMLD3) está localizada en la membrana del RE cuyos extremos N y C-terminal se localizan en la cara citoplasmática. Alteraciones en esta proteína se ha asociado con enfermedades inflamatorias como el asma. Esta proteína regula la actividad de la SERCA, por lo que es capaz de provocar estrés del RE alterando la señalización del Ca^{2+} . Además, también es un regulador negativo de SOCE, pues se ha encontrado que su sobreexpresión en células Jurkat y HEK-293 disminuye I_{CRAC} y, en consecuencia, SOCE. Por el contrario, su silenciamiento aumenta I_{CRAC} y SOCE. En particular, es el extremo N-terminal de esta proteína quien juega un papel clave en esta regulación de SOCE, y lo hace a través de la inducción de la inactivación lenta dependiente de Ca^{2+} . Cabe destacar que STIM1 y ORMDL3 colocalizan en la membrana del RE en condiciones basales y en las estructuras de punctas tras el vaciado de los depósitos (Lopez et al. (2016)).

El ORMDL3 sobreexpresado es capaz de unirse a SERCA2b e inhibirla, por lo que se reduce la liberación de Ca^{2+} desde el RE, ya que entra menos. Es más, se ha descrito que ORMDL3 tiene un papel importante en la regulación de ATF6, el cual activa genes diana de la respuesta a proteínas mal plegadas que tiene lugar durante el estrés del RE, y que su sobreexpresión en las células epiteliales humanas incrementa ATF6 α y regula positivamente la expresión de *SERCA2b*, pues es uno de los genes diana de ATF6 α , y se favorece la expresión de quimioquinas y MMP (James, Milstien, and Spiegel (2019)). Ante la presencia de proteínas mal plegadas, se activa un mecanismo de respuesta a proteínas mal plegadas (UPR) que, en mamíferos, engloba a tres rutas de señalización: ATF6, PERK e IRE1. La activación de estas rutas conlleva el aumento de la capacidad de plegado de proteínas en el RE mediante el aumento de chaperonas, a la disminución de la traducción de proteínas y la activación de mecanismos de degradación de proteínas e, incluso, de la autofagia. De hecho, las tres rutas de señalización de UPR regulan la autofagia. No obstante, cuando el estrés del RE es excesivo y los mecanismos activados anteriormente no son suficientes para recuperar la homeostasis celular, o cuando el estímulo que ha generado el estrés del RE no cesa, se desencadenan procesos tales como la apoptosis y respuestas inflamatorias.

2.6.6.4. SARAF

El factor regulador asociado a SOCE (SARAF, del inglés *SOCE-associated regulatory factor*) es una proteína de 339 residuos con un único dominio TM localizada en el RE (aunque también existe evidencia de su localización en la MP), cuyo dominio N-terminal se localiza en la cara luminal del RE, mientras que su extremo C-terminal, el cual contiene regiones ricas en serina/prolina y arginina, lo hace en la citosólica. Así, mientras que el extremo N-terminal regula la actividad de la propia proteína, el C-terminal es capaz de regular negativamente SOCE a través de la interacción con STIM (STIM1 y STIM2) y evita su oligomerización. Además, en caso de que esté oligomerizado por haberse activado previamente tras el vaciado de los depósitos, SARAF promueve la inactivación lenta de SOCE a través de la desoligomerización de STIM tras el llenado de los depósitos con el fin de evitar un llenado excesivo. En particular, a través del dominio inhibitorio de STIM1 en su extremo C-terminal (CTID) que, a su vez, contiene dos zonas funcionales (448-490 y 490-530) que cooperan para favorecer la interacción de SARAF con la región SOAR en condiciones basales para evitar la activación espontánea de SOCE. No obstante, el vaciado de los depósitos conlleva la disociación de SARAF del dominio SOAR lo que favorece SOCE a través de STIM1-ORAI1 y, acto seguido, la zona funcional 490-530 de STIM1 interacciona nuevamente con SARAF que da lugar a la inactivación lenta de SOCE dependiente de Ca^{2+} . De esta manera, la regulación negativa de SOCE por parte de SARAF evita que la célula se sobrecargue de Ca^{2+} tras llenado de los depósitos a través de una inactivación lenta de la actividad de SOCE dependiente de STIM.

El SARAF localizado en la MP puede regular de forma positiva la entrada de Ca^{2+} a través de ORAI1 independiente de depósitos y dependiente de ARC, pues está involucrado en la formación de los canales de ARC, formados por ORAI1:ORAI3 en relación 3:2 que, además, requieren de STIM1 localizado en la MP. Se ha descrito que el ácido araquidónico induce la proliferación celular a través de una entrada que aumenta $[Ca^{2+}]_{cit}$, lo que podría explicar la mayor $[Ca^{2+}]_{cit}$ encontrada en las HT29 (Alessandra Fiorio Pla et al. (2008)). Cabe destacar que la entrada de Ca^{2+} a través de los canales ARC regulados por SARAF modulan la supervivencia de las células de neuroblastoma (Lopez et al. (2016)).

Cabe destacar que en células en las que la expresión de STIM1 está disminuida, SARAF interactúa con ORAI1 a través del extremo C-terminal de ORAI1 de forma transitoria y lo activa tras la estimulación con determinados agonistas, de tal manera que alcanza un máximo de 30 segundos de activación tras la estimulación (Lopez et al. (2016)). De hecho, se ha encontrado que SARAF y ORAI1 contribuyen a la activación de las células endoteliales y a la angiogénesis a través de una interacción funcional entre ellos por la que, en respuesta a VEGF, aumentan $[Ca^{2+}]_{cit}$ (Galeano-Otero et al. (2021)).

2.6.6.5. SIMTATE

Este regulador positivo de SOCE es una proteína localizada en la membrana del RE que participa en la activación y traslocación de STIM1 hacia las regiones de contacto RE-MP y, por ende, favorece la formación de las puntas que acaba por activar los canales CRAC (Lopez et al. (2016)). Se ha observado que el silenciamiento de este regulador conlleva a la disminución de la activación farmacológica de SOCE. De hecho, su delección reduce drásticamente la formación de puntas tras el vaciado de los depósitos, de tal manera que es clave en la formación de estas estructuras.

2.6.6.6. MBP

Golli es una isoforma de la proteína de mielina básica (MBP, del inglés *myelin basic protein*) generada por *splicing* alternativo del gen *MBP*, se expresa en diferente tejidos en humanos, tales como el corazón, riñón o pulmón. Se ha observado que la disminución de esta proteína puede incrementar la entrada de Ca^{2+} por SOCE y, por el contrario, que su sobreexpresión es capaz de inhibirla. Por lo tanto, se considera un regulador negativo de este mecanismo de transporte de Ca^{2+} . En concreto, esta proteína con el extremo C-terminal de STIM1 y participa en la inactivación dependiente de Ca^{2+} de los canales CRAC, por lo que ante el aumento de Ca^{2+} citosólico, esta proteína interacciona con los complejos STIM1-ORAI1 para disminuir SOCE (Lopez et al. (2016)).

2.7. Superfamilia de los TRP

Los TRP son una superfamilia de canales expresados en diferentes organismos pluricelulares. Estos canales se expresan tanto en la MP como en la de orgánulos intracelulares en diferentes tipos celulares. En la década de los 60, Cosens and Manning (1969) identificaron por primera vez a uno de los miembros de esta familia en *Drosophila*. Observaron que este tenía afinidad por Ca^{2+} y que se activaba a través de la PLC, por lo que se pensó que podría participar en SOCE. En 1995 se identificó el primer homólogo en vertebrados, TRPC1. En concreto, en humanos hay un total de 27 TRP divididos en 6 familias (figura 2.17): TRPC (canónicos), TRPM (melastatina), TRPML (mucolipina), TRPV (vaniloides), TRPA (anquirina) y TRPP (policistina) (Venkatachalam and Montell (2007), Hardie (2011), Samanta, Hughes, and Moiseenkova-Bell (2018)).

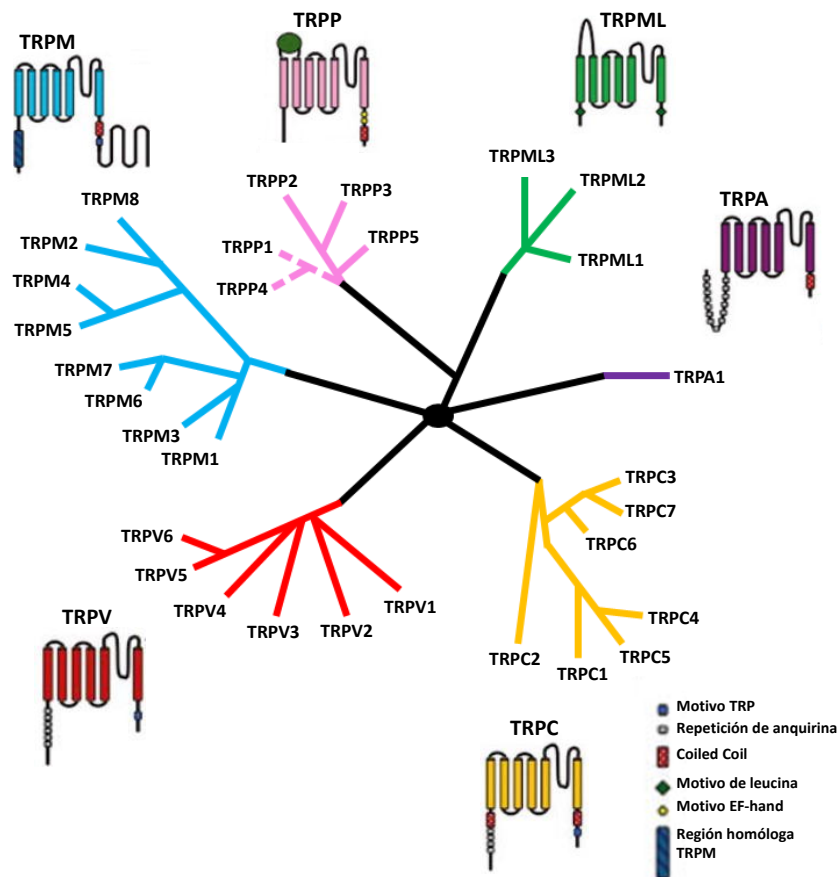


Figura 2.17: Superfamilia de los TRP. En la figura se muestran los 27 TRP humanos agrupados según su análisis filogenético, de tal manera que se agrupan en 6 familias diferentes. De ellos, el TRPC2 resulta ser un pseudogen. Para cada familia se muestra la estructura típica de cada una de las familias, si bien es cierto que todos ellos comparten unas características estructurales comunes: 6 dominios transmembrana cuyos extremos N y C-terminal están en la cara citoplasmática, los dominios TM 5 y 6 forman el poro. No obstante, cada familia presenta diferencias respecto a las otras, especialmente en los extremos N y C-terminal así como en los lazos que unen los diferentes TM: los TRPA, TRPC y TRPV poseen repeticiones de anquirina; los TRPC, TRPV y TRPM poseen el motivo TRP; los TRPML poseen el motivo de leucina; los TRPC, TRPA, TRPM y TRPP poseen regiones *coiled coil*; los TRPP poseen motivos mano EF, y los TRPM presentan la región homóloga TRPM (modificado de Hughes, and Moiseenkova-Bell (2018)).

Estos canales, poseen una estructura similar a los canales de K^+ operados por voltaje que consiste en 6 dominios TM (TM1-TM6), de los cuales son el 5 y el 6 quienes forman el poro (figura 2.17). Tanto sus extremos C y T-terminal como demás dominios estructurales se localizan en el citosol y difieren entre las distintas familias (Venkatachalam and Montell (2007), Hardie (2011), Mesquita et al. (2021)). Muchos de los TRP estrechamente relacionados son capaces de formar, a parte de homómeros, heterómeros, lo que les permite modular su función, su localización subcelular y las propiedades de los canales con quienes interactúan como sus conductancias. Por lo general, estas interacciones heteromultiméricas se favorecen entre miembros que están estrechamente relacionados entre sí (Venkatachalam and Montell (2007)).

Además de poder formar homo y heterómeros, estos canales también se asocian con diferentes proteínas andamio y les permite unirse a diferentes complejos supramoleculares para controlar su distribución subcelular. Por ejemplo, el TRPC3 e IP3R forman un complejo con este tipo de proteínas que podría modular tanto el tráfico como la apertura de los IP3R (Venkatachalam and Montell (2007)).

Los TRP son receptores no selectivos a cationes que dejan pasar Na^+ , K^+ y Ca^{2+} , a excepción de los TRPM4 y TRPM5 que no permiten el paso de este último (Hasan and Zhang (2018), Lefranc (2021)). Se ha encontrado que la CaM está involucrada en la regulación dependiente de Ca^{2+} de los TRPV, TRPM, TRPA1 y TRPC. Además, el propio Ca^{2+} también es capaz de regular los TRP por una unión directa a los TRP y por la activación de enzimas dependientes de Ca^{2+} (Hasan and Zhang (2018)). Son receptores polimodales que participan en numerosas funciones fisiológicas y patológicas en respuesta a diferentes estímulos: variaciones en la temperatura, pH, osmolaridad, presión, drogas, citoquinas y, por supuesto, al vaciamiento de los depósitos de Ca^{2+} . Además, también se ha observado la activación de algunos de estos canales bajo condiciones de hipoxia y se les ha relacionado con el estrés oxidativo y la muerte celular (Venkatachalam and Montell (2007), Hardie (2011), Mesquita et al. (2021)).

Estos canales se han relacionado con diferentes patologías como son enfermedades neurodegenerativas, cardiovasculares, metabopatías y, en especial, cáncer (Rizopoulos and Assimakopoulou (2021)). A continuación, se realizará una breve exposición de los TRP descritos en humanos dividida en función de las diferentes familias.

2.7.1. TRPC

La familia de los TRPC son un grupo de canales no selectivos para Ca^{2+} que debe su nombre a que son considerados como los TRP canónicos, pues fue el TRPC1 el primer miembro de esta superfamilia descubierto en mamíferos. Está constituida por siete miembros, TRPC1-7, de los cuales los humanos solo expresan 6 ya que el TRPC2 es un pseudogen. Además, en función de su similitud de secuencia, estos se dividen en cuatro subfamilias: TRPC1, TRPC2, TRPC3/6/7 y TRPC4/5 (Venkatachalam and Montell (2007), Mesquita et al. (2021), Lefranc (2021), Rizopoulos and Assimakopoulou (2021)).

Al igual que otros TRP, los TRPC pueden formar tanto homo como heterómeros, por lo general, con aquellos con los que guarda mayor relación. Cabe destacar que todas estas posibles combinaciones dan lugar a un amplio espectro de canales con propiedades y funciones muy diferentes. TRPC1 está más relacionado con TRPC4 y TRPC5; por ejemplo, TRPC1 y TRPC5 forman canales heteroméricos cuando se coexpresan y dan lugar a diferentes conductancias y propiedades diferentes respecto de los homodímeros correspondientes. También se han detectado heterómeros de TRPC1 con los miembros de la familia TRPC3/6/7. En definitiva, hay gran evidencia de interacciones entre TRPC1, TRPC3/6 y TRPC4/5 que da como resultado canales con diferentes funciones y propiedades biofísicas (Venkatachalam and Montell (2007)).

Se expresan tanto en células excitables como en no excitables. No obstante, mientras que TRPC1 se expresa ampliamente en distintos tejidos, TRPC3 y TRPC5 se expresa de forma predominante en el cerebro, lo que podría explicar su relación con diferentes patologías relacionadas con enfermedades neurodegenerativas como el Alzheimer o el Parkinson. También se expresan ampliamente en diferentes tipos de cáncer como, por ejemplo, en el cáncer de colon (Rizopoulos and Assimakopoulou (2021), Mesquita et al. (2021)).

Se han propuesto dos mecanismos de activación de estos canales: a través del DAG generado por la PLC o a mediante la liberación de Ca^{2+} de los depósitos (Lefranc (2021)). Por un lado, el mecanismo de activación a través de receptores operados por Ca^{2+} desencadena la activación de los TRPC a través de los GPCR y de los TKR. Así, en presencia de un estímulo, estos receptores se activan y reclutan la PLC, hidrolizan el PIP2 para dar IP3 y DAG, este último capaz de interactuar y activar los TRPC (Vazquez et al. (2001), Trebak et al. (2002)). Por otro lado, el IP3 se dirige al RE para activar los IP3R e inducir la liberación de Ca^{2+} hacia el citosol, por lo que se activará SOCE en la que, entre otros canales, participan los TRPC (Putney Jr (2004), Venkatachalam et al. (2001) Venkatachalam and Montell (2007)).

Tienen una amplia variedad de funciones celulares y regulación de diferentes procesos, como la proliferación y migración celular, la plasticidad sináptica, desarrollo neuronal, secreción de hormonas y neurotransmisores, activación de células T (Rizopoulos and Assimakopoulou (2021))... Muchos miembros de la familia TRPC participan en el mantenimiento de la homeostasis del Ca^{2+} intracelular. Además, el bloqueo de estos canales reduce de forma significativa la proliferación tumoral. Por ello, los TRPC se proponen como dianas terapéuticas para el tratamiento del cáncer. Los más estudiados en este área son TRPC1 y TRPC6. Cabe destacar que se ha encontrado fuerte evidencia de que el proceso de restitución epitelial, proceso en el que puede activarse de forma transitoria tanto la migración como la proliferación celular, podría estar mediado por la entrada de Ca^{2+} a través del TRPC1 (Gutiérrez et al. (2019)).

El TRPC1, que es el TRPC más estudiado, participa de forma activa en la entrada de Ca^{2+} desde el medio extracelular. Se ha descrito que TRPC1 podría formar complejos con ORAI1 y STIM1 (Ong et al. (2007)), y con NCX1 (J. Xu et al. (2018)). Es más, este canal, a través de su interacción con STIM1 y ORAI1 puede participar en SOCE (Xibao Liu, Singh, and Ambudkar (2003), J. P. Yuan et al. (2007), Sobradillo et al. (2014)). No obstante, el aumento en la relación STIM2/STIM1

favorece la formación de heterómeros entre estas dos isoformas que suprime la traslocación de STIM1 a la MP y, en consecuencia, se inhibe la interacción con TRPC1 (Sobradillo et al. (2014)). En cuanto a los miembros de la subfamilia TRPC4/5, estos también participan en SOCE a través de su interacción con STIM1 (Kretz et al. (2000), J. P. Yuan et al. (2007)). Cabe destacar que TRPC4 es activado en respuesta al estrés oxidativo (Duitama et al. (2020)).

La subfamilia TRPC3/6/7, aparte de ser activados por DAG generado por la PLC, la evidencia indica que podrían estar implicados en SOCE, al menos en combinación con TRPC1 formando un complejo como TRPC1/3/7 (Zagranichnaya, Wu, and Villereal (2005), J. P. Yuan et al. (2007), Brécharde et al. (2008), Alkhani et al. (2014)). Por ejemplo, el TRPC1 puede actuar como un regulador negativo del TRPV6, canal selectivo por Ca^{2+} activado por el vaciamiento de los depósitos (Schindl et al. (2012)). También, en líneas celulares de cáncer de mama, el TRPC6 podría no solo permitir la entrada de Ca^{2+} tras ser activado por DAG, sino que también podría activar SOCE, pues el flujo de Ca^{2+} a través del TRPC6 podría ser necesario para la adecuada localización de ORAI en la MP (I. Jardin et al. (2018)). De los tres, la menor evidencia recae sobre TRPC7 y la más fuerte sobre el TRPC3. Sin embargo, grupos como el de DeHaven et al. (2009) han mostrado que TRPC7 no se activa directamente por el vaciado de los depósitos, sino por la inhibición de SERCA de una forma independiente de STIM1 y ORAI1. Además, en las células esqueléticas se ha observado que TRPC3 también juega un papel importante en la migración a través de una interacción con los canales de K^+ sensibles a Ca^{2+} de conductancia intermedia, $\text{K}_{\text{Ca}3.1}$ (Mesquita et al. (2021)). También se ha observado que el estrés oxidativo provoca una expresión constitutiva de TRPC3 (Venkatachalam and Montell (2007)). Por el contrario, el canal TRPC6 se ha relacionado con diferentes patologías como el cáncer, la glomerulosclerosis segmental y focal, causada por mutaciones en TRPC6, que da lugar a una brecha en la barrera de permeabilidad en los pilares del riñón que favorece la progresión de la enfermedad (Venkatachalam and Montell (2007)).

De forma general, se ha observado que STIM1 puede unirse a TRPC1, TRPC4 y TRPC5. STIM1 también es capaz de regular TRPC3 y TRPC6, pues no interactúa directamente con ellos, pero sí facilita la formación de los heterodímeros TRPC3/TRPC1 y TRPC6/TRPC4. No obstante, STIM1 podría ser necesario para (o al menos regular) la activación de los TRPC a través de agonista pero no para que estos receptores ejerzan su función. Por un lado, experimentos de sobreexpresión de TRPC1, TRPC4 y TRPC5 no mostraron una activación de estos canales independiente de STIM1. Por otro lado, la sobreexpresión de TRPC3 y TRPC6 sí que mostraron actividad independiente de STIM1 y sin necesidad de que este mediase la formación de los heterodímeros TRPC3/TRPC1 o TRPC6/TRPC4. Es más, la delección de STIM1 mostró una inhibición en la activación de TRPC5 a través de la estimulación por agonista, pero no por La^{3+} (J. P. Yuan et al. (2007)). Cabe destacar que la isoforma STIML1 se asocia con los TRPC1, TRPC3, TRPC4 y TRPC6, y que, además, activa preferentemente a TRPC1 frente a ORAI1 (Grabmayr, Romanin, and Fahrner (2021)).

Debido a su capacidad para formar heterómeros, la sobreexpresión de algunos TRPC puede dar lugar a diferentes patrones de regulación de su actividad, dependiendo tanto de la célula en la que se de esté sobreexpresión como de los niveles de expresión basal, al igual que ocurre, por ejemplo, al sobreexpresar ORAI1,2. Esto podría deberse a diferentes formas de ensamblaje que podrían dar lugar a diferentes tipos de canales o a complejos ensamblados de forma inapropiada con diferencias en la regulación de su actividad (Putney Jr (2004)) (figura 2.18). Por ejemplo, se han descrito diferentes patrones de regulación de la actividad del TRPC3 en diferentes líneas celulares, como las células de riñón HEK293 o las DT40 pre-B, pues hay estudios que apoyan una regulación a través del vaciado de los depósitos (Vazquez et al. (2001), Trebak et al. (2002)) mientras que otros proporcionan evidencia de una regulación por PLC (Putney Jr (2004), Venkatachalam et al. (2001)).

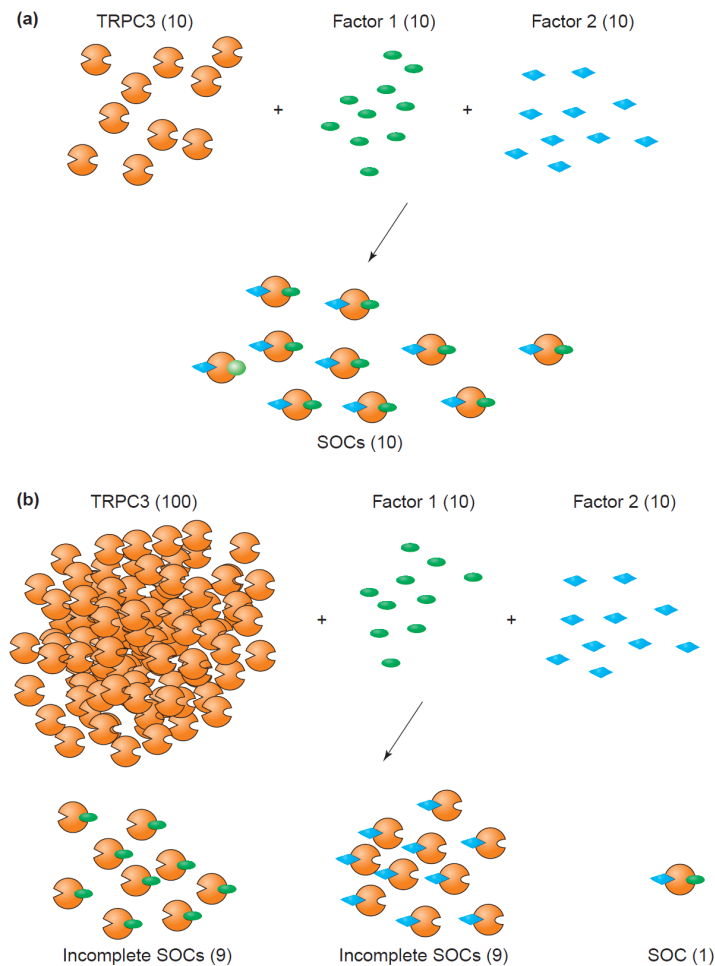


Figura 2.18: Sobreexpresión de componentes de complejos señalizadores. La sobreexpresión de alguno de los componentes de complejos señalizadores puede resultar en que estos se ensamblen de diferentes formas. En la figura se muestra un complejo para un complejo en el que TRPC3 está implicado, formado dicho complejo por un total de tres componentes: a) niveles bajos de expresión de TRPC3, se forman complejos estequiométricamente adecuados, b) a niveles altos de expresión, la probabilidad de que se los componentes se ensamblen adecuadamente para formar el complejo es baja, pudiendo dar lugar a complejos incompletos por repartirse el resto de componentes entre todos los TRPC3 sobreexpresados (modificado de Putney Jr (2004)).

2.7.2. TRPV

La familia TRPV está constituida por 6 miembros, TRPV1-6, y reciben su nombre por poder ser activados por compuestos vaniloideos. También se les denomina canales termoiónicos por detectar cambios térmicos en el ambiente. Son seis canales de Ca^{2+} no selectivos, excepto TRPV6, que forman homo o heterotetrámeros y que, por lo general, se localizan en la MP. Cabe destacar que los TRPV1-4 también se expresan en el RE (Rizopoulos and Assimakopoulou (2021), Lefranc (2021)). Estos se expresan en diferentes tejidos como, por ejemplo, en neuronas del sistema periférico, cerebrales y ganglios sensoriales (TRPV1-4), piel (TRPV1, 3, 4), páncreas (TRPV1, 5, 6), tracto gastrointestinal (TRPV1-6) (Hasan and Zhang (2018), Rizopoulos and Assimakopoulou (2021))... En cuanto a su activación, se ha propuesto que los canales TRPV4 y TRPV6 están implicados en la SOCE (X. Ma et al. (2011), Abeele et al. (2003)).

La capsaicina es capaz de actuar sobre TRPV1, de tal manera que una activación persistente de este ligando y la entrada excesiva de Ca^{2+} conlleva la desensibilización y la taquifilaxia, proceso dependiente de Ca^{2+} que estaría detrás del efecto analgésico de la capsaicina (Hasan and Zhang (2018)). Se ha implicado a la CaM en la desensibilización del TRPV1 dependiente de Ca^{2+} , de tal manera que un influjo de este catión conlleva la formación de $\text{Ca}^{2+}/\text{CaM}$ que se une al N-terminal de TRPV1 y lo inhibe (Hasan and Zhang (2018)). Se han relacionado diferentes patologías con la disfunción de TRPV1 como dolor inflamatorio, hiperalgesia termal, depresión hipocampal, diabetes, obesidad, hipertensión... También se ha encontrado que es un regulador importante de la expresión de EGFR, de tal manera que la sobreexpresión de TRPV1 disminuye los niveles de este (Mesquita et al. (2021)).

En cuanto al TRPV2, posee un mecanismo de desensibilización dependiente de Ca^{2+} el cual está mediado por una reducción de PIP2 en la membrana como resultado de la entrada de Ca^{2+} a través del canal, lo que puede deberse a que la entrada de este catión a través de TRPV2 active la PLC (Hasan and Zhang (2018)). Se ha observado que es crítico en la fagocitosis de los macrófagos y en la inmunidad innata (Hasan and Zhang (2018)). Sus alteraciones han sido estudiadas en la distrofia muscular de Duchenne, la hipertrofia cardíaca y la miocarditis (Rizopoulos and Assimakopoulou (2021)).

El TRPV3 es un canal termosensible que se activa entre los 30 y 33 °C. Una característica de este canal es que la activación repetida sensibiliza su propia respuesta, proceso que depende de Ca^{2+} , a diferencia de la desensibilización observada en los TRPV1 y TRPV2. Estos mecanismos de sensibilización de TRPV3 podrían explicar la sensibilización de la piel a alérgenos (Hasan and Zhang (2018)). De hecho, el TRPV3 se ha relacionado con desórdenes dermatológicos (Rizopoulos and Assimakopoulou (2021)).

El TRPV4 es otro canal termosensible que se activa tanto por baja osmolaridad como por temperaturas superiores a 27 °C. Participa en la transducción mecánica, el dolor y el picor (Hasan and Zhang (2018)). En estos canales, el Ca^{2+} ejerce un efecto regulador dual, pues, en primer lugar, promueve la activación del canal y, a continuación, lo inactiva. La potenciación de este canal se puede conseguir a través del Ca^{2+} liberado por los depósitos intracelulares o a través de la

entrada de Ca^{2+} extracelular (Hasan and Zhang (2018)). Mutaciones en TRPV4 pueden provocar diferentes enfermedades en humanos que afectan al sistema esquelético y al nervioso periférico, como la enfermedad de Charcot-Marie-Tooth o la atrofia muscular espinal (Hasan and Zhang (2018), Rizopoulos and Assimakopoulou (2021)).

Tanto TRPV5 como TRPV6 tienen una alta selectividad por Ca^{2+} , a diferencia del resto de los TRP que no muestran selectividad por cationes. Tras su apertura, el Ca^{2+} entra e inactiva rápidamente a ambos canales, pero es TRPV6 quien muestra una inactivación más rápida, lo que supone un mecanismo de retroalimentación negativa para evitar una entrada excesiva de Ca^{2+} . No obstante, esta inactivación por Ca^{2+} puede ser modulada por fosforilación mediante la PKC o PIP2 (Hasan and Zhang (2018)). El TRPV5 se encuentra expresado de forma predominante en las células epiteliales y en el riñón, donde juega un papel clave en la reabsorción renal de Ca^{2+} . En cuanto al TRPV6, este se expresa en las células epiteliales del intestino y es el responsable de la absorción de Ca^{2+} (Hasan and Zhang (2018)). La expresión de TRPV5 y TRPV6 está relacionada con la acidosis glomerular, la osteoporosis y la osteomalacia, puesto que ambos están implicados en el metabolismo óseo y la función osteoclástica.

2.7.3. TRPM

Esta familia está formada por 8 miembros, TRPM1-8, que pueden agruparse en 4 grupos: TRPM1/3, TRPM6/7, TRPM4/5 y TRPM2/8. Estos pueden formar homómeros y heterómeros, donde cada subunidad consiste en 6 TM. Todos los TRP, a excepción de los TRPM4 y TRPM5, son permeables a Ca^{2+} . Se expresan ampliamente en una gran variedad de tejidos y tipos celulares, como los órganos sensoriales, las células β pancreáticas, las células del sistema inmune, la lengua, el corazón y el riñón. Algunos tejidos muestran una elevada expresión de algunos miembros de esta familia, como por ejemplo el cerebro (TRPM1,2,3,6), la médula ósea (TRPM2), la pituitaria (TRPM3,7) o el intestino (TRPM4,5,6) (Rizopoulos and Assimakopoulou (2021)). En cuanto a su activación sensorial, algunos de sus activadores más habituales son el calor para el TRPM3, el frío para el TRPM8, el sabor para el TRPM5, la luz para TRPM1, o el estrés oxidativo para TRPM2,4,7.

También participan en la liberación de insulina (TRPM2, TRPM4 y TRPM5), en la homeostasis de Mg^{2+} (TRPM6 y TRPM7), o el daño isquémico y respuestas inflamatorias (TRPM2 y TRPM4). (Venkatachalam and Montell (2007), Hardie (2011), Hasan and Zhang (2018), Lefranc (2021), Rizopoulos and Assimakopoulou (2021)). Los genes que codifican para estos canales son considerados como oncogenes, pues participan en la regulación del crecimiento de las células cancerígenas, la proliferación, la autofagia, la invasión y la EMT. Además, estos canales son una de las dianas terapéuticas de diferentes agentes quimioterápicos contra el CCR, como por ejemplo la ciclofosfamida, el 5-fluorouracil y el leucovorin (Rizopoulos and Assimakopoulou (2021)).

El TRPM2 es un sensor de temperatura cálida y está implicado en la sensación y regulación térmica. Cabe destacar que puede ser activado por ROS, $\text{TNF}\alpha$ y ADP-ribosa los cuales se producen durante el estrés oxidativo e isquemia (Shimizu, Takahashi, and Mori (2014)). Por lo tanto, también

tiene el papel de sensor de estrés oxidativo y está implicado en la muerte celular por necrosis bajo estrés (Venkatachalam and Montell (2007), Lefranc (2021), Rizopoulos and Assimakopoulou (2021)). En consecuencia, la entrada de Ca^{2+} y su señalización a través de la activación de TRPM2 es clave en diferentes procesos patológicos tales como la producción de citoquinas, la muerte celular o enfermedades inflamatorias. El Ca^{2+} es capaz de activar a TRPM2 y, además, mejora la activación por ADP-ribosa (Hasan and Zhang (2018)).

TRPM3 es un sensor de calor nocivo el cual actúa junto con TRPV1 y TRPA1 para la nocicepción del calor. Este es permeable a Ca^{2+} y está constitutivamente activo. La entrada de Ca^{2+} a través de TRPM3 se ve potenciada tras el vaciado de los depósitos, por lo que la activación de este canal es dependiente de Ca^{2+} . En cuanto al papel modulador de los GPCR, en primer lugar, producirían una inhibición rápida tras la liberación de la subunidad $G_{\beta\gamma}$ y, posteriormente, una mejora de la activación de TRPM3 por Ca^{2+} por la activación de la PLC (Hasan and Zhang (2018)).

TRPM4 tiene el papel de sensor de estrés oxidativo. En los mastocitos modulan su migración y desgranulación. Cabe destacar que el TRPM4 es un regulador negativo de SOCE (Kappel, Borgström, et al. (2019)). Además, las poliaminas intracelulares son capaces de inhibir de forma reversible la corriente de TRPM4 (Nilius et al. (2004)). Por ello, en células en las que la síntesis de poliaminas sea elevada, como puede ser en células epiteliales durante la restitución epitelial, TRPM4 se inhibirá y dejará de ejercer su efecto de regulador negativo sobre SOCE y, en consecuencia, SOCE podría aumentar. En cuanto al TRPM5, intervienen en la liberación de opioides y en la detección de nutrientes. El TRPM6 tiene un papel importante en la absorción de Mg^{2+} a través de las células epiteliales del intestino (Rizopoulos and Assimakopoulou (2021)).

En cuanto al TRPM7, es permeable no solo a Ca^{2+} , sino también a Mg^{2+} y Zn^{2+} (Kádár et al. (2021)). De hecho, es clave en la absorción de Mg^{2+} y, además, tiene el papel de sensor de estrés oxidativo y está implicado en la muerte celular bajo estrés (Venkatachalam and Montell (2007), Rizopoulos and Assimakopoulou (2021)). También, gracias a su dominio quinasa del extremo C-terminal participa en procesos tales como la adhesión focal, la adhesión célula-célula o la formación del invadosoma y la movilidad. Respecto de la capacidad invasiva, también participa en la migración celular de las células cancerígenas (Lange and Koomoa (2014)), pues, entre otras funciones, este canal es que participa en la expresión de genes como la MMP13. También se ha descrito su participación como modulador positivo de la entrada de Ca^{2+} a través de ORAI1 en SOCE, si bien es cierto que para ello requiere de una previa activación de los ORAI (Bomfim et al. (2020), Kádár et al. (2021)). Además, también permite la entrada de Ca^{2+} dependiente de pH, pues la bajada de pH, tanto intra como extracelular, activa la entrada de Ca^{2+} a través de TRPM7 (Kádár et al. (2021)). Incluso, las poliaminas son capaces de regular su actividad a través de la interacción con los grupos amino de estas (Lange and Koomoa (2014)). Se ha descrito que la disminución intracelular de poliaminas es capaz de disminuir los niveles de $[\text{Ca}^{2+}]_{\text{cit}}$ con diferentes consecuencias, como la disminución en la liberación de insulina en las células β de los islotes de Langerhans en respuesta a glucosa o una menor liberación de acetilcolina en las células uroteliales de la vejiga en respuesta a oxotremorina (alcaloide sintético que actúa como una droga parasimpaticomimética pues activa los receptores muscarínicos).

En cuanto los TRPM8, este se activan por frío, a unas temperaturas inferiores a 30 °C y compuestos que dan sensación de frío, tales como el mentol (Bidaux et al. (2007)). Este canal es permeable a Ca^{2+} y, además, la entrada de este catión tras su apertura conlleva una rápida desensibilización de TRPM8, lo que podría estar relacionado con la adaptación al frío. En particular, esta desensibilización podría deberse a la activación de la PLC por Ca^{2+} (Hasan and Zhang (2018)). Se ha identificado su expresión no solo en la MP, sino también en la membrana del RE (X.-P. Dong, Wang, and Xu (2010)).

2.7.4. TRPA

En esta familia de TRP sólo hay un único miembro, el TRPA1, que recibe su nombre por contener varias repeticiones de anquirina. Estos canales son capaces de detectar diferentes estímulos nocivos de origen térmico, químico y mecánico, por lo que actúa como un sensor de daño. Además, también es un sensor de estrés oxidativo, pues se también se activa por ROS (Hayes, Dinkova-Kostova, and Tew (2020)). Este canal puede ser modulado negativamente por los protones y por las poliaminas. También se ha observado su activación dependiente de depósitos tras el vaciado con taspigargina (Patil, Jeske, and Akopian (2010)). Está implicado en procesos como el picor, la inflamación y la neuropatía. El Ca^{2+} activa directamente este canal y es esencial para las respuestas de TRPA1 en condiciones basales de este catión. En cuanto a su mecanismo de activación, se ha propuesto que existe una activación dependiente de Ca^{2+} ante un frío nocivo, pero también podría activar TRPA1 el aumento citosólico de este catión debido a la acción de la PLC. Además, la entrada de Ca^{2+} a través de otros canales cercanos como TRPV1 podría potenciar TRPA1, que forman un complejo entre ellos (Hasan and Zhang (2018)). Curiosamente, mientras que las poliaminas potencian la corriente a través de TRPV1 la inhiben a través de TRPA1 (Patil, Jeske, and Akopian (2010), Raisinghani et al. (2011)). Tras su activación, el Ca^{2+} inhibe las respuestas basales de TRPA1 y lo inactiva rápidamente a concentraciones superiores a 1 mM. Por lo tanto, el Ca^{2+} tiene un papel dual en la regulación de este canal (Venkatachalam and Montell (2007), Hardie (2011), Hasan and Zhang (2018)). El TRPA1 también puede actuar como regulador negativo de SOCE al regular negativamente la interacción STIM1-ORAI1 en células megacarioblásticas (Albarrán et al. (2013)). Aunque su papel en el cáncer no está muy claro, se ha encontrado que el silenciamiento de este canal puede mejorar la migración celular a través de un mecanismo independiente de su actividad como canal, pero también puede alterar la progresión del ciclo celular a través de su actividad como canal (Cojocarú et al. (2021)). Además, también se ha visto que su menor expresión en cáncer de colon está relacionado con un peor pronóstico (datos disponibles en <https://www.proteinatlas.org/ENSG00000104321-TRPA1/pathology>, v20.1proteinatlas.org).

2.7.5. TRPP

La familia TRPP, con cuatro miembros, TRPP(1, 2, 3, 5) que reciben su nombre de la enfermedad que provoca una mutación de estos, la policistina. Estos tienen 6 dominios TM. Tienen un papel clave en la mecanotransducción la maduración y diferenciación de los folículos, generalmente ubicados en las membranas intracelulares (Rizopoulos and Assimakopoulou (2021)). Pueden detectar el flujo de los fluidos, la osmolaridad y el estiramiento mecánico (Venkatachalam and Montell (2007)). Están expresados ampliamente en los tejidos humanos; desafortunadamente, TRPP3 y TRPP5 no están bien caracterizados. De hecho, de lo poco que se sabe de ellos, se conoce que son canales permeables a Ca^{2+} (Rizopoulos and Assimakopoulou (2021)).

TRPP1 y TRPP2 actúan como un complejo, pues TRPP1 es necesario para que TRPP2 pueda localizarse en la membrana (Venkatachalam and Montell (2007)). TRPP2 es un canal regulado por Ca^{2+} no selectivo permeable a Ca^{2+} , que puede activarse por factores de crecimiento. Mutaciones en este canal son la causa de la enfermedad poliquística renal autosómica dominante, lo que puede deberse a una señalización aberrante mediante este canal (Hasan and Zhang (2018)). El Ca^{2+} tiene un efecto dual sobre este canal, pues a bajas concentraciones, inferiores a $0.3 \mu\text{M}$, favorecen la activación y a concentraciones elevadas la inhibe (Hasan and Zhang (2018)). Se localiza en el cilio del epitelio renal y en múltiples compartimentos subcelulares como el RE, el aparato de Golgi o la MP; en particular, en el RE puede actuar como un canal de liberación de Ca^{2+} y modula la liberación de este catión mediada por IP3R y RyR y actuar, de igual modo que estos dos, como un canal más del mecanismo CICR y participar en la amplificación de la señal (Venkatachalam and Montell (2007), X.-P. Dong, Wang, and Xu (2010), Rizopoulos and Assimakopoulou (2021)). Además, la sobreexpresión de este canal puede contribuir al menor tamaño de los depósitos y al aumento de $[\text{Ca}^{2+}]_{\text{cit}}$ y a la disminución de $[\text{Ca}^{2+}]_{\text{RE}}$. En cuanto a TRPP1, se expresa en las células epiteliales del túbulo renal así como en otros tipos celulares durante el desarrollo y el crecimiento, pero está ausente en la edad adulta, por ejemplo, en el hígado y el riñón. Este canal participa en la proliferación celular y en las uniones entre células. Alteraciones en la expresión de estos dos canales puede provocar mecanotransducciones alteradas relacionadas con diferentes enfermedades como el cáncer, defectos cardiovasculares, osteoporosis, deformaciones de los huesos y enfermedades inflamatorias (Hasan and Zhang (2018), Rizopoulos and Assimakopoulou (2021)). Se ha identificado la expresión de TRPP1 en la membrana del RE, por lo que podría participar en la liberación de Ca^{2+} desde este (X.-P. Dong, Wang, and Xu (2010)).

En cuanto al TRPP5, se ha encontrado que está expresado durante el desarrollo testicular de los ratones y que su sobreexpresión incrementa la $[\text{Ca}^{2+}]_{\text{cit}}$ en las células MDKC (células adherentes de riñón de perro). También se ha identificado el aumento en la amplitud de la corriente de Ca^{2+} cuando se sobreexpresan en células HEK293 y varía el potencial de reversión hacia valores más negativos y, en consecuencia, se hacen menos permeables a Ca^{2+} . Por lo tanto, el aumento de Ca^{2+} citosólico observado se debe a la liberación de Ca^{2+} desde los depósitos. Además, su sobreexpresión también suprime la proliferación en células Hela a través de la inhibición de la replicación del ADN, pues se detecta la acumulación de células en la fase G1/S, e induce su apoptosis (Xiao et al. (2010)).

2.7.6. TRPML

En los mamíferos, la familia TRPML engloba a 3 miembros, TRPML1-3, y recibe su nombre al gen que los codifica, mucolipina. Se localizan en los endosomas y lisosomas. Estos participan en el control del pH, el tráfico a través de la membrana, la transducción de la señal, la autofagia, la exocitosis y el transporte de vesículas. Mutaciones en TRPML1 o TRPML3 pueden dar lugar a patologías como la mucopolipidosis tipo IV, que se caracteriza por retraso mental, degeneración de la retina, deficiencia de hierro, y anomalías gastrointestinales. Por el contrario, no se conocen enfermedades relacionadas con TRPML2 (Venkatachalam and Montell (2007), Hardie (2011), Rizopoulos and Assimakopoulou (2021), Mesquita et al. (2021)).

El TRPML1 es un canal permeable a protones y con selectividad por Ca^{2+} y Fe^{2+} y, además, es un sensor de estrés oxidativo. También se ha identificado su importante papel en la autofagia al promover la formación del autofagosoma a través de la activación de la ruta $\text{CaMKK}\beta/\text{VPS34}$ (Huang et al. (2020)). Está expresado en los endosomas/lisosomas tardíos, se encuentra expresado de forma ubicua en las células de mamíferos y, en especial, en el cerebro, glándula adrenal, pulmón, corazón, riñón, hígado, sistema inmune... y podría tener un papel importante en la regulación de los linfocitos y en la respuesta inmune. También participa en el control de la viabilidad celular y en la autofagia (Lefranc (2021)). Este puede tanto homo como multimerizar. En cuanto a TRPML3, se localiza en regiones ricas en vesículas de las células de la cóclea, endosomas tempranos, y se distribuye por órganos del sistema endocrino, ojos, tipo, intestino o pulmones. El TRPML2, que es el único para el que no se conocen enfermedades asociadas, tiene una expresión similar a TRPML3 y se ha encontrado en células inmunes, timo, corazón o riñón. Cabe destacar que TRPML1 podría regular los niveles de TRPML2 en los órganos linfoides y el riñón (Rizopoulos and Assimakopoulou (2021)).

Capítulo 3

Remodelado del Ca^{2+} intracelular en el cáncer

En los últimos años el papel del remodelado de la homeostasis del Ca^{2+} intracelular en el cáncer ha cobrado una gran importancia, pues se ha visto que está implicado en varios de sus *hallmarks*: autonomía proliferativa, transcripción de genes, capacidad replicativa ilimitada, resistencia a la muerte, angiogénesis, motilidad, invasión y metástasis... Además, se ha observado un remodelado de los sistemas de transporte de Ca^{2+} intracelular en el proceso tumoral, desde cambios en la expresión de bombas y canales, modificaciones postranscripcionales hasta mutaciones. De hecho, actualmente hay una gran cantidad de trabajos en esta dirección; por citar algunos: Tsavaler et al. (2001), Monteith et al. (2007), Sobradillo et al. (2014), Xie et al. (2016), Hernández-Morales et al. (2017), Villalobos et al. (2017), Elzamzamy, Penner, and Hazlehurst (2020), Huang HK (2020), D.-B. Jardin I. (2021). Por lo tanto, parece evidente que la regulación de los diferentes procesos carcinogénicos a través de la modulación de la señalización de Ca^{2+} intracelular podrían suponer una estrategia antitumoral interesante. A continuación, se van a resumir brevemente algunas de las alteraciones en estos sistemas de transporte y su implicación en el cáncer.

3.1. Remodelado de los VOCC

De las 10 isoformas de los VOCC, se han encontrado alteraciones en el cáncer en al menos 8 de ellas y, por lo general, suelen estar sobreexpresados en el fenotipo tumoral. Por ejemplo, en el cáncer de colon se ha encontrado sobreexpresión de las isoformas Cav1.1 (L. Zhang et al. (1997)), Cav1.2 (X.-T. Wang et al. (2000), C.-Y. Wang et al. (2015)), Cav1.3 (C.-Y. Wang et al. (2015)), Cav3.1 (C.-Y. Wang et al. (2015)). No obstante, estudios como Toyota et al. (1999) muestran una disminución de Cav3.1 en el tejido tumoral de colon que podría jugar un papel en el desarrollo tumoral, otros como Paz et al. (2003) observaron una infraexpresión del Cav3.3 tanto en líneas

celulares como en tejido humano de cáncer de colon. Además, Phan et al. (2017) también encontró que los Cav1.2, Cav2.1 y Cav3.1 está infraexpresado en el cáncer de colon.

Cabe destacar que, mediante análisis bioinformático de la base de datos ONCOMINE, C.-Y. Wang et al. (2015) o Phan et al. (2017) encontraron una clara expresión diferencial de estos canales analizando 21 y 19 subtipos de cáncer diferentes, respectivamente. No obstante, mientras que el primero se centró en la sobreexpresión de estos canales en el tejido tumoral, y encontró que los Cav1.2, Cav1.3 y Cav2.1 están sobreexpresados en la mayoría de ellos; el segundo se centró en la infraexpresión de estos, e identificó un patrón de infraexpresión de los VOCC en el cáncer de cerebro, riñón, mama y pulmón. Curiosamente, los autores de ambas publicaciones son los mismos, pero con dos años de diferencia y que, en cierto modo, publicaron en primer lugar los VOCC sobreexpresados y, dos años después, los infraexpresados. De estos análisis bioinformáticos se puede extraer, a grandes rasgos, las siguientes conclusiones (algunos de ellos contradictorios):

■ Cav1:

- Cav1.1: sobreexpresado leucemia o en cáncer cerebral; infraexpresado en cáncer de pulmón, renal y de cabeza y cuello.
- Cav1.2: sobreexpresado en leucemia o en cáncer de colon (adenocarcinoma, adenoma rectal y adenoma de colon), gástrico, páncreas, mama o próstata (carcinoma); infraexpresado en cáncer de colon (adenocarcinoma rectosigmoide) linfomas o próstata (carcinoma epitelial e intraepitelial. No obstante, aparecen ciertas contradicciones, pues se identifica tanto sobreexpresión como infraexpresión de este canal en glioblastomas, oligodendrogliomas.
- Cav1.3: sobreexpresado en cáncer de cerebro (glioblastoma), colon (adenocarcinoma, adenoma y adenocarcinoma rectosigmoide), pulmón, mama o próstata; infraexpresado en cáncer de cerebro (glioblastoma, astrocitoma difuso), pulmón, mieloma, renal o sarcoma.
- Cav1.4: sobreexpresado en cáncer de testículos; infraexpresado en linfoma.

■ Cav2:

- Cav2.1: sobreexpresado en leucemia o en cáncer de cerebro (glioblastoma, meduloblastoma...) o pulmón; infraexpresado en cáncer de cerebro (glioblastoma, astrocitoma...), colon (carcinoma y adenoma) o mama.
- Cav2.2: sobreexpresado en cáncer de próstata y mama; infraexpresado en cáncer de cerebro (glioblastoma, astrocitoma difuso...) o mama.
- Cav2.3: sobreexpresado en cáncer de sófago y útero; infraexpresado en leucemia o en cáncer cerebral (glioblastoma y astrocitoma), gástrico, de próstata o de pulmón.

■ Cav3:

- Cav3.1: sobreexpresado en cáncer colorrectal (adenocarcinoma rectosigmoide), de mama, de próstata o de pulmón; infraexpresado en cáncer cerebral (astrocitoma), de colon (carcinoma) o renal.
- Cav3.2: sobreexpresado en cáncer de renal, gástrico o en sarcoma; infraexpresado en cáncer cerebral (glioblastoma), pulmonar o de mama.
- Cav3.3: sobreexpresado en sarcomas o en cáncer de mama o esófago; infraexpresado en cáncer cerebral (glioblastoma y oligodendroglioma).

De lo anterior, está claro que las alteraciones en la expresión de las diferentes isoformas de los VOCC está presente en el cáncer, si bien es cierto que, en ocasiones, como se observa claramente en los glioblastomas, se pueden encontrar algunas de ellas tanto sobre como infraexpresadas. En cuanto al efecto de estas alteraciones, depende de la isoforma. Por ejemplo, la escisión del extremo C-terminal de los Cav1.2 y Cav1.3 puede contribuir a los *hallmarks* del cáncer a través de la modulación de la expresión de canales como TRPV4 o los canales de K^+ de corta conductancia, o de la entrada adicional de Ca^{2+} (Tajada and Villalobos (2020)). En cuanto al Cav3.2, se ha observado que pueden tener un papel clave en la proliferación celular en las células de cáncer de próstata a través de su relación con los canales de K^+ de larga conductancia (Gackiere et al. (2008)). Además, las subunidades accesorias de los canales de las familias Cav1 y Cav2 se han asociado a diferentes *hallmarks* de distintos cánceres: ovario, próstata, pulmón o colon (Tajada and Villalobos (2020)).

Cabe destacar que Cav1 es la que muestra una mejor señalización hacia el núcleo y es capaz de inducir la expresión de diferentes genes regulados por CREB, los cuales están estrechamente relacionados con la fisiopatología del cáncer (Sakamoto and Frank (2009)), lo que está relacionado con su colocalización con clústers de CaMKII (H. Ma et al. (2012)). De hecho, Cav2 no es capaz de activar esta señalización y no es tan eficaz para aumentar $[\text{Ca}^{2+}]_{\text{cit}}$ como Cav1, lo que podría explicar el mayor $[\text{Ca}^{2+}]_{\text{cit}}$ encontrado en las células de cáncer de colon HT29 (menor expresión de Cav2 y mayor de Cav1) respecto de las normales NCM460. No obstante, el Ca^{2+} introducido por Cav2, frente al Cav1, es preferiblemente captado por el RE y la mitocondria, lo que también podría explicar, en parte, el menor tamaño de los depósitos de las HT29 en comparación con las NCM460 (H. Ma et al. (2012)).

3.2. Remodelado de las PMCA

Las diferentes isoformas de PMCA muestran diferentes perfiles de expresión entre ellas así como entre diferentes tipos de cáncer. Se ha observado una reducción en la expresión de PMCA1 en el cáncer oral, tanto en líneas celulares como en muestras de tejido de pacientes (K. Saito et al. (2006)), aunque en líneas celulares de cáncer de mama está sobreexpresada (W. J. Lee et al. (2002)). En particular, se ha observado que el silenciamiento de esta bomba en líneas de cáncer de mama incrementa la necrosis inducida por ionomicina (Curry et al. (2012)). La expresión de PMCA2 tiene una asociación positiva con la proliferación celular en las líneas celulares de cáncer de mama (W. J. Lee, Roberts-Thomson, and Monteith (2005), Peters et al. (2016)); de hecho, en modelos de ratón KO para esta bomba el crecimiento del tumor se inhibe (Jeong et al. (2016)). PMCA4 ha mostrado una menor expresión en muestras de tejido de cáncer de colon. Esto podría resultar en el aumento de la tasa proliferativa, como se observó a través de la sobreexpresión de esta bomba en células HT29 (Aung et al. (2009)). Esto también podría tener efectos sobre el aumento de la angiogénesis, como muestra su inhibición con ATA en células endoteliales, pues esta bomba regula negativamente la angiogénesis vía VEGF al inhibir la ruta de señalización calcineurina/NFAT mediante la interacción con calcineurina (Kurusamy et al. (2017)). Por el contrario, Sritangos et al. (2020) encontraron una sobreexpresión de PMCA4 en el adenocarcinoma ductal pancreático y, además, vieron que en la línea celular MIA PaCa-2 juega un papel importante en la migración celular y en la resistencia a la apoptosis. Al igual que PMCA1, el silenciamiento de PMCA4 en líneas celulares de cáncer de mama incrementa la necrosis inducida por un inhibidor de Bcl-2 (Curry et al. (2012))

3.3. Remodelado de los NCX

En cuanto a los NCX, se ha visto que pueden estar sobreexpresados en algunos tipos de cáncer. Por ejemplo, se ha visto una sobreexpresión de NCX1 en tejidos de carcinoma de esófago de células escamosas. De hecho, J. Xu et al. (2016) observaron que la interacción entre Na/H^+ , NCX y CaM puede regular la proliferación mediada por IL6, la migración y la invasión en células de carcinoma hepatocelular. Además, se ha observado que en estas células se forma el complejo heteromérico TRPC6/NCX1 el cual promueve el efecto de $\text{TGF}\beta$ sobre la migración, invasión y metástasis (J. Xu et al. (2018)). De igual modo, en células de cáncer de páncreas, $\text{TGF}\beta$ induce la entrada de Ca^{2+} a través del complejo TRPC1/NCX1, de tal manera que este catión acaba por activar la $\text{PKC}\alpha$ dependiente de Ca^{2+} que promueve la invasión celular (Hui Dong et al. (2010)). Aunque para las otras dos isoformas hay menor evidencia, se ha observado una sobreexpresión de NCX3 en líneas celulares de carcinoma de ovario humano (Rodrigues, Estevez, and Santos Tersariol (2019), Chovancova et al. (2020)).

3.4. Remodelado de las SERCA

En diferentes tipos de cáncer se ha observado una reducción de SERCA2, como por ejemplo en muestras de pacientes con carcinoma de células escamosas (Endo et al. (2004)), en líneas celulares de tiroides (Pacífico et al. (2003)) o en muestras de tejido de pacientes de cáncer de pulmón. Esta infraexpresión y la consecuente disminución de los depósitos de Ca^{2+} podría inducir la resistencia a la apoptosis (Prevarskaya et al. (2014)). Además, la expresión del oncogen *KRAS* regula negativamente la expresión de SERCA2, como se comprobó al aumentar la expresión de SERCA2 tras suprimir este oncogen (Pierro et al. (2014)). Es más, la actividad de SERCA2 también está condicionada por otro gen sumamente implicado en el cáncer, el supresor tumoral *TP53*, pues la interacción entre ambas proteínas conlleva la mayor entrada de Ca^{2+} al RE y es capaz de regular la apoptosis dependiente de este catión (Giorgi et al. (2015)). Por el contrario, en muestras de tejido de pacientes de CCR se ha observado un aumento de esta isoforma (Chung et al. (2006)). De hecho, en líneas celulares de cáncer de colon se ha visto que la sobreexpresión de esta isoforma aumenta la tasa proliferativa y su capacidad para metastatizar (Fan et al. (2014)).

En cuanto a SERCA3, se ha visto una disminución tanto en líneas celulares como en muestras de tejido de pacientes de cáncer de colon (Gélébart et al. (2002), Brouland et al. (2005), Gou et al. (2014)). Esta disminución también se ha observado en muestras de tejido de cáncer de mama (Papp and Brouland (2011)) y en líneas celulares de adenocarcinoma de pulmón (Arbajian et al. (2013)). No obstante, durante la diferenciación celular en células de cáncer de pulmón aumenta la expresión de SERCA3, lo que podría explicar la disminución en el tamaño de los depósitos (Arbajian et al. (2013)), pues esta isoforma tiene menor afinidad por el Ca^{2+} respecto de SERCA2.

3.5. Remodelado de la liberación de Ca^{2+} desde el RE

La mayor liberación de Ca^{2+} desde los depósitos se ha relacionado con la inducción de la apoptosis dependiente de Ca^{2+} . Sobradillo et al. (2014) relacionaron la disminución de estos depósitos en células de CCR y el aumento en SOCE con el incremento en su proliferación, invasión y supervivencia. Es más, observaron que la SOCE generada por la liberación de Ca^{2+} de los depósitos intracelulares inducida por agonistas (ATP y carbacol) tenía lugar únicamente en las líneas celulares de CCR, posiblemente a la mayor facilidad para alcanzar el umbral de activación de SOCE favorecida por el menor contenido basal de los depósitos en las células tumorales.

La inhibición de la apoptosis a través de los IP3R puede llevarse a cabo por la acción de la proteína proapoptótica Bcl-2 expresada en el RE, pues es capaz de interactuar con IP3R e inhibir la liberación de Ca^{2+} al citosol y, en consecuencia, la apoptosis (Rong et al. (2009)). Otros genes relevantes en el cáncer, como el supresor tumoral *BRCA1*, también regulan la actividad de los IP3R favoreciendo la liberación de Ca^{2+} y, en consecuencia, induciendo la apoptosis (Hedgepeth, Garcia, and Wagner (2015)). Además, el oncogén *KRAS* regula la expresión de los IP3R (Pierro et al. (2014)).

La expresión de IP3R1 se ha visto disminuida en muestras de tejido de pacientes con glioblastoma (S. S. Kang et al. (2010)). De hecho, Tsunoda et al. (2005) observó que el silenciamiento de IP3R1 en células de cáncer de vejiga induce la resistencia a la apoptosis. En cuanto al IP3R3, a diferencia de la isoforma IP3R1, está sobreexpresado en las muestras de tejido de pacientes con glioblastoma (S. S. Kang et al. (2010)).

Además, Sakakura et al. (2003) observó el aumento de IP3R3 en muestras de tejido humano y líneas celulares de CCR, al igual que Shibao et al. (2010). También se ha observado que una mayor expresión de esta isoforma podría estar relacionada con una mayor agresividad. IP3R3 también podría participar en la capacidad metastásica de las células tumorales y en su mayor supervivencia, como comprobó S. S. Kang et al. (2010) al tratar células de glioblastoma con cafeína, pues la capacidad de invasión y supervivencia disminuyeron. No obstante, se ha visto que la delección del oncogén *KRAS* en células de cáncer colorrectal conllevó el aumento de expresión de IP3R3, la disminución de IP3R1 y el aumento en la liberación de Ca^{2+} desde el RE y de su sensibilidad a la apoptosis, por lo que la expresión de *KRAS* podría disminuir la liberación de Ca^{2+} desde los depósitos (Pierro et al. (2014)). Cabe destacar que durante la EMT, la expresión tanto de la isoforma IP3R1 como de la 3 aumenta en líneas celulares de cáncer de mama (Davis et al. (2013)).

En cuanto a IP3R2, hay poca evidencia de su participación en el cáncer. Heighway et al. (1996) observaron un aumento en la expresión de IP3R2 y coexpresión de esta isoforma con KRAS en muestras de pacientes con cáncer de pulmón de células no pequeñas. En cuanto a los receptores de RyR Kusner, Mygland, and Kaminski (1998) observó una disminución de la isoforma RyR1 en muestras de pacientes con tioma. En líneas celulares de cáncer de mama se observó un aumento de la isoforma RyR2 durante la EMT (Davis et al. (2013)).

3.6. Remodelado de la liberación de Ca^{2+} de la vía secretora

Grice et al. (2010) encontraron un aumento en la expresión de SPCA1 en el cáncer de mama del subtipo basal (Grice et al. (2010)). Estos autores observaron que la inhibición mediante siARN de esta bomba en este tipo de cáncer (línea celular MDA-MD-231) produjo notables cambios en la proliferación, morfología, regulación de enzimas dependientes de Ca^{2+} de la ruta secretora (como las convertasas) y el procesamiento del receptor del factor de crecimiento similar a la insulina, este último altamente correlacionado con un peor pronóstico del cáncer de mama (Dang and Rao (2016)). Es más, observaron cómo algunos transportadores de Ca^{2+} son capaces de regular el procesamiento de proteínas relevantes en la progresión tumoral sin repercutir en la señalización del Ca^{2+} citosólico (Grice et al. (2010)).

En cuanto a la isoforma SPCA2, M. Feng et al. (2010) encontraron que está sobreexpresada en el cáncer de mama, próstata y colon (M. Feng et al. (2010), Dang and Rao (2016)). De hecho, encontraron que su supresión disminuye la $[\text{Ca}^{2+}]_{\text{cit}}$ y la tumorigenicidad. Como resultado, la formación del complejo ORAI1/SPCA2 provoca una señalización constitutiva de Ca^{2+} independiente de los depósitos que promueve la tumorigénesis. Además, aunque este mecanismo

de transporte de Ca^{2+} a través de ORAI1 independiente de depósitos es importante en diferentes procesos fisiológicos, su desbalance debido a una sobreexpresión de SPCA2 conlleva el aumento de la proliferación y la tumorigénesis, mientras que la entrada de Ca^{2+} por SOCE parece jugar un papel más modesto en este sentido (M.-Y. Feng and Rao (2013)). De hecho, el aumento de $[\text{Ca}^{2+}]_{\text{cit}}$ se correlaciona con la activación constitutiva de la ruta RAS/ERK al favorecer la fosforilación de ERK1/2 y la expresión de CycD. No obstante, también se ha observado que SOCE mediada por STIM1 es indispensable para la invasión y metástasis donde, a la par, la expresión de SPCA2 se reduce, por lo que ambos procesos podrían jugar un papel clave en el proceso tumoral en diferentes estadios de la enfermedad (M.-Y. Feng and Rao (2013)).

3.7. Remodelado del transporte de Ca^{2+} a través de los TRP

Estos canales se han relacionado con diferentes patologías, entre ellas el cáncer. A continuación, se presentan algunas de las evidencias encontradas divididas en función de las diferentes familias.

3.7.1. TRPC

La sobreexpresión de TRPC1 podría explicar la mayor SOCE mostrada por las células de CCR respecto de las normales. También se ha observado su sobreexpresión en muestras de tejido de pacientes de cáncer de mama (Dhennin-Duthille et al. (2011), Monteith, Davis, and Roberts-Thomson (2012)). Es más, se ha visto que TRPC1 está implicado en los *hallmarks* de diferentes tipos de cáncer (proliferación y migración celular, o la angiogénesis) y se le asocia con un mal pronóstico. De hecho, este canal está implicado en el proceso de restitución epitelial, mecanismo regulado por la síntesis de poliaminas en el que, de forma transitoria, se activa la migración y/o la proliferación celular, y, puesto que la síntesis de poliaminas está sobreactivada en el CCR, el papel del TRPC1 podría ser relevante en *hallmarks* del cáncer tales como la autonomía proliferativa y la metástasis (Gutiérrez et al. (2019)). Además, se ha encontrado evidencia de un cambio hacia TRPC1 en cáncer de colon. También se ha observado que la inducción de TRPC1 aumenta la apoptosis en el epitelio intestinal a través de la inhibición de la activación de NF- κ B (Sobradillo et al. (2014), Rizopoulos and Assimakopoulou (2021), Lefranc (2021)). En cuanto a la pérdida de función del TRPC1, se ha observado que podría dificultar el crecimiento y proliferación tumoral en células de glioblastomas con el TRPC1 silenciado (Bomben et al. (2011)). En cuanto a su silenciamiento, en células de cáncer nasofaríngeo se observa una atenuación de su capacidad invasiva causada por una disminución de SOCE. (B. He et al. (2012)). Además, también tiene un papel importante en la migración de gliomas tras ser estimulados por EGF (Bomben et al. (2011)). Autores como Sobradillo et al. (2014) observaron en células de CCR mostraban una mayor SOCE y una disminución de los depósitos intracelulares de Ca^{2+} relacionado con cambios en TRPC1. Es más, también se ha observado que la interacción entre STIM1 y TRPC1 es clave para la migración celular.

Los niveles de expresión de TRPC3 elevados se han podido observar en gliomas y, además, que los pacientes con altos niveles de este canal tienen una esperanza de vida muy reducida (H.-H. Chang et al. (2018), Lefranc (2021)). Por el contrario, en pacientes con adenocarcinomas de pulmón la expresión de este canal es señal de un mejor pronóstico (H. Saito et al. (2011)). En las células del músculo esquelético se ha observado que TRPC3 juega un papel importante en la migración a través de una interacción con los canales de K^+ sensibles a Ca^{2+} de conductancia intermedia, $\text{K}_{\text{Ca}3.1}$ (Mesquita et al. (2021)). También se ha observado que el estrés oxidativo provoca una expresión constitutiva de TRPC3 (Venkatachalam and Montell (2007)). La sobreexpresión de este gen también se ha encontrado en muestras de tejido de pacientes con cáncer de mama, de ovario y de esófago (Aydar et al. (2009), SL Yang et al. (2009), Monteith, Davis, and Roberts-Thomson (2012)).

En cuanto a TRPC4, se ha observado que su sobreexpresión participa en la proliferación en células de cáncer de ovario (B. Zeng et al. (2013)), pues se ha observado que la activación de TRPC3 favorece la proliferación y la migración celular en cáncer de ovario (S. Li et al. (2020)). No obstante, su infraexpresión favorece la angiogénesis en células de carcinoma renal (Veliceasa et al. (2007)). De hecho, la sobreexpresión de TRPC5 está relacionada con el grado del tumor y el peor pronóstico de este. Además, también se ha observado que su sobreexpresión induce la EMT a través de la ruta de señalización HIF-1 α -Twist y promueve la metástasis en el cáncer de colon (Z. Chen et al. (2017a), Z. Chen et al. (2017b)). El TRPC5, al igual que otros muchos canales de Ca^{2+} , se han relacionado con la resistencia a diferentes terapias anticancerígenas. En este sentido, se ha observado quimiorresistencia dependiente de GLUT1 y a 5-Fluoruracil en pacientes de CCR que muestran sobreexpresión de este canal (T. Wang et al. (2017)). También se ha observado que TRPC5 induce la p-glicoproteína implicada en la resistencia a doxorubicin en líneas celulares de cáncer de mama, por lo que la inhibición de este (y otros canales de Ca^{2+} que participen en este tipo de resistencia a drogas antitumorales) puede volver a sensibilizar a estas células al fármaco (X. Ma et al. (2012)).

Por el contrario, el canal TRPC6 se ha visto infraexpresado en pacientes con CCR a nivel de transcrito (Sozucan et al. (2015)). No obstante, en pacientes con glioblastoma se ha encontrado una sobreexpresión de este canal y, además, mayor con el grado del tumor (Lefranc (2021)). También se ha encontrado que la sobreexpresión de este canal en miofibroblastos podría promover la formación de fibras en pacientes con enfermedad de Crohn (Rizopoulos and Assimakopoulou (2021)). Se ha observado que en condiciones de hipoxia la activación de Notch1 aumenta el cual, a su vez, induce la expresión de TRPC6 en líneas celulares de glioblastoma. No obstante, si se entorpece la actividad de TRPC6 entonces se observa el arresto del ciclo celular en la fase G2/M (Lefranc (2021)), lo que también se ha observado en células pancreáticas de ratón (Mesquita et al. (2021)), en pacientes con cáncer de esófago (Shi et al. (2009)), gástrico (Cai et al. (2009)) y glioblastoma (Ding et al. (2010)). En tejidos humanos de cáncer de mama también se ha encontrado que este canal está sobreexpresado (Guilbert et al. (2008), Dhennin-Duthille et al. (2011)). De hecho, en líneas celulares de cáncer de mama, en las que también se detectó la sobreexpresión de este canal, se ha visto que TRPC6 juega un papel importante en la proliferación, migración e invasión que podría estar relacionado con su papel regulador de la entrada de Ca^{2+} (I. Jardin et al. (2018)). En concreto, su papel podría ser más

importante como activador SOCE que como canal de Ca²⁺. Así, en las líneas celulares de cáncer de mama MCF7 donde SOCE está mediada por STIM1, STIM2 y ORAI3, o en las MDA-MB-231 en las que está mediada, principalmente, por STIM1 y ORAI1, TRPC6, el flujo de Ca²⁺ a través del TRPC6 activado por DAG podría ser necesario para la adecuada localización de ORAI3 u ORAI1 a la MP, respectivamente. En cuanto al TRPC7, no se conoce mucho su papel en el cáncer, si bien es cierto que se ha encontrado evidencia de que podría jugar un papel importante en el inicio del tumor, en la inestabilidad genómica y en el envejecimiento celular (Hsu et al. (2020)).

En la tabla 3.1 se muestra un resumen de algunos de los cambios en los TRPC encontrados en diferentes tipos de cáncer.

Tabla 3.1: Implicación de los TRPC en el cáncer.

| TRPC | Tipo de tumor | Modelo | Expresión diferencial | | Referencia |
|--------------|---------------|------------------|-----------------------|----------|--------------------------------|
| | | | ARNm | Proteína | |
| <i>TRPC1</i> | Mama | Tejido humano | ↑ | ↑ | Dhennin-Duthille et al. (2011) |
| | Colon | Líneas celulares | ↑ | ↑ | Sobradillo et al. (2014) |
| <i>TRPC3</i> | Mama | Tejido humano | ↑ | | Aydar et al. (2009) |
| | Glioma | Líneas celulares | ↑ | ↑ | Chang et al. (2018) |
| | Ovario | Tejido humano | | ↑ | SL Yang et al. (2009) |
| <i>TRPC4</i> | Riñón | Líneas celulares | ↓ | | Veliceasa et al. (2007) |
| | Ovario | Líneas celulares | ↑ | ↑ | Zeng et al. (2013) |
| | | Tejido humano | | ↑ | S. Li et al. (2020) |
| <i>TRPC5</i> | Colon | Tejido humano | ↑ | ↑ | Z. Chen et al. (2017a) |
| | | Líneas celulares | ↑ | ↑ | T. Wang et al. (2017) |
| <i>TRPC6</i> | Mama | Tejido humano | ↑ | | Aydar et al. (2009) |
| | | Tejido humano | ↑ | ↑ | Dhennin-Duthille et al. (2011) |
| | | Tejido humano | ↑ | ↑ | Guilbert et al. (2008) |
| | Esófago | Tejido humano | ↑ | ↑ | Shi et al. (2009) |
| | Gástrico | Tejido humano | ↑ | ↑ | Cai et al. (2009) |
| | Glioblastoma | Tejido humano | ↑ | ↑ | Ding et al. (2010) |
| | Colon | Tejido humano | ↓ | | Sozucan et al. (2015) |

3.7.2. TRPV

TRPV1 es un regulador importante de la expresión de EGFR, de tal manera que la sobreexpresión de TRPV1 disminuye los niveles de este (Mesquita et al. (2021)). En el cáncer de colon se ha encontrado que TRPV1 sirve como supresor tumoral, lo que está de acuerdo con la disminución en la expresión de este en este tipo de tumores (N. Hou et al. (2019), Rizopoulos and Assimakopoulou (2021)). De hecho, en el CCR se ha encontrado que la proteína de la matriz extracelular fibulin-5, implicada en la metástasis e invasión, promueve la producción de ROS y la apoptosis por la vía ROS/MAPK y por la señalización por Akt a través de la regulación negativa de TRPV1 (Y. Chen et al. (2019)). También se ha encontrado que el TRPV1 es capaz de inducir apoptosis en las células de CCR activando la ruta de señalización Calcineurina-NFAT-p53 (N. Hou et al. (2019)). De forma similar, también se ha encontrado que en pacientes con cáncer gástrico TRPV1 está infraexpresado y, puesto que en este tipo de cáncer TRPV1 también actúa como supresor tumoral a través de una vía $\text{Ca}^{2+}/\text{CaMKK}\beta/\text{AMPK}$, su regulación negativa está correlacionada con un peor pronóstico en la supervivencia de los pacientes con cáncer gástrico (N. Gao et al. (2020)). También se observa su disminución progresiva en gliomas a la par que se incrementa la malignidad de estos (Amantini et al. (2007)). De forma similar, se ha identificado la infraexpresión del TRPV1 en el cáncer urotelial donde también, al igual que en los glioblastomas, su menor expresión está relacionada con un peor pronóstico y menor supervivencia de los pacientes (Kalogris et al. (2010)). Cabe destacar que se ha considerado el uso de la capsazepine, un análogo sintético de la capsaicina que actúa como antagonista de TRPV1, como estrategia terapéutica para diferentes tumores y, entre ellos, el CCR. De hecho, se ha observado que este compuesto puede inducir la apoptosis en células CCR (Rizopoulos and Assimakopoulou (2021)) y en células de glioma (Amantini et al. (2007)).

En cuanto al TRPV2, en los gliomas se ha observado la disminución en la expresión de este canal a medida que progresa la enfermedad. De hecho, TRPV2 controla de forma negativa la supervivencia y proliferación celular de estas células a través de la apoptosis inducida por Erk (Lefranc (2021)). Por el contrario, se han encontrado altos niveles de este canal en hepatoblastoma humano o en cáncer de próstata. De hecho, se ha visto que estos altos niveles pueden favorecer la progresión del cáncer de próstata hacia un estadio más agresivo (Mesquita et al. (2021)). Curiosamente, aunque se ha encontrado que en cáncer de mama triple negativo está sobreexpresado el TRPV2, dentro de los tejidos tumorales de mama, la menor expresión está relacionada con un peor pronóstico; de hecho, la sobreexpresión de TRPV2 ha mostrado un aumento de la captación del quimioterápico doxorubicin y, en consecuencia, de la apoptosis de células de cáncer de mama triple negativas (Elbaz et al. (2018)). También se ha identificado la sobreexpresión de TRPV2 en cáncer de esófago, cuya mayor expresión se relaciona con una menor supervivencia de los pacientes (K. Zhou et al. (2014)). En cuanto al TRPV3, se ha observado su infraexpresión en el CCR (Sozucan et al. (2015)).

Respecto del TRPV4, en el CCR hay evidencia de que puede estar tanto sobre como infraexpresado. Por un lado, se ha observado su sobreexpresión y se ha relacionado con un pronóstico desfavorable de los pacientes. Su silenciamiento induce la autofagia y la apoptosis en células de CCR y detiene el desarrollo tumoral a través de la ruta PTEN (Xiaoyu Liu et al. (2019)). Además, también se ha encontrado que mutaciones en las regiones de unión a CaM le confieren un fenotipo invasivo (Arbaban et al. (2020)). Por otro lado, autores como Sozucan et al. (2015) han identificado la infraexpresión del TRPV4 en tejido de pacientes con cáncer de colon. También se ha identificado la sobreexpresión de este canal en tejidos de cáncer de mama (W. Lee et al. (2017)). De hecho, este canal parece tener un importante papel en la capacidad migratoria de este tipo de cáncer y, además, podría tener un papel relevante en tumores endoteliales humanos (W. H. Lee et al. (2016)). En este sentido, A. Fiorio Pla et al. (2012) observaron que su activación mediante forbol, un agonista de este canal, participa en el aumento de la capacidad migratoria en líneas celulares de cáncer de mama al favorecer la angiogénesis. También, W. H. Lee et al. (2016), en cuatro líneas celulares isogénicas derivadas de un xenoinjerto con potencial metastásico creciente, identificaron el papel del TRPV4 en su capacidad de extravasación al participar en la reducción de la rigidez celular a través del control del citoesqueleto en la corteza celular. Más tarde, estos mismos autores (W. Lee et al. (2017)) observaron que TRPV4 no solo induce, a través de un mecanismo dependiente de Ca²⁺, la fosforilación de la PKB/AKT y la regulación negativa de la expresión de proteína E-cadherina de la corteza celular, sino que también es necesario para la secreción de proteínas implicadas en la remodelación de la matriz extracelular. En otro tipo de tumores, como el de piel o el de vejiga, el TRPV4 está infraexpresado (Fusi et al. (2014), Mizuno et al. (2014)).

En el CCR, el TRPV5 muestra una menor expresión (Sozucan et al. (2015)), mientras que el TRPV6 muestra un comportamiento contrario, pues, incluso, se ha encontrado una sobreexpresión ya desde estadios tempranos de la enfermedad (Zhuang et al. (2002), Rizopoulos and Assimakopoulou (2021)). Es más, TRPV6 correlaciona con la patogeneidad de los tumores. En el caso de silenciar su gen, se observa una inhibición en la proliferación e inducción de la apoptosis en células de CCR. Curiosamente, la supresión de TRPV6 debida a una dieta rica en calcio ejerce un efecto protector frente al CCR. De igual forma que con TRPV1 en CCR, en las células de cáncer gástrico TRPV6 mediaría la apoptosis inducida por capsaicina lo que dependería de la abundancia de este canal (Rizopoulos and Assimakopoulou (2021)). Además, también se ha encontrado que mutaciones en las regiones de unión a CaM le confieren un fenotipo invasivo (Arbaban et al. (2020)).

En la tabla 3.2 se muestra un resumen de algunos de los cambios en los TRPV encontrados en diferentes tipos de cáncer.

Tabla 3.2: Implicación de los TRPV en el cáncer.

| TRPV | Tipo de tumor | Modelo | Expresión diferencial | | Referencia |
|--------------|---------------|---------------------------------------|-----------------------|----------|---|
| | | | ARNm | Proteína | |
| <i>TRPV1</i> | Urotelial | Tejido humano | ↓ | ↓ | Kalogris et al. (2010) Gao et al. (2020) |
| | Gástrico | Tejido humano y Líneas celulares | ↓ | ↓ | |
| | Colon | Tejido humano | | ↓ | N. Hou et al. (2019) |
| <i>TRPV2</i> | Esófago | Tejido humano y Líneas celulares | ↑ | | K. Zhou et al. (2014) |
| | Mama | Tejido humano | ↑ | ↑ | Elbaz et al. (2018) |
| | Vejiga | Tejido de ratón y Líneas celulares | ↑ | | Mizuno et al. (2014) |
| <i>TRPV3</i> | Colon | Tejido humano | ↓ | | Sozucan et al. (2015) |
| <i>TRPV4</i> | Mama | Tejido humano y Líneas celulares | ↑ | | W. Lee et al. (2017) |
| | Colon | Tejido humano y Líneas celulares | ↑ | ↑ | Xiaoyu Liu et al. (2019) |
| | | Tejido humano | ↓ | | Sozucan et al. (2015) |
| | Piel | Tejido humano | ↓ | ↓ | Fusi et al. (2014) |
| | Vejiga | Tejido de ratón y Líneas celulares | ↓ | ↓ | Mizuno et al. (2014) |
| <i>TRPV5</i> | Colon | Tejido humano | ↓ | | Sozucan et al. (2015) |
| <i>TRPV6</i> | Mama | Tejido humano | ↑ | ↑ | Dhennin-Duthille et al. (2011) |
| | Colon | Tejido humano | | ↑ | Zhuang et al. (2002) |
| | | Líneas celulares | ↓ | | Sobradillo et al. (2014) |
| | Próstata | Tejido humano | | ↑ | Zhuang et al. (2002) |
| | Tiroides | Tejido humano | | ↑ | Zhuang et al. (2002) |
| | Ovario | Tejido humano | | ↑ | Zhuang et al. (2002) |

3.7.3. TRPM

Los genes que codifican para estos canales son considerados como oncogenes, pues participan en la regulación del crecimiento de las células cancerígenas, la proliferación, la autofagia, la invasión y la EMT. Además, estos canales son una de las dianas terapéuticas de diferentes agentes quimioterápicos contra el CCR, como por ejemplo la ciclofosfamida, el 5-fluorouracilo y el leucoverin (Rizopoulos and Assimakopoulou (2021)).

El TRPM1 se ha descrito como un supresor tumoral, por lo que su infraexpresión se ha relacionado con un peor pronóstico del cáncer en pacientes con metástasis de melanoma y, más en concreto, el miR-211 codificado por el sexto intrón de este gen, pues reduce la expresión de genes relacionados con la capacidad metastásica como IGF2R, TFGBR2 o NFAT5 (Fioro Pla and Gkika (2013)). En el contexto del melanoma, este canal es el más relevante de los miembros de TRPM. En las metástasis se ha encontrado que está infraexpresado en las metástasis malignas no diferenciadas, mientras que la mayor expresión de este está correlacionado con la

diferenciación de los melanocitos. De hecho, puede servir como marcador de melanoma maligno. El TRPM1 ha mostrado ser altamente relevante en la homeostasis del Ca^{2+} intracelular en los melanocitos, pues controla positivamente la proliferación, diferenciación y la melanogénesis (Sulk and Steinhoff (2015)). Se ha identificado la sobreexpresión en diferentes tipos de cáncer, tales como el de próstata (X. Zeng et al. (2010)), de mama (Sumoza-Toledo, Espinoza-Gabriel, and Montiel-Condado (2016)) o en la leucemia mieloide aguda (S. Chen et al. (2020)).

El TRPM2 responde a situaciones de estrés celular, pues se activa por ROS, $\text{TNF}\alpha$ y ADP-ribosa, por lo que tiene un papel como sensor de estrés oxidativo y está implicado en la muerte celular por necrosis bajo estrés (Venkatachalam and Montell (2007), Hasan and Zhang (2018), Lefranc (2021), Rizopoulos and Assimakopoulou (2021)). De hecho, autores como S. Chen et al. (2020) han encontrado que este está altamente expresado en leucemia mieloide aguda donde promueve la proliferación celular y la supervivencia a través de la modulación de ROS, de la función mitocondrial y de la autofagia. En consecuencia, la entrada de Ca^{2+} y su señalización a través de la activación de TRPM2 es clave en diferentes *hallmarks* relacionados con la producción de citoquinas, la muerte celular o enfermedades inflamatorias. Además, se ha observado que regula la migración y la invasión en células de cáncer gástrico y la EMT (D. Yang and Kim (2020)). Se ha observado, en líneas de cáncer de colon Caco-2, que el efecto apoptótico de los agentes quimioterápicos ciclofosfamida, 5-fluorouracilo y leucoverin podría deberse a un efecto directo del TRPM2, los cuales tendrían un papel importante en la apoptosis al incrementar el Ca^{2+} intracelular y la despolarización de la mitocondria (Rizopoulos and Assimakopoulou (2021)). Respecto de TRPM3, se ha observado que un aumento en la tasa de mutación de este gen en regiones no codificantes podría estar relacionado con el riesgo de metástasis en CCR (Rizopoulos and Assimakopoulou (2021)).

La expresión de TRPM4 se ha asociado con la EMT y el patrón de crecimiento invasivo, pues regula la mecanotransducción en la migración celular ya que favorece el desarme de las adhesiones focales, la activación de FAK y Rac1 o la reorganización del citoesqueleto (D. Yang and Kim (2020)). En la regulación del TRPM4 también participan las poliaminas, pues su desensibilización es promovida por la depleción de PIP2 con poliaminas, de tal manera que la espermina es capaz de bloquear dicho canal. En particular, en CCR se ha encontrado el aumento de su expresión (Kappel, Stokłosa, et al. (2019)), aunque otros autores encontraron todo lo contrario (Sozucan et al. (2015)). También se ha encontrado sobreexpresión de este canal en cáncer de próstata, donde se asocia con el aumento de la capacidad metastásica (Holzmann et al. (2015)). El TRPM6, que participa en la absorción de Mg^{2+} a través de las células epiteliales del intestino, está infraexpresado en CCR (Ibrahim et al. (2019)). Es más, se ha descrito que unos niveles de expresión bajos de este canal en las células LoVo¹ resistentes a doxorubicina respecto de las sensibles a dicho tratamiento podría estar relacionado con las diferentes tasas de proliferación entre ellas y, por ende, con la quimiorresistencia a este tratamiento (Castiglioni et al. (2015), Rizopoulos and Assimakopoulou (2021)).

¹La línea celular LoVo proceden de una metástasis de adenocarcinoma de colon extraídas de un nodo linfático izquierdo subclavicular de un paciente varón de 56 años.

La evidencia encontrada para el TRPM7, permeable a $\text{Ca}^{2+}/\text{Mg}^{2+}/\text{Zn}^{2+}$, es algo contradictoria, pues mientras que unos estudios han encontrado sobreexpresión en CCR, otros revelan todo lo contrario. También se ha encontrado la sobreexpresión de este en cáncer de vejiga (Mizuno et al. (2014), S.-L. Gao et al. (2017)), de mama (Dhennin-Duthille et al. (2011)) o de próstata (L. Chen et al. (2017)). En el cáncer de vejiga se ha encontrado que este canal favorece la proliferación, migración, invasión y crecimiento tumoral (S.-L. Gao et al. (2017)). En el de próstata se ha visto que su inhibición suprime la migración e invasión a través de la regulación EMT (L. Chen et al. (2017)). En cuanto al CCR, existe evidencia que relaciona el consumo de Mg^{2+} con el riesgo de padecer esta enfermedad, sobretodo en pacientes con dietas bajas en Ca^{2+} , pues se ha observado que el consumo de este catión y polimorfismos en TRPM7 aumentan el riesgo de sufrir CCR. Además, controla la proliferación de las células HT29. También se ha observado que el aumento en la expresión de este canal en los tejidos de CCR está asociado con una mayor infiltración del tumor, metástasis en los nodos linfáticos y en regiones distantes, y un estadio de la enfermedad avanzado. En el contexto de la metástasis, el TRPM7 es el miembro de la superfamilia de los TRP mejor estudiado, pues regula procesos clave de la motilidad celular, tales como la polarización, la adhesión o la migración (D. Yang and Kim (2020)); por ejemplo, se ha observado que regula la expresión de diferentes MMP (Mesquita et al. (2021)). La infraexpresión de este canal *in vitro* ha es capaz de arrestar el ciclo celular, suprimir la migración e invasión, revertir la EMT e inducir la apoptosis (Fioro Pla and Gkika (2013), L. Chen et al. (2017)). En este sentido, el uso de inhibidores de TRPM7, como el carbacol, puede inhibir la proliferación celular, la migración y la invasión (Lefranc (2021)). Por el contrario, otros estudios revelan que la expresión de TRPM7 es menor en CCR respecto del tejido normal (Ibrahim et al. (2019)). Además, al igual que TRPM6, TRPM7 también presenta unos bajos niveles de expresión en células LoVo resistentes a doxorubicina lo que explicaría la superior tasa proliferativa respecto de las células sensibles a dicho tratamiento (Castiglioni et al. (2015), Locatelli et al. (2021)). Cabe destacar que las discrepancias entre los diferentes estudios podrían deberse a diferentes patrones de expresión de transportadores de Mg^{2+} (Fioro Pla and Gkika (2013)).

En cuanto al TRPM8, se ha observado una sobreexpresión en diferentes tipos de cáncer, como el de próstata, mama, colon, pulmón, piel, páncreas y glioblastomas (Tsavaler et al. (2001), Yee, Zhou, and Lee (2010), Klumpp et al. (2017)). Este canal podría estar relacionado con el inicio de la progresión de diferentes tumores, como el CCR, en el que se ha observado una sobreexpresión (Rizopoulos and Assimakopoulou (2021)). También se ha encontrado que este también podría tener un importante papel en la diseminación, supervivencia y radiorresistencia de glioblastomas humanos (Klumpp et al. (2017)). Se ha visto que su sobreexpresión en el cáncer de próstata podría estar regulada por diferentes miARN (Erdmann et al. (2014)) y por hormonas androgénicas (Bidaux et al. (2005)). No obstante, este tipo de cáncer puede progresar hacia una independencia de hormonas androgénicas, lo que le vuelve más agresivo (Prevarskaya et al. (2007)). En cuanto a su diferente localización celular, el TRPM8 del RE podría inducir la resistencia a la muerte (Monteith et al. (2007), Bidaux et al. (2007)). El uso de cannabigeron, derivado canabinoide no psicotrópico, podría bloquear TRPM8 (aunque es agonista de TRPA1, TRPV1 y TRPV2) e inhibir el crecimiento de las células de CCR (Borrelli et al. (2014)).

En la tabla 3.3 se resumen algunas de las alteraciones de los TRPM en diferentes tipos de cáncer.

Tabla 3.3: Implicación de los TRPM en el cáncer.

| TRPV | Tipo de tumor | Modelo | Expresión diferencial | | Referencia |
|--------------|-------------------------|------------------------------------|-----------------------|---------------------------|---------------------------------|
| | | | ARNm | Proteína | |
| <i>TRPM2</i> | Leucemia Mieloide aguda | Líneas celulares | ↑ | | Chen et al. (2020) |
| | Próstata | Tejido humano y Líneas celulares | | ↑ | Zeng et al. (2010) |
| <i>TRPM4</i> | Mama | Tejido humano | ↑ | | Sumoza-Toledo et. al (2016) |
| | Colon | Tejido humano | ↓ | | Sozucan et al. (2015) |
| | | Tejido humano y Líneas celulares | ↑ | ↑ | Kappel, Stokłosa, et al. (2019) |
| <i>TRPM7</i> | Próstata | Tejido humano | | ↑ | Holzmann et al. (2015) |
| | Mama | Tejido humano | ↑ | ↑ | Dhennin-Duthille et al. (2011) |
| | Vejiga | Tejido humano | ↑ | | Gao et al. (2017) |
| | | Tejido de ratón y Líneas celulares | ↑ | | Mizuno et al. (2014) |
| <i>TRPM8</i> | Gástrico | Líneas celulares | ↑ | ↑ | Kim and Hong (2016) |
| | Próstata | Líneas celulares | ↑ | | Chen et al. (2017) |
| | Mama | Tejido humano | ↑ | ↑ | Dhennin-Duthille et al. (2011) |
| | | Tejido humano | ↑ | ↑ | Tsavalier et al. (2001) |
| | Colon | Tejido humano | ↑ | | Tsavalier et al. (2001) |
| | | Líneas celulares | ↓ | | Sobradillo et al. (2014) |
| | Pulmón páncreas | Tejido humano | ↑ | ↑ | Tsavalier et al. (2001) |
| | Tejido humano | ↑ | ↑ | Yee, Zhou, and Lee (2010) | |

3.7.4. TRPP

La sobreexpresión de TRPP1 y TRPP2 se ha relacionado con un fenotipo más agresivo de CCR y un peor pronóstico. También se ha visto, en las líneas celulares de cáncer de colon HT29, que TRPP1 regula rutas de señalización como mTOR y JAK, activadas en el cáncer. Además, TRPP1 y TRPP2 son considerados nuevos marcadores y dianas terapéuticas del CCR (Rizopoulos and Assimakopoulou (2021)).

3.7.5. TRPML

Se ha encontrado una correlación entre la disminución de TRPML1 (ARNm) con la menor supervivencia en pacientes con glioblastoma o cáncer páncreas; esto podría deberse a que TRPML1 es un sensor de estrés oxidativo capaz de activar de forma irreversible la autofagia lo cual que conlleva la muerte celular. En cuanto a TRPML2, se ha observado su aumento en líneas celulares de glioblastomas de alto grado en la expresión tanto en astrocitos como en células madre progenitoras neuronales. Además, la pérdida de este canal inhibe la proliferación e induce la apoptosis (Mesquita et al. (2021), Lefranc (2021)).

3.8. Remodelado de SOCE

En este apartado se va a hablar del remodelado de SOCE, si bien es cierto que, puesto que ya se ha hablado de ellos en el apartado anterior, los miembros de la superfamilia TRP no se tratarán en detalle en este apartado. Además, también se va a mencionar su posible implicación en los *hallmarks* del cáncer. Diferentes estudios han mostrado el remodelado de SOCE en diferentes tipos de cáncer. Por ejemplo, Sobradillo et al. (2014) encontraron que SOCE provocada por la liberación de Ca^{2+} de los depósitos intracelulares inducida por agonistas (ATP y carbacol) sólo tenía lugar en las líneas celulares de cáncer de colon, pero no en las normales, lo que podría deberse al mayor tamaño de los depósitos en las tumorales. Esto podría ser explicado por el menor contenido basal de Ca^{2+} en los depósitos de las células tumorales, pues en tal situación se requiere de un menor vaciado de estos depósitos para alcanzar el umbral a partir del cual se activa SOCE. A nivel molecular, estos autores observaron que esta mayor SOCE en las células tumorales podría deberse a la sobreexpresión, a nivel de proteína, de ORAI1-3, STIM1, TRPC1, y a la disminución de STIM2, lo que favorece la traslocación de STIM1 a la MP y su interacción con TRPC1 (figura 3.1).

Sobradillo et al. (2014) también encontraron diferencias en las corrientes de los canales SOCC entre líneas celulares de cáncer de colon (HT29) y normales (NCM460) con la técnica *patch-clamp*. En concreto, el vaciado de los depósitos de Ca^{2+} en estas células normales de colon provocado por la acción de la tapsigargina es capaz de inducir una corriente similar a la I_{CRAC} observada en las células T. Es decir, una corriente de entrada rectificadora pequeña independiente del voltaje altamente selectiva para Ca^{2+} que, además, es sensible a La^{3+} y a bajas concentraciones de 2-ABP. Estos y otros datos indican que en las células normales los componentes de SOCE son los canónicos, STIM1 y ORAI1, aunque STIM2 también parece jugar un papel importante. Por el contrario, en las células de cáncer de colon este mismo vaciado mostró un perfil de corriente intensidad vs. voltaje diferente al mostrado por las células normales, pues en las células de CCR se pueden apreciar dos tipos de corriente diferentes. Por un lado, induce una corriente similar a la I_{CRAC} observada en las células normales de colon, salvo que, en estas células, se observa una densidad de corriente mayor. Por otro lado, y a diferencia de las células normales, también se registra una mezcla de dos corrientes, I_{CRAC} con I_{SOC} , esta última caracterizada por tener componente de salida, no ser selectiva para Ca^{2+} y ser sensible a grandes [2 – ABP]. Por lo tanto, además de los componentes canónicos que participan en las células normales de colon, en las HT29 también podrían participar miembros de la superfamilia TRP tales como TRPC1 (figura 3.1).

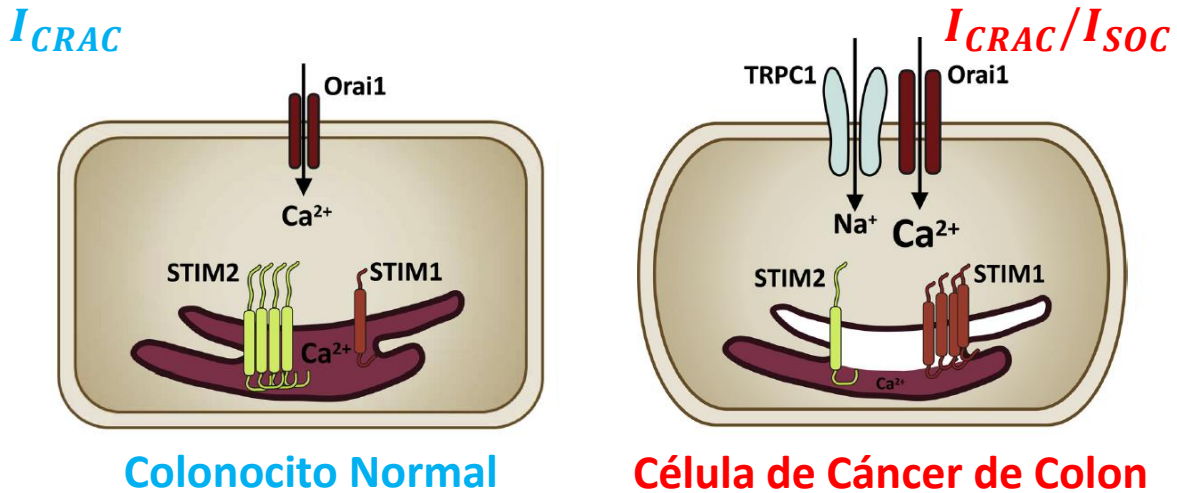


Figura 3.1: Remodelado de SOCE y de los depósitos en el cáncer de colon. En las líneas celulares de cáncer de colon se ha encontrado un menor tamaño de los depósitos y una mayor SOCE respecto de las normales. El menor tamaño de los depósitos en las células tumorales respecto de las normales, posiblemente debido a la menor expresión de STIM2, permite alcanzar más fácilmente el umbral de vaciado del RE necesario para activar SOCE. La participación de dos corrientes diferentes ($I_{\text{CRAC}}/I_{\text{SOC}}$) en las células tumorales frente a I_{CRAC} presente en las normales puede deberse a la mayor expresión de ORAI1, STIM1 y TRPC1 (modificado de Villalobos et al. (2019)).

En esta misma línea, Gutiérrez et al. (2019) observó que estas mismas células, HT29, las cuales sobreexpresan ODC, revierten parcialmente su fenotipo tumoral tras el tratamiento con DFMO. En concreto, aumentan el tamaño de sus depósitos y disminuye SOCE, lo que puede explicarse por la disminución en la expresión de TRPC1 a nivel tanto de ARNm como de proteína, o de STIM1 a nivel de proteína (figuras 3.2). Además, Hernández-Morales et al. (2017) observó que el tratamiento con sulindac sobre estas células es capaz de inhibir SOCE.

En relación a la corriente I_{SOC} en HT29, Gutiérrez et al. (2019) encontró que el tratamiento con DFMO es capaz de inhibir la componente de salida de I_{SOC} de una forma dependiente tanto del tiempo como de la dosis, mientras que la componente de entrada puede ser inhibida por el AINE sulindac. Por un lado, en relación al tratamiento con DFMO, observaron que tras 6 horas del tratamiento con 5 mM DFMO la componente de salida de I_{SOC} desapareció sin afectar a la componente de entrada, pues se consiguió un perfil intensidad vs. voltaje similar al I_{CRAC} de las NCM460. No obstante, tras 24 horas esta inhibición de la componente de salida sólo es parcial y, tras 96 horas dicha inhibición se pierde por completo. No obstante, a una concentración menor de DFMO, 500 μM , a pesar de que también se pierde la componente de salida, el perfil intensidad vs. voltaje es el de una corriente I_{SOC} en lugar de I_{CRAC} , puesto que el potencial de reversión se alcanza a 0 mV. Cabe destacar que este efecto inhibitorio de DFMO sobre la componente de salida de I_{SOC} tiene lugar a través de la inhibición de ODC y, por ende, de la inhibición de la síntesis de poliaminas, pues el tratamiento combinado de DFMO con putrescina (>500 μM) sobre las HT29 revierte dicha inhibición (figura 3.2).

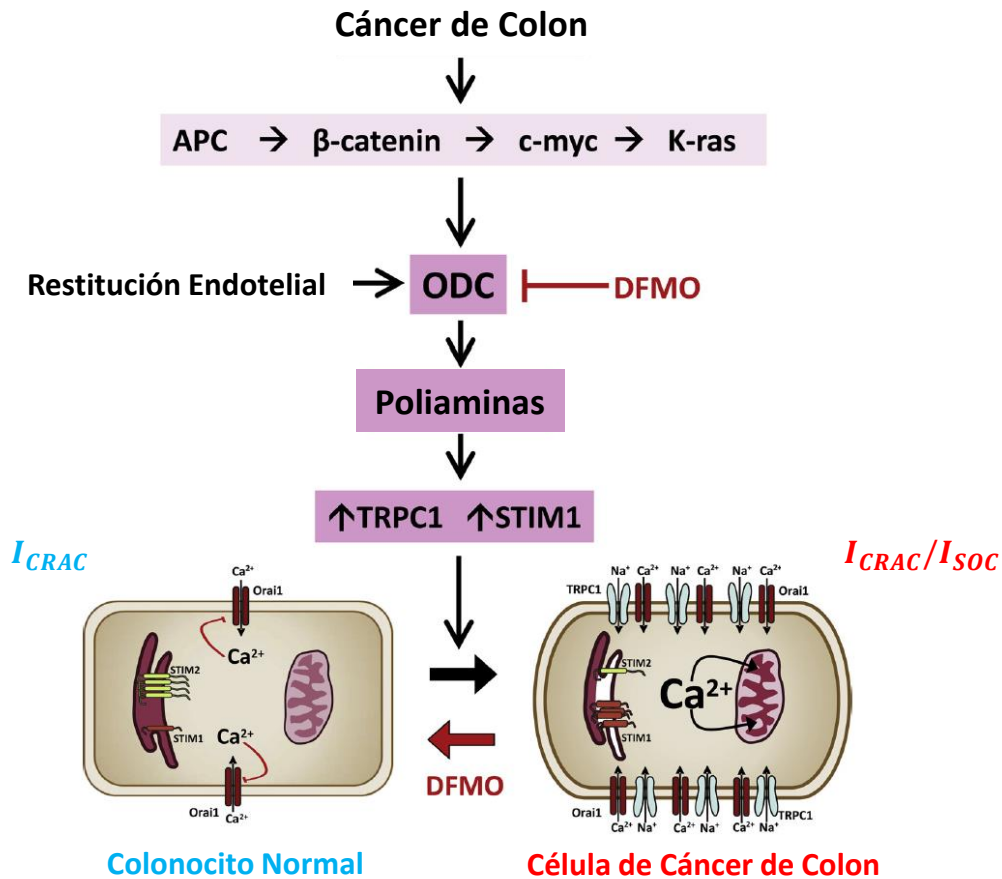


Figura 3.2: Reversión parcial del remodelado de SOCE en el cáncer de colon por DFMO. El tratamiento con DFMO es un inhibidor de ODC, enzima clave en la síntesis de poliaminas, proceso sobreactivado en el cáncer de colon consecuencia de alteraciones en la ruta Wnt/ β -catenina, como pueden ser mutaciones en APC. La síntesis de poliaminas promueve la sobreexpresión de STIM1 y TRPC1 en las células tumorales y, en consecuencia, su característica corriente $I_{\text{CRAC}}/I_{\text{SOC}}$, de tal manera que la inhibición de ODC con DFMO revierte parcialmente el fenotipo tumoral, pues reduce la expresión de TRPC1, de STIM1, aumenta el tamaño de los depósitos y da lugar a una corriente similar a CRAC (modificado de Villalobos et al. (2019)).

Por otro lado, en relación a la componente de entrada de I_{SOC} selectiva de Ca^{2+} en HT29, Hernández-Morales et al. (2017) observaron que el sulindac es capaz de inhibirla a través de la despolarización de la mitocondria, puesto que de esta manera pierde su capacidad para actuar como tampón de Ca^{2+} y, por ende, se promueve la inactivación de I_{SOC} (ver apartado “3.9 Remodelado del transporte de Ca^{2+} mitocondrial”). Posteriormente, Gutiérrez et al. (2019) observó que el tratamiento combinado de sulindac con DFMO en estas células consiguió la inhibición tanto de la componente de entrada como de salida de I_{SOC} . No obstante, resultados similares fueron obtenidos con anterioridad por Núñez et al. (2006), en los que se observó que el salicilato en concentraciones terapéuticas, que también provoca la despolarización parcial de la mitocondria, inhibe la captación de Ca^{2+} por parte de la mitocondria y, en consecuencia, acaba por inhibir SOCE y, como resultado, disminuye la proliferación celular de las células de cáncer de colon HT29.

No obstante, la contribución de los SOCC dependen del tipo de cáncer, así como de las isoformas tanto de este canal como de los sensores STIM (Chalmers and Monteith (2018)). Es más, como se verá más adelante, en un mismo tipo de tumor puede suceder que en función del estadio de la enfermedad unas isoformas estén sobre o infraexpresadas y, en cada caso, estos perfiles de expresión tan diferentes conferirán a las células tumorales diferentes capacidades. Las mutaciones en ORAI1 de ganancia de función se han encontrado en diferentes tipos de cáncer, como el CCR, el de estómago o de útero (Chalmers and Monteith (2018)). Se ha encontrado que estas podrían remodelar SOCE. De hecho, se han empleado inhibidores de ORAI1 para tratar el linfoma No Hodgkin. En muestras de pacientes de cáncer de mama basal se ha observado un aumento en la expresión de ORAI1, y lo diferencia del cáncer de mama no basal. También se ha encontrado sobreexpresión de este en cáncer de pulmón, donde se asocia a una mayor proliferación celular y un peor pronóstico de los pacientes (Zhan et al. (2015)). Cabe destacar que también se ha encontrado en células de cáncer de mama un aumento de la participación de esta isoforma en un mecanismo de entrada de Ca^{2+} independiente de los depósitos y que depende de la SPCA2 (M. Feng et al. (2010)). También se han observado diferentes mutaciones en TM2 de ORAI1 de ganancia de función que dan lugar a un ORAI constitutivamente activo e independiente de STIM1. En concreto, la mutación en el residuo A137 está relacionada con el cáncer de colon (Network and others (2012), Frischauf et al. (2017)).

En cuanto a ORAI3, se ha observado una sobreexpresión de este en líneas celulares de cáncer de mama positivos a receptor de estrógenos, así como en muestras de pacientes (Chalmers and Monteith (2018)). También se ha encontrado sobreexpresión de esta isoforma en cáncer de pulmón, donde podría favorecer la metástasis y la menor supervivencia de los pacientes (Benzerdjeb et al. (2016)). STIM1 aparece sobreexpresado en el cáncer de mama basal, a la par de una infraexpresión de STIM2 (Chalmers and Monteith (2018)). En muestras de pacientes de CCR también se observa la sobreexpresión de STIM1 (J.-Y. Wang et al. (2015)) y de STIM2 (Aytes et al. (2012)). También se observa una expresión de esta isoforma en el cáncer de cervix (Y.-F. Chen et al. (2011)), de pulmón (W. Li et al. (2013)), hepáticos (N. Yang et al. (2013)), en melanomas (Umemura et al. (2014)) o en glioblastomas (Scrideli et al. (2008)). En cuanto a su expresión CCR, algunos autores han encontrado sobreexpresión de STIM2 (Aytes et al. (2012)), que podría explicar la supresión del crecimiento tumoral, mientras que otros infraexpresión (Sobradillo et al. (2014)) en los cuales se observa una mayor expresión de estas células.

En general, SOCE participa en diferentes *hallmarks* del cáncer: invasión, proliferación, migración y resistencia a la apoptosis y a los estímulos citotóxicos. En la figura 3.3 se muestra un resumen del papel de las isoformas de ORAI y STIM en los *hallmarks* del cáncer (figura 3.3):

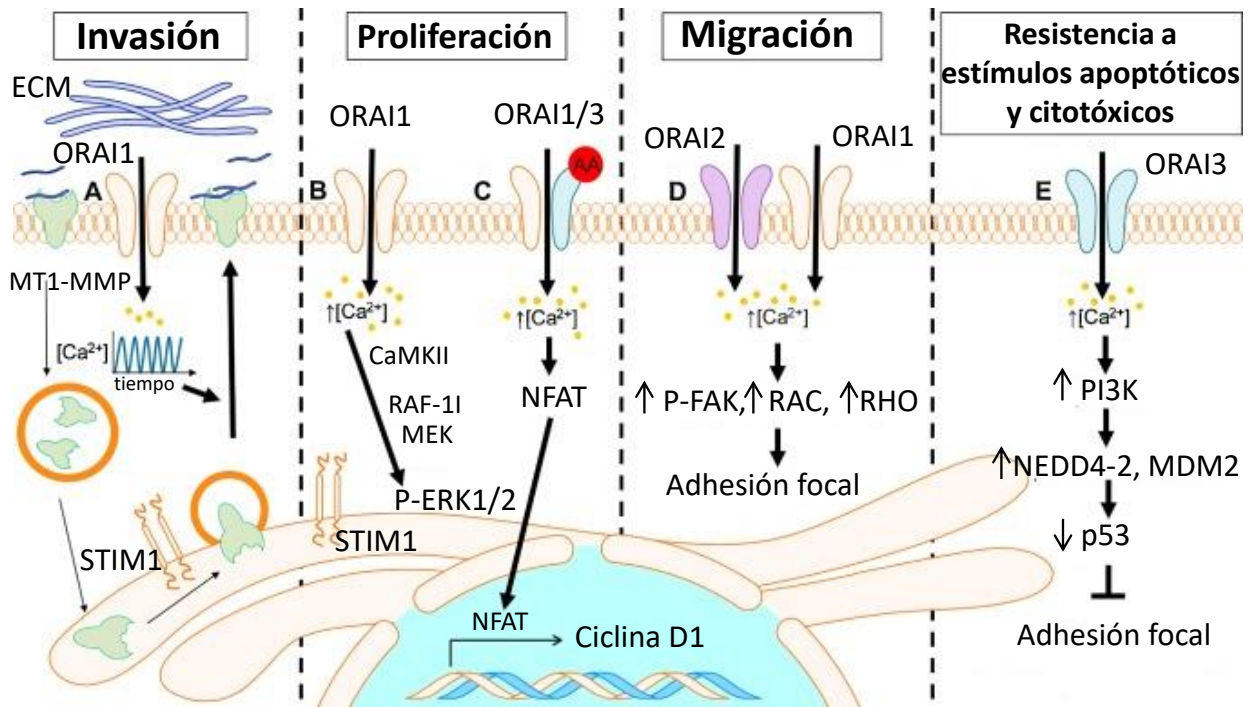


Figura 3.3: Implicación de ORAI y STIM en los hallmarks de cáncer. En la figura se muestra un esquema de la participación de la SOCE en algunos de los hallmarks del cáncer: invasión, proliferación, migración, resistencia a la muerte y a la citotoxicidad (Modificado de Chalmers and Monteith (2018)).

- **Invasión y metástasis:** Uno de los hallmarks donde la remodelación de SOCE está implicada es en la capacidad de las células cancerígenas de invadir, migrar y metastatizar de las células cancerígenas. Autores como Sobradillo et al. (2014) demostraron la relación de SOCE con la capacidad invasiva, pues a través de la inhibición farmacológica de SOCE con el antagonista 2APB observaron la disminución de la invasión de células de carcinoma de colon. Estudios como el realizado por Shengyu Yang, Zhang, and Huang (2009) mostraron que en líneas celulares de cáncer de mama el silenciamiento de ORAI reduce la capacidad metastásica de estas. El silenciamiento de ORAI1 también es capaz de inhibir la degradación de la matriz extracelular a través de la inhibición del reciclado de la MMP MT1 (J. Sun et al. (2014)). También se ha observado que el silenciamiento de ORAI1 en líneas celulares de adenocarcinoma de pulmón reduce los marcadores de inducción de EMT (Qi et al. (2016)). Para ORAI2 y 3 los estudios son más escasos. Por un lado, R Diez-Bello (2012) vieron que al silenciar ORAI2 en una línea celular de leucemia aguda la migración celular se atenuó, lo que podría deberse al efecto sobre la fosforilación de la *quinasa de las adhesiones focales* (FAK, del inglés, *focal adhesion kinase*). En cuanto a ORAI3, también se ha visto que su silenciamiento en líneas celulares de cáncer de mama reduce su capacidad invasiva (Motiani et al. (2013)).

En cuanto a la inhibición de STIM1, se ha visto que mejora la adhesión focal y disminuye la renovación de la adhesión focal a través de la inhibición de FAK en células de hepatocarcinoma, por lo que la sobreexpresión de STIM1 explica la capacidad migratoria de estas células cancerígenas (N. Yang et al. (2013)). De igual modo, el silenciamiento de STIM1 disminuye la capacidad migratoria de células de melanoma (Umemura et al. (2014)). También se ha observado que el aumento de su expresión en células de cáncer de mama, consecuencia del silenciamiento de *OCT4*, induce la EMT (Hu et al. (2011)). En las células de cáncer de cérvix se ha observado la participación de esta isoforma en la activación de la proteasa calpanina, lo que favorece la metástasis (Y.-F. Chen et al. (2011)). También en las células de CCR se observó que el aumento de expresión de STIM1 está relacionado con el aumento de su capacidad metastásica y con su tamaño, pues el silenciamiento de esta conlleva la disminución de la SOCE y de su capacidad migratoria (J.-Y. Wang et al. (2015)).

- *Autonomía proliferativa*: En cuanto a la implicación de SOCE en la autonomía proliferativa de las células tumorales, estudios como los realizados por Núñez et al. (2006), en los que se inhibió SOCE con 2-ABP, se observó que dicha inhibición fue suficiente para detener la proliferación celular en la línea celular de cáncer de colon HT29 y en otras como Jurkat. También, en estudios de silenciamiento de ORAI1 y ORAI3 en líneas celulares de cáncer de mama se observó la disminución de la proliferación, lo que puede explicarse por una atenuación en la actividad de ERK1/2, la quinasa de adhesión focal, la actividad transcripcional del factor nuclear para células T activadas (NFAT) y cMyc (M. Feng et al. (2010), Motiani et al. (2013), Faouzi et al. (2013)). Cabe destacar que también se ha encontrado un mecanismo de activación de la entrada de Ca^{2+} a través de ORAI1 independiente de los depósitos pero que depende de la SPCA2 y que participa en la proliferación celular (M. Feng et al. (2010)). Estudios de silenciamiento de ORAI2 en líneas celulares de leucemia mieloide aguda también muestran una reducción en la tasa proliferativa (29). En cuanto al cáncer de colon, se ha observado que el silenciamiento de ORAI1 disminuye la capacidad del EGF para inducir un aumento de COX2 (J.-Y. Wang et al. (2012)). En cuanto a ORAI3, mediante la infraexpresión de esta isoforma en células de cáncer de mama, se ha visto que se arresta el ciclo celular y se induce la apoptosis, por lo que la sobreexpresión de ORAI3 participa en la progresión de la fase G_1 a la G_1/S en un mecanismo que depende de SOCE mediada por ORAI3 a través de la regulación de CycD1, CycE, CDK4 y CDK2 así como de p21 y p53 mediante la regulación de cMyc (Faouzi et al. (2011)). También se ha observado la participación de ORAI3 en células de cáncer de próstata, en la que la sobreexpresión de esta isoforma favorece la formación del heterodímero ORAI1/ORAI3 que permite la entrada de Ca^{2+} independiente de los depósitos pero regulada por ácido araquidónico y que participa en la proliferación y vuelve menos susceptibles a la apoptosis a las células (Dubois et al. (2014)).

También se ha visto que STIM1 participa en la proliferación celular, pues la disminución de su expresión en células de cáncer cervical arresta el ciclo celular en la fase G_1/G_0 (Y.-F. Chen et al. (2011)). En cuanto a STIM2, se ha encontrado que tiene una regulación negativa del crecimiento y de la capacidad migratoria en células de melanoma, pues su silenciamiento reduce la $[\text{Ca}^{2+}]_{\text{cit}}$ y conlleva el aumento de la tasa proliferativa y la disminución de la capacidad migratoria e invasiva (Stanisz et al. (2014)). En este mismo sentido, la sobreexpresión de esta isoforma observada en los tumores colorrectales puede suprimir la proliferación celular y el crecimiento de estos, pues el efecto contrario se observa con su silenciamiento (Aytes et al. (2012)). Por el contrario, autores como Sobradillo et al. (2014) han encontrado que esta isoforma está infraexpresada en CCR y que dicha infraexpresión podría explicar, en parte, la mayor proliferación.

- *Resistencia a la muerte celular:* SOCE también está implicado en la muerte celular. Por ejemplo, en cáncer de próstata que progresa hacia un fenotipo más invasivo y resistente a la terapia androgénica se observa la infraexpresión de ORAI1 y la disminución de SOCE, lo que podría favorecer su resistencia a la apoptosis ante agentes como el $\text{TNF}\alpha$ o el cisplatino (Flourakis et al. (2010), Verbe (2016)). Además, la inhibición farmacológica de SOCE o a través del silenciamiento de ORAI1 o STIM también muestra la disminución de la inducción de la muerte celular de diferentes tumores, como sucede tras la inhibición de STIM2 en líneas celulares de cáncer de colon (Sobradillo et al. (2014)). No obstante, en otros tipos de tumores el silenciamiento de ORAI1 o la inhibición de SOCE puede conllevar la muerte celular, como muestran los experimentos de Vacher et al. (2015) al silenciar ORAI1 en células B de linfoma No Hodgkin, o los de H. Liu et al. (2011) en glioblastomas de rata tratados con Rituximab en los que se reduce la SOCE. Vacher et al. (2015) también observó este mismo efecto al silenciar ORAI3.

En cuanto a los sensores, H. Liu et al. (2011) mostró que el silenciamiento de STIM1 en células de glioblastoma aumentó la apoptosis, por lo que su mayor expresión podría proteger a estas células de la muerte celular. Por el contrario, S. Sun et al. (2012) mostró que el silenciamiento de STIM1 y la consecuente reducción de la SOCE conlleva a la reducción de la apoptosis en células de cáncer de colon. Gu et al. (2014) mostró un resultado similar mediante el silenciamiento de STIM1 en células de cáncer de próstata disminuye la apoptosis a través de la disminución de Bcl-2/BAX que acaba por activar la caspasa-3.

- *Resistencia a terapias antitumorales:* SOCE también parece jugar un papel importante. Por ejemplo, se ha encontrado que ORAI3 podría jugar un papel clave en la resistencia a fármacos, como se evidencia al sobreexpresar este canal en líneas celulares de cáncer de mama que conlleva la resistencia a cisplatino, al 5-fluorouracilo y al paclitaxel. Este mecanismo de resistencia a la quimioterapia es dependiente de p53, pues la sobreexpresión de ORAI3 conlleva una menor inducción de p53 por estos fármacos (Hasna et al. (2018)). También se ha observado el papel de STIM1 en la resistencia a la muerte a cisplatino en células de cáncer de pulmón (W. Li et al. (2013)).

Cabe destacar que el ácido araquidónico, que induce la entrada de Ca^{2+} desde el medio extracelular a través de los canales ARC formados por ORAI1 y ORAI3, y que requieren del STIM1 de la MP, puede inducir el aumento de $[\text{Ca}^{2+}]_{\text{cit}}$ a través del aumento de la entrada de Ca^{2+} y, a su vez, inhibir SOCE inducida por ATP o taspigargina (Alessandra Fiorio Pla et al. (2008)). No obstante, esta entrada a través de ARC está relacionada con la progresión de las fases tempranas de la angiogénesis a través de la proliferación y la tubulogénesis, aunque disminuye durante la reorganización y las células endoteliales derivadas del tumor en las estructuras similares a capilares. Por el contrario, su inhibición puede contribuir al efecto antiangiogénico del carboxiamidotriazol.

A parte de las proteínas más clásicas protagonistas de SOCE, como son los ORAI, algunos TRP o STIM, hay otras proteínas que juegan un papel clave en este mecanismo de transporte, como CRACR2A y MS4A12 que, al igual que las anteriores, alteraciones en estas también están implicadas en el cáncer:

- *CRACR2A*: Srikanth et al. (2010) propusieron que en condiciones basales CRACR2A se encuentra en el citoplasma y tanto ORAI1 como STIM1 están en las membranas plasmática y del RE, respectivamente, sin formar complejos ni oligomerizar con ellos mismos. No obstante, tras el vaciado de los depósitos, STIM1 y ORAI1 se reagrupan hacia zonas de contacto entre las membranas del ER y la MP, proceso en el que CRACR2A podría participar ya sea favoreciendo la oligomerización de STIM1, la de ORAI1, la de ambos y, simplemente, estabilizando el complejo ORAI1-STIM1 una vez formado. De esta forma, se genera un complejo ternario que permanece unido hasta que aumenta $[\text{Ca}^{2+}]_{\text{cit}}$, momento en el que se disocia el CRACR2A de dicho complejo. Además, también se ha observado que la unión de CaM al extremo N-terminal de ORAI1 da lugar a la inactivación de los canales CRAC y, curiosamente, estos sitios de unión a CaM tanto de STIM1 como de ORAI1 se solapan con las regiones de unión a CRACR2A. Es más, debido a que CRACR2A y CaM tienen una afinidad similar por Ca^{2+} , a bajas $[\text{Ca}^{2+}]_{\text{cit}}$ CRACR2A estaría activando ORAI1 mientras que al aumentar más esta concentración CaM ejercería un efecto inactivador dependiente de Ca^{2+} de este canal. Por lo tanto, el papel del CRACR2A como posible modulador del complejo STIM1-ORAI1 en la activación de SOCE sería dependiente de Ca^{2+} . No obstante, Srikanth et al. (2010) sugieren que CRACR2A podría tener un papel importante para unos niveles fisiológicos de expresión de ORAI1 y STIM1, pero no cuando estos están sobreexpresados, como por ejemplo sucede en diferentes tipos de cáncer.

En conclusión, el papel de CRACR2A a través de su unión a ORAI1 y STIM1 es polifacético. Por un lado, podría estabilizar la interacción entre ambos en áreas de contacto entre la MP y el RE, por lo que juega un papel clave en la formación del complejo STIM1-ORAI1 tras el vaciado de los depósitos. Por otro lado, a diferencia de STIM1, CRACR2A no es suficiente para inducir grandes corrientes CRAC cuando se co-expresa con ORAI1, lo que indica que no es un sustituto de STIM1. Sin embargo, sí actúa como sensor de Ca^{2+} gracias a sus dominios mano EF y, por ende, también podría participar en SOCE ya que, ante el aumento de $[\text{Ca}^{2+}]_{\text{cit}}$,

este se disocia tanto de ORAI1 como de STIM1 y el complejo se desestabiliza y, además, permite la unión de CaM-Ca^{2+} para inducir la inactivación de ORAI1 dependiente de Ca^{2+} (Srikanth et al. (2010), Dai et al. (2019a)).

Dai et al. (2019a) encontraron que pacientes con cáncer de próstata presentaban una elevada metilación en islas CpG en la región promotora de *CRACR2A*. Además, también observaron que estos niveles de metilación están inversamente correlacionados con la expresión de dicho gen. Es más, también encontraron una asociación entre la metilación de este gen en su región promotora y el tiempo de progresión hacia la metástasis letal. Afortunadamente, también observaron que los niveles de metilación de su región promotora de este gen disminuyen con la actividad física, lo que evidencia la relevancia de la actividad física como tratamiento preventivo frente al desarrollo de la metástasis. Por lo tanto, esta proteína podría actuar como un regulador de la progresión del cáncer de próstata a través de su interacción con ORAI1 y STIM1 que, recordemos, participan en la proliferación, migración e invasión celular en varios tipos de cáncer. En experimentos con células T en los que se emplean *CRACR2A* mutados en sus dominios EF mejoraron la oligomerización de STIM1 y, en consecuencia, se indujo la muerte de las células T como consecuencia de un aumento en $[\text{Ca}^{2+}]_{\text{cit}}$ (Srikanth et al. (2010), Dai et al. (2019a)).

En células endoteliales, como son las células de colon, se ha observado la presencia de una variante larga de esta proteína, *CRACR2A-L*, cuyo extremo C-terminal contiene una GTPasa Rab que podría participar en la angiogénesis a través de la formación del tubo endotelial (Wilson et al. (2015)). Curiosamente, la familia de GTPasas Rab, aparte de mediar tanto el tráfico a través de la membrana como la integridad de esta, también regulan el posicionamiento intracelular y la actividad de rutas de señalización relacionadas con el crecimiento celular, su supervivencia y la apoptosis. Es más, la expresión aberrante de proteínas Rab se ha descrito en diferentes tipos de cáncer tales como de pulmón, cerebro o mama. En cuanto a sus mutaciones, pueden alterar la red de tráfico vesicular que da lugar a una modulación del potencial tumorigénico, la migración celular y la capacidad metastásica. No obstante, estas proteínas también tienen un papel supresor tumoral ya que pueden inducir la apoptosis e inhibir la angiogénesis (Gopal Krishnan et al. (2020)).

- *MS4A12*: Como se dijo anteriormente, otro miembro de la misma subfamilia de *MS4A12*, el CD20 o *MS4A1*, que se expresa de forma específica en las células B, es un marcador del linfoma de células B y una diana terapéutica del tratamiento con mAb (Reff et al. (1994)). Por ello, Koslowski et al. (2008) encontraron, mediante el uso de métodos de minería de datos, una proteína estructuralmente similar a la proteína CD20 pero expresada de forma específica en las células epiteliales de colon, la proteína *MS4A12*. Desafortunadamente, al principio se pensó que esta proteína estaba sobreexpresada en el tejido de cáncer de colon (Koslowski et al. (2008)) e, incluso a día de hoy diferentes bases de datos como GeneCards inducen a pensar que su sobreexpresión favorece la progresión tumoral al mostrar en su descripción resultados experimentales sobre el efecto de su silenciamiento en el cáncer de colon: inhibición de la proliferación, motilidad e invasión celular (Koslowski et al. (2009)).

No obstante, en los últimos años se ha visto que, al contrario de lo que estos resultados indican, la expresión de MS4A12 está reducida en el cáncer de colon y, además, que a menor expresión pero es su pronóstico (J. Han et al. (2020), J. W. Han et al. (2021)). Es más, la disminución de su expresión inhibe la diferenciación de las células de cáncer de colon, característico de las células cancerígenas (L. He, Deng, and others (2017)). En cuanto a su localización celular, junto con STIM1 y ORAI1, parece ser diferente entre el tejido de colon normal y cancerígeno, pues a diferencia del tejido cancerígeno, en el tejido normal MS4A12 se localiza en la región apical (J. W. Han et al. (2021)). Es más, su localización celular aberrante, junto con la de ORAI1 y STIM1, podría contribuir a la progresión tumoral (J. W. Han et al. (2021)), al igual que sucede con otros tipos de canales de Ca^{2+} en los que su localización aberrante podría contribuir a la progresión tumoral, como por ejemplo el TRPM8 en el cáncer de próstata (Monteith et al. (2007), Bidaux et al. (2007)) o el Cav1.2 en el de colon (X.-T. Wang et al. (2000)). No obstante, es cierto que la expresión en el tejido de cáncer de colon de MS4A12, tanto a nivel de mRNA como de proteína, está muy reducida. Es más, el peor pronóstico y la menor supervivencia de los pacientes parece estar directamente relacionado con la menor expresión de MS4A12 (J. Han et al. (2020)).

3.9. Remodelado del transporte de Ca^{2+} mitocondrial

La mitocondria es un orgánulo clave en la homeostasis del Ca^{2+} intracelular y en procesos sumamente implicados en el cáncer, como el metabolismo o la apoptosis. No obstante, el mecanismo por el cual el Ca^{2+} mitocondrial regula los procesos tumorales no está todavía claro (Pathak et al. (2020)). Como se ha comentado en anteriores ocasiones, la mitocondria también juega un papel importante como regulador de SOCE evitando la inactivación de los SOCC dependiente de Ca^{2+} . En condiciones normales, la liberación del Ca^{2+} desde los depósitos activa SOCE a través de la interacción entre STIM1 y ORAI1, lo que da lugar a la corriente I_{CRAC} . Tras la entrada de Ca^{2+} desde el medio extracelular a través de ORAI1, esta puede ser inactivada por diferentes mecanismos (figura 3.4). Por un lado, el aumento de $[\text{Ca}^{2+}]_{\text{cit}}$ en un microambiente cercano a ORAI1 promueve una inactivación dependiente de Ca^{2+} . Por otro lado, la captación de Ca^{2+} por la mitocondria aumenta ROS en su interior que inactiva ORAI1 a través de la oxidación de su residuo Cys 195 (Ben-Kasus Nissim et al. (2017), Villalobos et al. (2018), Villalobos et al. (2019)). Además, se ha observado que en colonocitos normales las mitocondrias son importantes para la actividad de los SOCC, pues son capaces de modular la amplitud de I_{CRAC} , pero no pueden prevenir su inactivación dependiente de Ca^{2+} (Hernández-Morales et al. (2017)).

Sin embargo, en muchos tipos de cáncer se ha observado el aumento de SOCE y, además, que esta permanece mantenida en el tiempo o que, por lo menos, su inactivación es más lenta, tal y como se comentó en el apartado anterior, “3.8 Remodelado de SOCE”, lo que podría estar, en parte, favorecido por la acción de la mitocondria (Hernández-Morales et al. (2017)). Esta diferencia respecto de las células normales podría ser debida al efecto Warburg y a la acción conjunta del TRPC1 y del NCLX. Por un lado, el efecto Warburg conlleva un aumento del $\Delta\Psi$ que junto con el aumento en la expresión de MCU mejora la capacidad tampón de la mitocondria. Por lo tanto, es capaz de captar una mayor cantidad de Ca^{2+} citosólico y, por ende, de evitar la inactivación lenta de ORAI1 dependiente de Ca^{2+} (Villalobos et al. (2018), Villalobos et al. (2019)) (figura 3.4).

Por otro lado, la sobreexpresión de TRPC1 en muchos tipos de cáncer, como en el cáncer de colon, da lugar a una corriente mixta $I_{\text{CRAC}}/I_{\text{SOC}}$, de tal manera que ya no es selectiva por Ca^{2+} , sino que también entra Na^+ desde el medio extracelular. Esta entrada de Na^+ introducida por TRPC1 es aprovechada por el NCLX para sacar Ca^{2+} de la mitocondria y evitar así la sobrecarga de este catión en su interior. De hecho, se ha observado la sobreexpresión de NCLX en diferentes tipos de cáncer, entre ellos el de colon (Piao et al. (2009)). En consecuencia, este mecanismo podría evitar la formación de ROS provocada por el aumento de $[\text{Ca}^{2+}]_{\text{mit}}$ y, por ende, no tendrá lugar la inactivación de ORAI1 vía oxidación de la Cys 195 (Ben-Kasus Nissim et al. (2017), Villalobos et al. (2018), Villalobos et al. (2019)) (figura 3.4). Hernández-Morales et al. (2017) observó que en células de cáncer de colon (HT29) la mitocondria tiene un papel menos importante en la modulación de la amplitud de la corriente de entrada de Ca^{2+} ($I_{\text{CRAC}}/I_{\text{SOC}}$), pero sí lo tiene en la prevención de su inactivación dependiente de Ca^{2+} , lo que explicaría por qué estas células presentan una mayor SOCE y mantenida en el tiempo. Esto confiere a las células tumorales una serie de ventajas, como la resistencia a la apoptosis, pues evita la sobrecarga de Ca^{2+} en la mitocondria, y el aumento de la proliferación, pues se favorece una SOCE mayor y mantenida en el tiempo. Cabe destacar que la sobreexpresión de NCLX también podría regular la necrosis y la producción de ATP, lo que podría deberse a que esta proteína podría actuar de anclaje para la formación de un complejo con los ribosomas mitocondriales y regular la biogénesis mitocondrial (Piao et al. (2009)).

Se ha observado que la despolarización de la mitocondria con determinados AINE con salicilato, metabolito de la aspirina, o con AINE evita que la mitocondria impida la inactivación lenta dependiente de Ca^{2+} de SOCE (Villalobos et al. (2017), Villalobos et al. (2018)). Por ejemplo, Hernández-Morales et al. (2017) observó que el tratamiento de células de cáncer de colon con sulindac, que despolariza sus mitocondrias, promueve la inactivación lenta de I_{SOC} dependiente de Ca^{2+} y, como resultado, se reduce su amplitud y se inactiva su componente de entrada, lo que se traduce en la inhibición de SOCE.

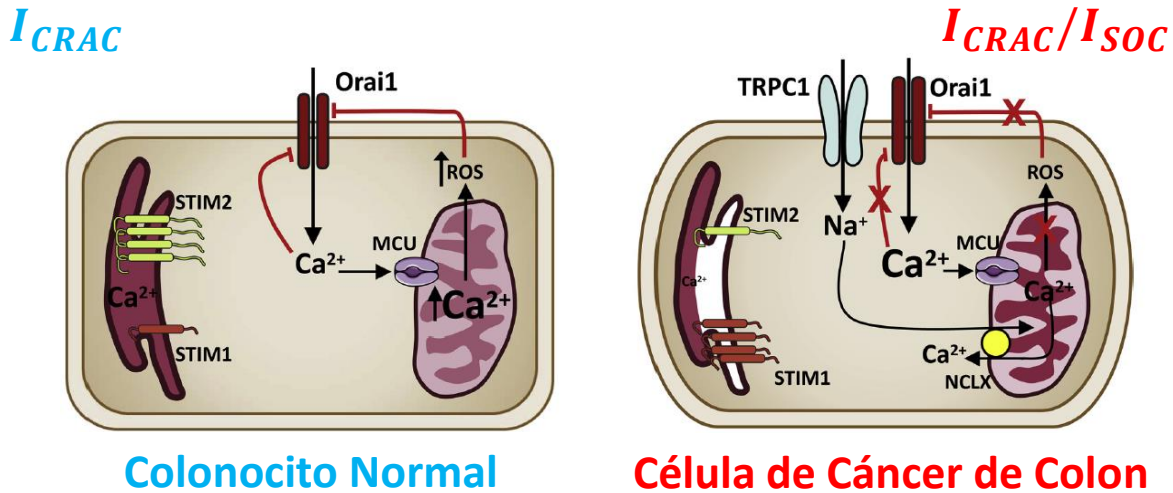


Figura 3.4: Control mitocondrial de SOCE en células normales vs. células de cáncer de colon. En los colonocitos normales, el vaciado de los depósitos intracelulares de Ca^{2+} promueve la interacción entre el sensor STIM1 de la membrana del RE y el ORAI1 de la MP y, por ende, activa SOCE. Posteriormente, ORAI1 se inactiva a través de una vía dependiente de Ca^{2+} o a través de la oxidación de la Cys 195 de ORAI1 por acción de los ROS generados en la mitocondria como consecuencia de la sobrecarga de Ca^{2+} mitocondrial. Además, la mitocondria es capaz de modular la amplitud de I_{CRAC} . Por el contrario, en las células de cáncer de colon, que presentan un menor tamaño de los depósitos debido posiblemente a la infraexpresión de STIM2, el vaciado de sus depósitos activa la entrada de Ca^{2+} a través no solo de ORAI1, sino también de TRPC1, sobreexpresado en estas células, lo que da lugar a una mezcla de corrientes $I_{\text{CRAC}}/I_{\text{SOC}}$. Debido a la sobreexpresión de MCU y al efecto Warburg, que genera un mayor $\Delta\Psi$ mitocondrial, aumenta la capacidad tampón de la mitocondria, lo que impide la inactivación lenta dependiente de Ca^{2+} de la corriente de entrada de Ca^{2+} . Puesto que TRPC1 permite la entrada de Na^+ , este se intercambia por el Ca^{2+} mitocondrial a través del NCLX y se impide la sobrecarga de Ca^{2+} mitocondrial, se evita la generación de ROS y no se inactiva ORAI1 por la oxidación de la Cys 195. En consecuencia, las células tumorales presentan una SOCE superior y mantenida en el tiempo respecto de las normales (modificado de Villalobos et al. (2019)).

No obstante, no siempre se observa una sobreexpresión del NCLX o del MCU en las células cancerígenas. Por un lado, en estadios más avanzados de la enfermedad se puede observar una reducción en la expresión de NCLX. Esta baja expresión se ha asociado con mutaciones en TP53 y tumores BRAF wild type. También se ha observado una reducción de NCLX desde el estadio I al IV. Mediante experimentos NCLX KO en ratones, se ha observado que esta menor expresión provoca sobrecarga de Ca^{2+} en la mitocondria y su despolarización y reduce la expresión de genes del ciclo celular en ratones. También conduce a la diseminación metastásica, quimiorresistencia y expresión de rutas de señalización epitelio-mesenquimal, hipoxia y células madre. En términos mecanísticos, el aumento de Ca^{2+} mitocondrial conduce al incremento de ROS que activa la señalización por HIF1 α que favorece la metástasis de las células con baja expresión de NCLX (Pathak et al. (2020)). Por otro lado, también se ha observado la disminución en la expresión de MCU como consecuencia de la sobreexpresión del miARN25 en cáncer de colon cuya consecuencia directa es la menor capacidad de este para actuar como tampón de Ca^{2+} y, en consecuencia, ser más resistente a la muerte celular apoptótica inducida por una sobrecarga de Ca^{2+} mitocondrial (Marchi et al. (2013)).

En cuanto a los VDAC, se ha encontrado una estrecha relación entre estos canales y el cáncer, pues el aumento del flujo de Ca^{2+} a través de los VDAC está estrechamente relacionado con el estrés oxidativo y la apoptosis, lo que podría explicar por qué la mayoría de las células cancerígenas hayan perdido estos canales. No obstante, en otros tumores se ha encontrado el aumento en la expresión de estos, lo que podría deberse al papel de estos en la regulación metabólica y de la apoptosis. En concreto, se ha observado que VDAC1 es capaz de unirse a HK (de las dos isoformas, HK-II suele estar sobreexpresada en varios tipos de cáncer) para proteger a las células tumorales de la apoptosis y favorecer su proliferación (Camara et al. (2017)). En este sentido, dicha unión podría aumentar la glucólisis y promover el crecimiento celular. Es más, también se ha observado que previene la apoptosis por la vía intrínseca y atenúa la inducción de esta por ROS al disminuir la producción de mROS; por el contrario, si HK se separa entonces se induce la liberación de citocromo c. Así, se han investigado diferentes estrategias terapéuticas basadas en impedir dicha unión. Otra unión que también es una posible diana terapéutica es VDAC-tubulina. También se ha encontrado que la liberación del anión O_2^- tiene preferencia por VDAC1, por lo que este canal también podría regular la señalización redox en las células tumorales al promover el flujo de mROS. Por lo tanto, se ha investigado el uso de bloqueantes de VDAC para el tratamiento del cáncer los cuales han mostrado inducir la apoptosis al impedir el flujo de sustratos a su través, como el ATP/ADP (Fang and Maldonado (2018), Delierneux et al. (2020), Varughese, Buchanan, and Pitt (2021)). En cuanto VDAC2, actúa en sentido contrario a VDAC1 en la apoptosis, pues, por ejemplo, se une a ceramida y desencadena la apoptosis mitocondrial (Dadsena et al. (2019)).

Por tanto, numerosas evidencias indican un gran remodelado de la homeostasis del Ca^{2+} intracelular en cáncer con grandes diferencias entre los distintos tipos de cáncer. Se hace necesario, pues, un estudio sistemático del remodelado del remodelado que abarque la expresión de todos los genes implicados en la homeostasis del Ca^{2+} intracelular.

Parte II

Hipótesis y Objetivos

“La verdadera ciencia nunca es especulativa, sino que emplea hipótesis que sugieren puntos de investigación, pero nunca adopta las hipótesis como si fueran proposiciones demostradas”.

Cleveland Abbe

“La verdad en ciencia puede ser definida como la hipótesis de trabajo que mejor se ajusta para abrir el camino a la siguiente mejor ajustada”.

Konrad Lorenz

Capítulo 4

Hipótesis de trabajo y objetivos

La hipótesis de partida del trabajo presentado es que el cáncer de colon o colorrectal está asociado a un remodelado de la homeostasis del Ca^{2+} intracelular caracterizado por diferencias en la expresión y/o la actividad de genes implicados en el transporte de Ca^{2+} intracelular. Estos cambios, además, podrían ser esenciales en la adquisición del fenotipo tumoral caracterizado por *cancer hallmarks* que incluirían a la proliferación descontrolada, la adquisición de la capacidad de migración e invasión y la resistencia a la muerte celular. Finalmente, el remodelado de la homeostasis del Ca^{2+} intracelular podría deberse a la inactivación del gen supresor de tumores APC que conlleva la activación de la ruta Wnt/ β -Catenina y el oncogén *MYC* que promueve la sobreexpresión de la ornitina descarboxilasa (ODC), la enzima limitante en la síntesis de poliaminas que es inhibida de modo suicida por DFMO que provoca la depleción de poliaminas y podría revertir parcialmente el remodelado de la homeostasis del Ca^{2+} intracelular. Por tanto, el objetivo fundamental de esta tesis es llevar a cabo el análisis transcriptómico del posible remodelado de la homeostasis del Ca^{2+} intracelular en el cáncer colorrectal y su posible reversión por depleción de poliaminas. Para alcanzar este objetivo global, nos hemos planteado abordar los siguientes objetivos específicos:

1. Establecer el posible remodelado de la homeostasis del Ca^{2+} intracelular mediante el análisis transcriptómico masivo y análisis estadístico de los genes implicados en el transporte de Ca^{2+} en células modelo normales y tumorales de colon humano utilizando diferentes metodologías de secuenciación masiva y microarrays.
2. Determinar los efectos de la depleción de poliaminas por tratamiento con DFMO sobre la homeostasis del Ca^{2+} intracelular y sus bases transcriptómicas en células modelo normales y tumorales de colon humano.
3. Establecer el posible remodelado de la homeostasis del Ca^{2+} intracelular mediante el análisis transcriptómico masivo y análisis estadístico así como la expresión de proteínas implicadas en la entrada capacitativa de Ca^{2+} en muestras pareadas de colon normal y tumoral procedentes de pacientes de cáncer colorrectal.

4. Utilizar los datos crudos de repositorios públicos de análisis transcriptómicos de muestras de cáncer colorrectal humano junto con nuestro análisis transcriptómico para llevar a cabo el primer metanálisis del remodelado de la homeostasis del Ca^{2+} intracelular en cáncer de colon a partir de los datos reanalizados extraídos de dichos repositorios.

Parte III

Material y Métodos

“Consultar a un estadístico después de que haya concluido el experimento es, muy a menudo, pedirle que realice un examen post-mortem. Quizás pueda decir de qué murió el experimento”.

Ronald Fisher (1938)

Capítulo 5

Material

En este capítulo se van a presentar los materiales empleados en los diferentes experimentos realizados durante la investigación de la presente tesis.

5.1. Modelos de estudio

En este apartado se van a presentar los diferentes modelos empleados a lo largo de la tesis: modelos celulares, modelos de tejido humano procesados por nuestro grupo y modelos de tejido humano extraídos de repositorios públicos.

5.1.1. Modelos celulares

Como modelos celulares humanos se han empleado dos ampliamente usados y validados en el campo de la investigación fisiopatológica del cáncer de colon: células HT29 como modelo de cáncer de colon y células NCM460 como modelo de colon normal. Por un lado, las células HT29, usadas como modelo de adenocarcinoma y adquiridas en LONZA (Basilea, Suiza), proceden de la parte recto-sigmoidea del intestino de una mujer de 44 años (Fogh and Trempe (1975), Kawai et al. (2002)). Estas células expresan marcadores típicos de cáncer de colon como COX-2 y mutaciones en APC, BRAF, PIK3CA o TP53 (Ahmed et al. (2013), Brocardo, Näthke, and Henderson (2005)). Por otro lado, las NCM460, usadas como modelo de células epiteliales de colon normal y obtenidas a través de acuerdo con INCELL Corporation (San Antonio, TX), proceden de tejido normal de un donante hispano varón de 68 años con cáncer gástrico. Estas expresan marcadores característicos del epitelio como citoqueratina y vilina; además, no crecen en agar¹, lo que confirma su fenotipo no tumoral (Moyer et al. (1996), Fuertes et al. (2013), Gil, López, and Pellin (1989)).

¹El cultivo *in vitro* en agar semisólido ha sido uno de los procedimientos más usados para mejorar el rendimiento de los cultivos tumorales en el sentido de obtener una menor presencia de células normales tanto fibroblásticas como epiteliales (Gil, López, and Pellin (1989)).

5.1.2. Modelos de tejido humano

Como modelos más cercanos a la realidad del CCR humano respecto de las líneas celulares se emplearon muestras de tejido humano de cáncer de colon y de mucosa normal. No obstante, en este trabajo estos modelos pueden dividirse en dos tipos según su procedencia: propios y externos.

- Los modelos considerados como propios son muestras de tejido humano de cáncer de colon y de mucosa normal proporcionadas por el Banco de Tumores de Castilla y León. En concreto, se emplearon 20 muestras pareadas donadas por 10 pacientes; es decir, que de cada uno de los 10 pacientes que han donado la muestra de tejido de cáncer de colon también se dispone de la muestra de mucosa normal de ese mismo paciente. En adelante, a los experimentos realizados con dichas muestras se les denominará, en adelante, como *IBGM*.
- Como externas se consideran a aquellas muestras de tejido humano de cáncer de colon y normal que han sido empleadas en experimentos transcriptómicos que no han sido procesadas por nuestro laboratorio, sino por otros grupos de investigación y cuyos datos originales o experimentos han sido publicados y depositados en el repositorio público NCBI GEO. Una de las entidades de NCBI GEO se denomina *Series*², cada una de las cuales tiene un identificador único de la forma *GSE[1-9]{3,6}* (en el enlace Series pueden consultarse las *Series* disponibles). En cada una de las *Series* se encuentran agrupadas muestras relacionadas entre sí (de un mismo experimento, por ejemplo), se describe el experimento y el diseño, se da información sobre cada una de las muestras y proporciona los enlaces de descarga de diferentes tipos de datos, entre ellos los datos originales sin procesar. En concreto, de todos los publicados se seleccionaron aquellos experimentos que cumplieron los siguientes de criterios de inclusión:
 1. Cada experimento debe contar con muestras de tejido humano tanto de cáncer de colon como normal.
 2. Cada condición experimental ha de contar con al menos 5 réplicas.
 3. Los microarrays empleados han de ser de *Affymetrix*.
 4. Los datos transcriptómicos originales *.CEL* deben ser accesibles para poder reanalizarlos desde cero.

En concreto, los datos originales sin procesar (*.CEL* en el caso de microarrays de *Affymetrix*) son los que realmente se usan en la presente tesis, puesto que estos son los que proporciona el *software* del equipo que escanea los microarrays y que son los datos a partir de los cuales se comienza cualquier análisis de este tipo de experimentos. Así, se han extraído un total de 1231 microarrays de 5 experimentos diferentes (tabla 5.1).

²Una de las entidades de NCBI GEO se denomina *Series*, y en ella pueden encontrarse agrupadas muestras relacionadas entre sí (de un mismo experimento, por ejemplo), se describe el experimento y el diseño, se da información sobre cada una de las muestras y proporciona los enlaces de descarga de diferentes tipos de datos, entre ellos los datos originales sin procesar, que son los que nos interesan. Cada serie tiene asignado un identificador único de la forma *GSE[1-9]{3,6}* (en el enlace Series pueden consultarse las *Series* disponibles).

Tabla 5.1: Series GEO seleccionadas.

| Series | Microarrays | | Pareado | Plataforma | Referencia |
|----------|-------------|---------|-----------------|---|-------------------------------|
| | Normal | Tumoral | | | |
| GSE8671 | 32 | 32 | Sí | Affymetrix Human Genome U133 Plus 2.0 Array | Sabates-Bellver et al. (2007) |
| GSE39582 | 19 | 566 | No | Affymetrix Human Genome U133 Plus 2.0 Array | Marisa et al. (2013) |
| GSE33113 | 6 | 90 | No | Affymetrix Human Genome U133 Plus 2.0 Array | Felipe de Sousa et al. (2011) |
| GSE44076 | 148 | 98 | Si ¹ | Affymetrix Human Genome U219 Array | Solé et al. (2014) |
| GSE41258 | 54 | 186 | Si ¹ | Affymetrix Human Genome U133A Array | Sheffer et al. (2009) |

Notas:

¹ No todas las muestras son pareadas.

5.2. Medios, reactivos y materiales de cultivo celular

A continuación, se especifican los diferentes materiales empleados en los cultivos celulares, tanto los generales como los específicos de los experimentos:

- Medio DMEM con 1 g/L glucosa, suero bovino fetal (FBS, del inglés *fetal bovine serum*), los antibióticos (penicilina/estreptomicina), L-glutamina, un tampón fosfato salino (PBS, del inglés *phosphate buffer saline*) y la tripsina y ácido etilendiaminotetraacético (EDTA), el cual ha sido suministrado por LONZA (Verbiers, Bélgica).
- Poli-L-lisina y cubreobjetos de 12 mm de diámetro proporcionados por Marlenfeld GmbH (Lauda-Könlghshofen, Alemania).
- Frascos de cultivo de 25 cm² y las placas de cultivo de 4 pocillos de la casa comercial NUNC (Thermo Scientific, Waltham, MA).
- Placas de cultivo de 10 cm de diámetro de Corning (Corning, NY).
- DFMO de TOCRIS (Bristol, UK).

5.3. Medios de imagen de Ca²⁺

- Medio externo completo (MEC): NaCl 145 mM, KCl 5 mM, MgCl₂ 1 mM, CaCl₂ 1 mM, D-glucosa 10 mM, HEPES 10 mM, pH 7.42.
- MEC libre de Ca²⁺: NaCl 145 mM, KCl 5 mM, MgCl₂ 1 mM, ácido etilenglicol tetraacético (EGTA) 0.5 mM, D-glucosa 10 mM, HEPES 10 mM, pH 7.42.

5.4. Reactivos y materiales para la extracción de proteínas y *Western Blot*

- Tampón de muestra 2x: Tris-HCl 0.5 M (pH 6.8), SDS (del inglés, *sodium dodecyl sulfate*) 2%, glicerol 10%, β -mercaptoetanol 5%, Azul de bromofenol 0.05%.
- Tampón de electroforesis 10x: Trizma base 0.3 M, glicina 192 mM, SDS 0.1% (pH 8.3)
- Tampón de transferencia: Trizma base 0.25 M, glicina 192 mM, metanol 20% (pH 8.3)
- TBS 10x: Tris-HCl 0.25 M, NaCl 0.9% (pH 7.5)
- TTBS: TBS 10%, Tween20 0.1%.
- Buffer de bloqueo BSA 5%: leche desnatada en polvo, TTBS
- Tampón de electroforesis 10x: Trizma base 0.24 mM, glicina 192 mM, SDS 0.1% (pH 8.3)
- SDS-PAGE (10%, del inglés, *sodium dodecyl sulfate polyacrylamide gel electrophoresis*): 30% Acrilamida/BisAcrilamida, Tris-HCL 0.5 M (pH 6.8), SDS 10%, APS y TEMED

La Acrilamida/bisacrilamida 30% es de BioRad (Hercules, CA). La membrana de PVDF es de Millipore (Bedford, MA). El kit PIERCE ECL Western Blotting substrate es de Thermo Scientific (Waltham, MA). El TRIZOL y el Tween20 son de Merck (Steinheim, Alemania). El APS y el TEMED son Fluka (Merck, Steinheim, Alemania)

Los anticuerpos primarios utilizados fueron los siguientes (tabla 5.2):

Tabla 5.2: Anticuerpos primarios utilizados.

| Proteína | Peso Molecular (KDa) | Anticuerpo | Tipo | Dilución | Casa Comercial |
|-----------------|----------------------|---------------------|------------|----------|----------------|
| STIM1 | 77.423 | Ratón anti-hSTIM1 | monoclonal | 1:5000 | Abcam |
| STIM2 | 83.971 | Conejo anti-hSTIM2 | monoclonal | 1:500 | Abcam |
| ORAI1 | 32.668 | Conejo anti-hORAI1 | Policlonal | 1:500 | Alomone Labs |
| ORAI2 | 28.570 | Conejo anti-hORAI2 | Policlonal | 1:1000 | Abcam |
| TRPC1 | 91.212 | Conejo anti-hTRPC1 | Policlonal | 1:1000 | Abcam |
| β -actina | 41.737 | Ratón anti-B-actina | monoclonal | 1:500 | Abcam |

El anticuerpo secundario conejo anti-ratón HRP (Abcam, Cambridge, UK) fue utilizado a una dilución 1:5000 para el revelado de las proteínas TRPC1 y B-actina, mientras que para el resto de proteínas se usó el anticuerpo ECL burro anti-conejo HRP (es de GE Healthcare, UK) se utilizó a una concentración de 1:10000. Tanto los anticuerpos primarios como los secundarios fueron diluidos en una solución de TTBS con BSA 0.5%.

5.5. Reactivos y materiales para la extracción de ARN

- Kit de extracción de ARN GeneMATRIX Universal RNA Purification Kit de EURx (Gdansk, Polonia).

5.6. Sondas, reactivos de las soluciones y otros compuestos

- Sonda de Ca^{2+} Fura2/AM es de Invitrogen (Eugene, OR).
- NaCl, NaOH, CaCl_2 , D-glucosa son de Merck (Darmstadt, Alemania).
- Etanol absoluto, isopropanol, cloroformo, Trizma base, β -mercaptoetanol, glicina, SDS, Azul de Bromofenol y BSA, HCl, metanol, SDS, EGTA y glicerol obtenidos de Sigma-Aldrich (Steinheim, Alemania).
- Ácido ciclopiazónico (CPA) es de Sigma-Aldrich (Steinheim, Alemania).
- El H_2O DEPC son de Invitrogen (Carlsbad, CA).

5.7. Bases de datos de conjuntos de genes

Las técnicas transcriptómicas de alto rendimiento proporcionan grandes cantidades de datos e información. Por un lado, se tiene una matriz de expresión de una gran dimensión y, por otro lado, una lista de miles de genes en los que se muestra su expresión diferencial entre las condiciones experimentales evaluadas. Una de las estrategias más habituales es realizar un análisis de componentes principales (ver apéndice “A.3 Análisis de componentes principales”). No obstante, en los datos transcriptómicos es posible usar la información generada por la comunidad científica sobre diferentes conjuntos de genes que tienen un significado biológico: rutas de señalización, conjuntos de genes relacionados con alguna patología, localización celular de las proteínas que codifican, etc.

Existen multitud de bases de datos con esta información y, además, dentro de cada una puede haber diferentes colecciones de conjuntos de genes agrupados en función de la temática que los relaciona. Así, un mismo gen puede pertenecer a multitud de diferentes conjuntos de genes, de tal manera que a cada gen se le asigna un identificador o anotación por cada conjunto de genes al que pertenezca y que será diferente en función de la base de datos que describa cada conjunto. En particular, en la presente tesis se ha recurrido a las siguientes:

- KEGG, del inglés *Kyoto Encyclopedia of Genes and Genomes*, es una colección de bases de datos y herramientas web. En particular, una de estas bases de datos es KEGG PATHWAY la cual contiene redes de interacciones moleculares dentro de las células y, en consecuencia, agrupa los genes en función de diferentes criterios: rutas metabólicas, procesamiento de la información genética, procesos celulares, sistemas orgánicos, enfermedades o desarrollo de fármacos. Para cada uno de estos criterios hay definidos diferentes conjuntos de genes. En total, contiene 345 conjuntos de genes.
- GO, del inglés *gene ontology*, contiene un criterio para describir a los diferentes genes así como los productos de estos contiene tres ontologías en base a las cuales se pueden clasificar los genes: función molecular de los productos de los genes (*molecular function*), proceso biológico en los que participan (*biological process*) o su localización componente celular (*cellular component*). Para cada una de estas ontologías hay definidos numerosos conjuntos de genes. De ellas, en la presente tesis se han usado *biological process* y *molecular function* con 12342 y 4245 conjuntos de genes, respectivamente.
- MSigDB, del inglés *The Molecular Signatures Database*, es una colección de conjuntos de genes que, a su vez, pueden dividirse en 9 colecciones de conjuntos de genes (Subramanian et al. (2005), Liberzon et al. (2011)):
 - *Hallmark (H)*: representan estados o procesos biológicos bien definidos.
 - *Positional gene sets (C1)*: para cada cromosoma humano y banda citogenética.
 - *Curated gene sets (C2)*: de bases de datos de rutas de señalización online, de publicaciones en PubMed y del conocimiento de expertos en la materia.
 - *Regulatory target gene sets (C3)*: basado en predicciones de genes diana para secuencias de semillas de microARN y sitios de unión de factores de transcripción predichos.
 - *Computational gene sets (C4)*: definidos por minería de grandes colecciones de datos de microarrays orientados al cáncer.

 - *Ontology gene sets (C5)*: formados por genes anotados con la misma ontología term.
 - *Oncogenic signature gene sets (C6)*: definidos a partir de microarrays de genes alterados en el cáncer.
 - *Immunologic signature gene sets (C7)*: representan estados celulares y alteraciones del sistema inmune.
 - *Cell type signature gene sets (C8)*: extraídos de agrupaciones generadas en estudios *single-cell* en tejidos humanos.

De las anteriores, en la presente tesis se han usado H, C2, C4, C6 y C7, con 50, 6290, 858, 189 y 5219 conjuntos de genes, respectivamente.

Capítulo 6

Métodos

En este capítulo se va a presentar la metodología empleada a lo largo de la tesis. Puesto que en un sólo experimento pueden emplearse diferentes metodologías, estas se muestran de forma independiente sin relacionarlas directamente con ninguno de ellos.

6.1. Cultivos Celulares

En los cultivos celulares se han empleado las líneas celulares HT29 y NCM460 (ver apartado “5.1.1”). Puesto que se han empleado en diferentes experimentos, en función de cada uno de ellos el soporte sobre el que son sembradas y su densidad es diferente. Así, mientras que para los experimentos de imagen de Ca^{2+} , sobre cada cubreobjetos pretratado con poli-L-lisina se siembran 3000 células por pocillo, para los experimentos de extracción de ARN se siembran $10 \cdot 10^5$ células HT29 o $10.5 \cdot 10^5$ células NCM460 en placas de 10 cm de diámetro. Además, para el propio mantenimiento de los cultivos, las células se siembran en *flask* de 25 cm². En todos los casos, para el mantenimiento de los cultivos se ha utilizado el medio DMEM 1g/L Glucosa suplementado con 1 % de penicilina-estreptomicina, 1 % L-Glutamina y 10 % FBS en un incubador a 37 °C y con una atmósfera de 10 % CO₂.

6.2. Tratamiento de células NCM460 y HT29 con DFMO

Las alteraciones en la ruta de señalización Wnt son la forma más frecuente de aparición de CCR, razón por la cual numerosos tratamientos están dirigidos contra esta. Una de las consecuencias directas de alteraciones en esta ruta es el aumento en la expresión del oncogen *MYC* el cual, a su vez, regula otros genes como *ODC* que participa en la ruta de síntesis de las poliaminas. Como consecuencia, esta ruta está sobreactivada en el CCR, por lo que su inhibición con DFMO se ha propuesto como posible diana terapéutica.

Por ello, tanto a nivel transcriptómico -microarrays- como a nivel funcional -imagen de Ca^{2+} -, se estudió el efecto de la inhibición de la síntesis de poliaminas con DFMO en los modelos celulares HT29 y NCM460. No obstante, puesto que uno de los intereses es encontrar un posible perfil de expresión capaz de explicar los resultados obtenidos en los experimentos de imagen de Ca^{2+} , las unidades experimentales de ambos experimentos fueron sometidas a las mismas condiciones externas no controlables. Para ello, cada una de las réplicas con un tratamiento “g” (NCM460.Control, NCM460.DFMO, HT29.Control o HT29.DFMO) de un experimento “A” (imagen de Ca^{2+}) fue procesada en paralelo a una muestra para ese mismo tratamiento del experimento “B” (microarrays). Es decir, que a partir del mismo cultivo o *flask* se siembran células para la misma condición experimental o tratamiento de ambos experimentos: una placa de 10 cm de diámetro para la condición “g” del experimento transcriptómico y dos cristales de la misma condición “g” para el experimento de imagen de Ca^{2+} (figura 6.1). Además, las unidades experimentales de una misma línea celular, tanto para la condición experimental tratada como para la no tratada, provienen también del mismo *flask*.

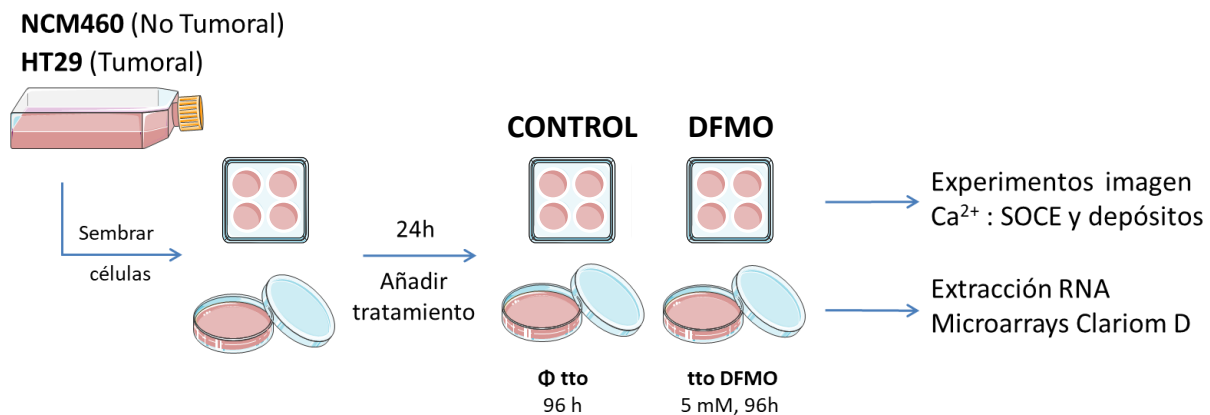


Figura 6.1: Diseño experimental para el análisis transcriptómico y funcional del efecto del DFMO en las líneas celulares HT29 y NCM460. Para cada línea celular, a partir de un mismo cultivo (*flask*) se siembran cuatro cubreobjetos, en una placa de seis pocillos, y dos placas de 10 cm de diámetro. Pasadas 24 horas, se añade el tratamiento (control o DFMO 5 mM). Tras 96 horas, de las unidades muestrales correspondientes a las placas de 10 cm se extrae su ARN para los experimentos transcriptómicos, mientras que las unidades muestrales sembradas en los cubreobjetos se emplean para los experimentos de imagen de Ca^{2+} .

Hay que destacar que ambos experimentos comparten una parte común del protocolo. En concreto, la parte común del protocolo es la siguiente: i) para cada condición experimental las células (HT29 o NCM460) se siembran sobre una placa de 10 cm de diámetro (para el experimento transcriptómico) y 2 cristales (para imagen de Ca^{2+}) ii) tanto las 4 placas de 10 cm como los 8 cristales se cultivan en un incubador a 37 °C y 10% de CO_2 iii) tras 24 horas, en función de la condición experimental, se les añade o bien el tratamiento con DFMO (DFMO 5 mM en DMEM 1 g/L Glucosa más 10% FBS dializado) o bien el medio sin tratamiento (DMEM 1 g/L Glucosa más

10 % FBS dializado) iv) pasadas 96 horas tras el tratamiento (o el medio sin tratamiento), se extrae el ARN de las células sembradas en placas y, ese mismo día, se realizan experimentos de imagen de Ca^{2+} con las células sembradas en los cubreobjetos o cristal.

El diseño experimental que se planteó es muy similar para ambos experimentos salvo ligeras sutilezas. Por un lado, el diseño del experimento de imagen corresponde con un modelo lineal mixto con dos factores fijos -*Línea Celular* a dos niveles (*NCM460* y *HT29*) y *Tratamiento* a dos niveles (*Control* y *DFMO*)- y dos aleatorios -*Día*, correspondiente al día de realización del experimento, y *Cristal*, que se refiere a la unidad experimental y que agrupa los registros de todas las células sembradas en él. Por otro lado, el experimento transcriptómico es igual que el anterior salvo que carece del factor aleatorio *Cristal*, puesto que el ARN de las diferentes células de cada una de las placas se extrae de forma conjunta. Por lo tanto, se distinguen cuatro condiciones experimentales: *HT29.Control*, *HT29.DFMO*, *NCM.Control* y *NCM.DFMO*. En cuanto a la parte específica de cada uno de los experimentos, a continuación se describe el procedimiento llevado a cabo:

6.3. Imagen de Ca^{2+} intracelular en célula única

En este apartado se va a realizar una introducción a la técnicas de imagen de Ca^{2+} intracelular en célula única y, a continuación, se detallará el procedimiento y diseño experimental seguido en los experimentos realizados por nuestro grupo de investigación. Por último, se explicará brevemente el análisis de datos llevado a cabo.

6.3.1. Fundamentos teóricos

La técnica de imagen de Ca^{2+} es una técnica microscópica de fluorescencia que permite monitorizar los cambios en la $[\text{Ca}^{2+}]_{\text{cit}}$ a nivel de célula única; es decir, los cambios que tienen lugar en cada una de las células que caen dentro del campo óptico. No obstante, para registrar estos cambios en la concentración de Ca^{2+} libre intracelular es necesario el uso de sondas fluorescentes y, en particular, para registrar el Ca_{cit} la sonda más comúnmente empleada es Fura2 pentaacetoximetilester (Fura2/AM). Esta se basa en la molécula Fura2 la cual es un derivado del quelante de Ca^{2+} EGTA cuyos grupos aromáticos le otorgan las propiedades fluorescentes. No obstante, debido a su polaridad, la molécula Fura2 es incapaz de atravesar la MP pero sí su derivado acetoximetil éster, Fura2/AM, el cual fue desarrollado en la década de los 80 por el grupo de Tsien, Premio Nobel de química en 2008 por el desarrollo de la proteína fluorescente verde (GFP, del inglés *green fluorescent protein*) (Grynkiewicz, Poenie, and Tsien (1985)). En consecuencia, el Fura2/AM es capaz de difundir a través de la membrana celular a su interior donde las esterasas endógenas hidrolizan, escinden el grupo AM y, en consecuencia, devuelven como producto de la reacción a la molécula Fura2 en su forma de ácido libre la cual queda retenida en su interior (figura 6.2a).

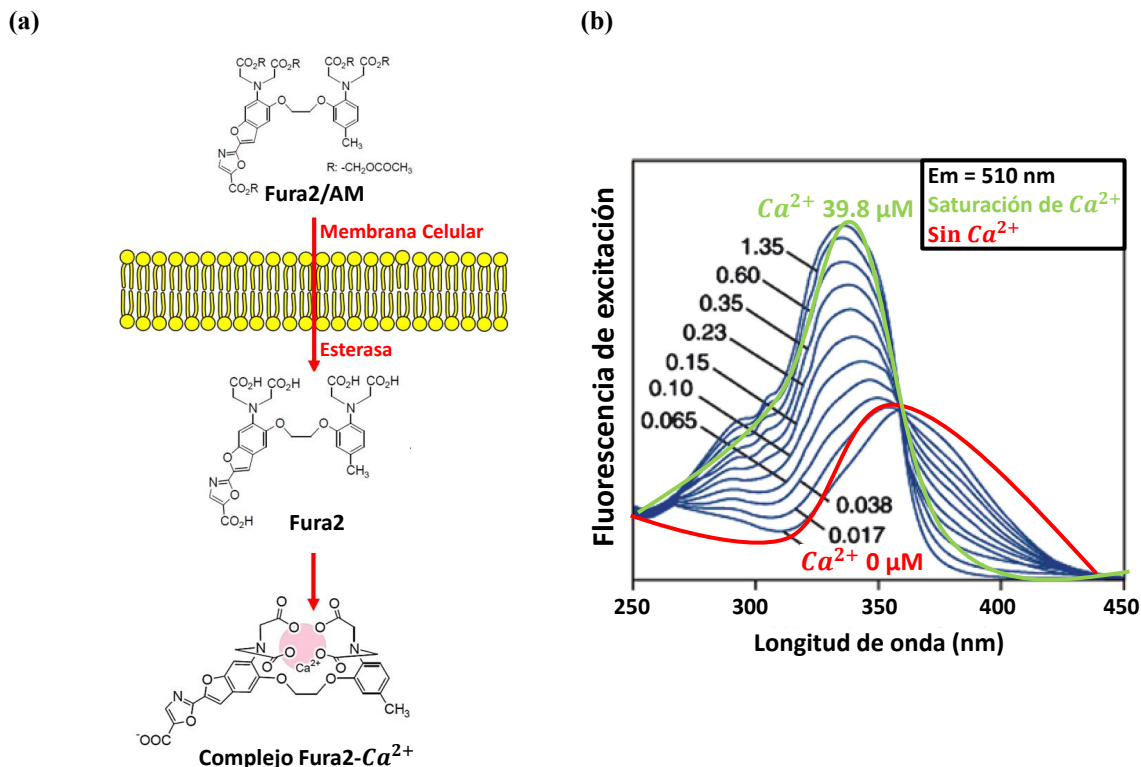


Figura 6.2: Sonda ratiométrica Fura2 y el proceso de entrada a la célula de su forma Fura2/AM. a) **Entrada de Fura2/AM a la célula:** la forma Fura2/AM de la sonda Fura2 es capaz de atravesar la MP celular mediante difusión simple; una vez dentro, las esterasas intracelulares hidrolizan Fura2/AM y se obtiene la sonda Fura2 capaz de complejar Ca²⁺. b) **Espectro de excitación del Fura2:** El espectro de excitación registrado a 510 nm de esta sonda le confiere su propiedad ratiométrica, ya que la forma complejada Fura2-Ca²⁺ muestra un máximo de excitación a 340 nm (línea verde), mientras que la forma libre Fura2 lo hace a 380 nm (línea rojo). Esta dualidad permite observar cómo el aumento de [Ca²⁺]_{cit} aumenta la intensidad de fluorescencia registrada a 340 nm mientras que disminuye a 380 nm, y viceversa (modificado de Shaalan, Carpenter, and Proctor (2017) y de Mammano and Bortolozzi (2010)).

Una característica muy interesante de la sonda Fura2 es su espectro de excitación, pues en su forma libre de Ca²⁺ el espectro de excitación registrado a 510 nm muestra un máximo de excitación a 380 nm para la forma libre Fura2 (figura 6.2b). En consecuencia, el aumento de [Ca²⁺]_{cit} se traduce en un aumento en la intensidad de fluorescencia a 340 nm y una disminución a 380 nm, mientras que la disminución produce el efecto contrario (Shaalan, Carpenter, and Proctor (2017), Mammano and Bortolozzi (2010), Grynkiewicz, Poenie, and Tsien (1985)). Debido a esta peculiaridad, el Fura2 es considerada como una sonda ratiométrica, puesto que es posible estimar la [Ca²⁺]_{cit} a partir del cociente de intensidades de fluorescencia 340/380. Una consecuencia directa de su propiedad ratiométrica es que muchas variables no controlables afectan por igual a la intensidad de fluorescencia correspondiente a ambas longitudes de onda de excitación, por lo que permite corregir ciertos artefactos debidos a estas variables, tales como el espesor celular, la cantidad de colorante acumulado en su interior, al desigualdad en la distribución del colorante o *photobleaching*, de las fluctuaciones en la intensidad de luz de excitación debido al deterioro progresivo de la lámpara.

Para llevar a cabo un experimento de imagen típico, las células son previamente sembradas en cubreobjetos de 12 mm de diámetro pretratados con poli-L-lis (0,01 mg/mL). Transcurrido el tiempo necesario para su adhesión, estas se incuban con la sonda ratiométrica Fura2/AM 4 μ M en medio MEC durante 45 minutos. Transcurrido este tiempo, los cubreobjetos se colocan en una cámara de perfusión que, a su vez, se fija en la pletina del microscopio de fluorescencia invertido Zeiss Axiovert S100TV. Además, el microscopio cuenta con un sistema de perfusión (Warner instruments, Hamden CT), lo que permite estimular a las células con el medio deseado a una velocidad y temperatura constante (37 °C).

En cuanto a la excitación de la sonda, las células se iluminan con una lámpara de Xenon que, con ayuda de una rueda de filtros de excitación (Sutter Instruments, Novato, CA), es capaz de alternar entre luz monocromática a 340 y a 380 nm que, gracias a un espejo dicróico, esta se desvía hacia la cámara que contiene el cubreobjetos con las células. La fluorescencia emitida por las células es recogida mediante un objetivo 40x (de inmersión en aceite y de apertura numérica 1.3) y esta atraviesa el espejo dicróico el cual refleja longitudes de onda superiores a 520 nm y las de menor longitud de onda que logran pasar son filtradas por un filtro de emisión a 520 nm. Finalmente, la fluorescencia que ha llegado hasta este punto es la imagen final la cual es capturada por una cámara Orca-ER de Hamamatsu Photonics (Hamamatsu, Japón) y procesada por un ordenador con un *software* capaz de procesar imágenes digitales y de analizarlas (figura 6.3).

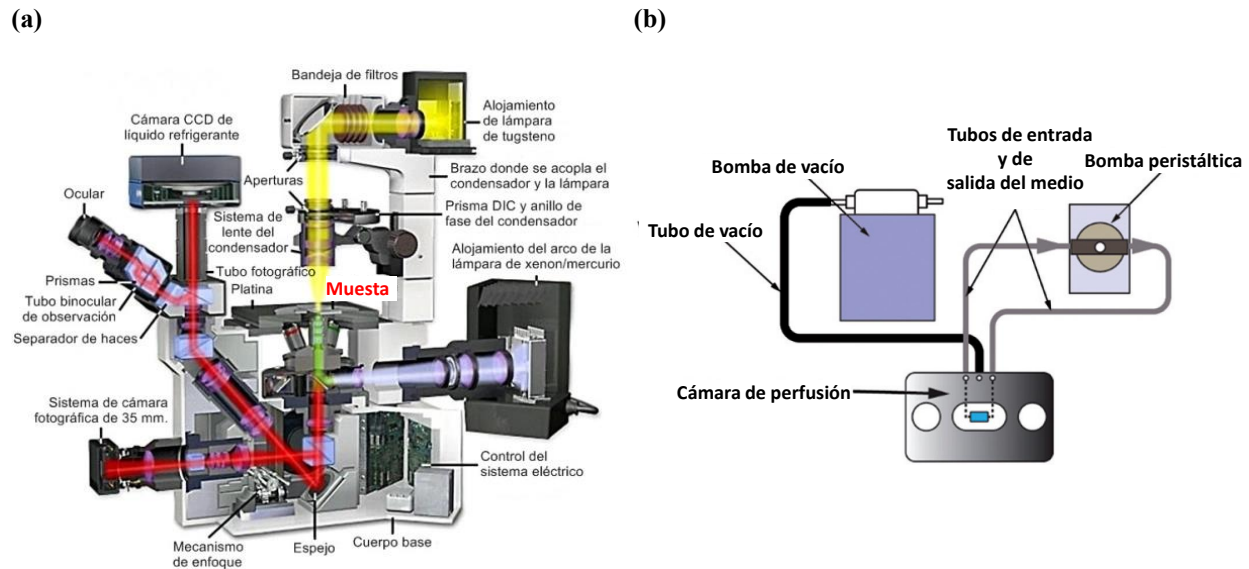


Figura 6.3: Esquema del equipo de imagen de fluorescencia. a) Microscopio de fluorescencia: los equipos de fluorescencia, a parte de los componentes necesarios para poder obtener esta, también cuentan con un sistema de iluminación para registrar las imágenes de campo claro -haz amarillo-: alojamiento de la lámpara de tungsteno, brazo de acople del condensador y la lámpara, prisma DIC y anillo de fase del condensador, bandeja de filtros, aperturas y sistemas de lentes del condensador. En lo que respecta a la fluorescencia *per se*, los cubreobjetos se colocan en una cámara de perfusión que, a su vez, se fija en la pletina del microscopio de fluorescencia invertido Zeiss Axiovert S100TV. Además, **(6.3b)** el microscopio cuenta con un sistema de perfusión (Warner instruments, Hamden CT), lo que permite estimular a las células con el medio deseado a una velocidad y temperatura constante (37 °C). **a)** En cuanto a la excitación de la sonda, las células se iluminan con una lámpara de Xenon que, con ayuda de una rueda de filtros de excitación (Sutter Instruments, Novato, CA), es capaz de alternar entre luz monocromática a 340 y a 380 nm que gracias a un espejo dicróico esta se desvía hacia la cámara que contiene el cubreobjetos con las células. La fluorescencia emitida por las células es recogida mediante un objetivo 40x (de inmersión en aceite y de apertura numérica 1.3) y esta atraviesa el espejo dicróico el cual refleja longitudes de onda superiores a 520 nm y las de menor longitud de onda que logran pasar son filtradas por un filtro de emisión a 520 nm. Finalmente, las fluorescencia que ha llegado hasta este punto es la imagen final la cual es capturada por una cámara Orca-ER de Hamamatsu Photonics (Hamamatsu, Japón) y procesada por un ordenador con un *software* capaz de procesar imágenes digitales y de analizarla (modificado de <https://histoptica.wordpress.com/microscopio-de-fluorescencia/> y de Queval et al. (2010)).

6.3.2. Procedimiento y diseño experimental para el análisis funcional del efecto del DFMO en los modelos celulares HT29 y NCM460

Tras realizar el tratamiento con DFMO tal y como se explicó en el apartado “6.2 Tratamiento de células NCM460 y HT29 con DFMO,” se procedió a realizar el experimento de imagen de Ca^{2+} . El fin de estos experimentos ha sido estudiar tanto el tamaño de los depósitos de Ca^{2+} como SOCE. En concreto, para estudiar el tamaño de los depósitos de Ca^{2+} se usa el inhibidor reversible de la SERCA ácido ciclopiazónico (CPA) 10 μM en medio libre de Ca^{2+} para conseguir el vaciado del RE y, a continuación, para revelar SOCE se perfunde Ca^{2+} 1 mM (figura 7.13a).

Para ello, las células sembradas en cubreobjetos se incuban en FURA2/AM 4 μM durante 45 min previo al inicio del experimento y, posteriormente, se montan sobre la cámara con perfusión a 37 °C con medio MEC. A continuación, se iluminan las células con la fuente de Xenon y, tras un breve periodo para que se establezca la señal, se comienza con la perfusión de CPA 10 μM en medio libre de Ca^{2+} el cual inhibe de forma reversible SERCA, lo que permite registrar el contenido de los depósitos de Ca^{2+} . A continuación, se perfunde con medio MEC con CPA 10 μM con el fin de observar y registrar SOCE. Los resultados fueron analizados utilizando el software Aquacosmos 2.6 (Hamamatsu Photonics, Hamamatsu, Japón). En cada uno de las unidades experimentales (cristales o cubreobjetos) se registraron alrededor de 30-40 células. Los resultados proceden de la medida de 4 experimentos independientes. En cuanto al diseño experimental, se tienen dos factores:

- *Línea Celular*: este factor se corresponde con las diferentes líneas celulares evaluadas, por lo que tiene dos niveles: *HT29* y *NCM460*.
- *Tratamiento*: este factor se corresponde con el tratamiento o no con DFMO, por lo que tiene dos niveles: *Control* y *DFMO* (tratadas con DFMO).

En consecuencia, combinando ambos factores se obtienen cuatro condiciones experimentales diferentes: *NCM460.Control*, *NCM460.DFMO*, *HT29.Control* y *HT29.DFMO*. Para evitar factores de confusión debidos al día de realización de cada réplica, en un mismo día se realiza una réplica (de un total de cuatro) de cada una de las cuatro condiciones experimentales. En cuanto a los factores aleatorios que pueden afectar al experimento y que hay que controlar, son:

- *Cristal*: correspondiente a las células registradas en un mismo cubreobjetos que, obviamente, serán todas de una misma condición experimental
- *Día*: correspondiente al día de procesamiento de la muestra que, como se ha dicho antes, en un mismo día se procesa una réplica de cada condición experimental de un total de cuatro réplicas para cada una.

Es importante recordar que de forma simultánea a la preparación de cada una de las muestras empleadas en el experimento de imagen de Ca^{2+} para estudiar el efecto del DFMO también se han preparado las muestras equivalentes para estudiar este mismo efecto a nivel transcriptómico a partir del ARN de estas últimas.

6.3.3. Análisis de los datos de experimentos de imagen de Ca^{2+}

A continuación, se muestra el protocolo seguido resumido del análisis de los experimentos de imagen. Además, en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos” se presenta un caso práctico de cómo analizar los datos e interpretar los resultados de un experimento de imagen de Ca^{2+} extrapolable tanto a las áreas bajo la curva como a los incrementos máximos y para cada uno de los estímulos. Para más detalle de algunos de los conceptos más generales relacionados con este tipo de análisis, se recomienda leer el apéndice “A Fundamentos básicos para el análisis de datos.”

De cada uno de los registros en célula única de los experimentos de imagen de Ca^{2+} se calcula el área y el incremento máximo correspondiente a cada estímulo. Posteriormente, cada uno de esos estímulos se considera una variable respuesta y se analizarán independientemente. Para contrastar las diferencias entre los tratamientos, es decir, entre las células HT29.Control, HT29.DFMO, NCM.Control y NCM.DFMO, se emplean modelos lineales mixtos. Así, se podrá considerar en el modelo los efectos aleatorios correspondientes al día de realización de los registros, de tal manera que en cada día se han realizado dos réplicas de cada una de las condiciones experimentales evaluadas (tratamientos), y el efecto del cristal; es decir, a la correlación que existe entre las células de un mismo cristal. Para seleccionar las variables del modelo se emplean los criterios cAIC, AIC y BIC. Posteriormente, se realizaron las comparaciones dos a dos empleando el método de Tukey.

Para evaluar los supuestos del modelo se realizó el diagnóstico de los residuos mediante métodos gráficos y basados en contrastes de hipótesis. Por un lado, se emplearon los gráficos Q-Q, histogramas de los residuos de Pearson junto con la estimación kernel de su densidad, gráficos de cajas y bigotes de los residuos de Pearson en función de las condiciones experimentales y diagramas de dispersión de los residuos de Pearson en función de los valores ajustados. Por otro lado, los test empleados para evaluar su normalidad fueron los de Shapiro-Wilks, Jarke Bera y Lilliefors, mientras que para su homocedasticidad se usó el método de Breush Pagan. En cuanto a los efectos aleatorios, también se evalúa su normalidad mediante gráficos Q-Q.

Por último, aparte de ajustar modelos lineales mixtos a los datos mediante métodos clásicos, también se han empleado métodos *bootstrapping* paramétricos para aquellos casos en los que se viole considerablemente alguno de los supuestos. No obstante, cabe decir que en ningún caso se ha violado ninguno de ellos de forma preocupante, como se verá al comparar los métodos clásicos con los *bootstrapping*, ya que apenas se observan diferencias entre ellos.

6.4. Extracción de ARN

A continuación, se muestran los protocolos seguidos para la extracción de ARN de los distintos modelos usados que, posteriormente, se usarán para experimentos transcriptómicos de alto rendimiento.

6.4.1. Extracción de ARN a partir de líneas celulares

La extracción de ARN a partir de líneas celulares se realiza con ayuda del kit de extracción *GeneMATRIX Universal RNA Purification Kit de EURx* y siguiendo exactamente las indicaciones de la casa comercial. Para ello, las células se siembran en placas de cultivo de 10 cm de diámetro. Al día siguiente se retira el medio de cultivo, se añade el tratamiento (medio con o sin tratamiento en función de si la muestra es o no control) y se incuban durante 96 h. Las células se levantan con tripsina-EDTA y, posteriormente, se sigue el protocolo indicado por el fabricante: i) se lisan las células en presencia de β -mercaptoetanol y tras la adición del tampón de lisis, las ARNs celulares se inactivan inmediatamente ii) a continuación, se transfiere el lisado celular a una columna de homogenización, en la cual quedará retenido el ADN genómico iii) posteriormente, se pasa el filtrado por otra columna la cual presenta una gran especificidad para el ARN, de tal forma que este queda adsorbido en la matriz de la columna iv) seguidamente, se eliminan los posibles contaminantes mediante sucesivos lavados con distintos buffers v) por último, se eluye el ARN con H_2O libre de ARNasas.

Tras extraer el ARN, es necesario determinar su calidad mediante el espectrofotómetro NanoDrop (Thermo Scientific, Waltham, MA) a través de la medida de la densidad óptica (DO) a 230, 260 y 280 nm. Además, mediante la medida de la DO a 260 nm se cuantifica la cantidad de ARN. En particular, para los microarrays empleados, Clariom D Human, los parámetros de calidad mínimos han de ser $A_{260}/280 \geq 1.8$, $A_{260}/230 \geq 1.5$ y entre 100 y 250 ng de ARN total (*Guía de Usuario de Microarray Affymetrix (EXPRESION)* (2021)). Además, también es importante comprobar los parámetros de integridad $28S/18S \geq 1.2$ y *número de integridad del ARN (RIN) ≥ 7* .

Este parámetro RIN mide la calidad del ARN mediante electroforesis capilar -Agilent Bioanalyzer...- en lugar del Nanodrop, puesto que este último es incapaz de distinguir entre ARN y el ADN contaminante y no permite estudiar la degradación de la muestra. En particular, con el equipo de Bioanalyzer, por ejemplo, se realiza la caracterización de las muestras en función de los patrones de movilidad electroforética se obtiene el RIN el cual ofrece un valor objetivo de la calidad del ARN basado en la cuantificación de los productos de degradación del ARN a lo largo de la migración electroforética que vendrá determinada por la conservación del tejido y del proceso de extracción. Este oscila entre 1 (ARN totalmente degradado) y 10 (ARN intacto), y suele considerarse aceptable por encima de 6-7 (figura 6.4). Así, cuando el ARN está degradado se observa una mancha a bajos pesos moleculares (figura 6.4d), mientras que si está intacto e íntegro se observan claramente los picos correspondientes a 18S y 28S con una relación 1:2 (figura 6.4a).

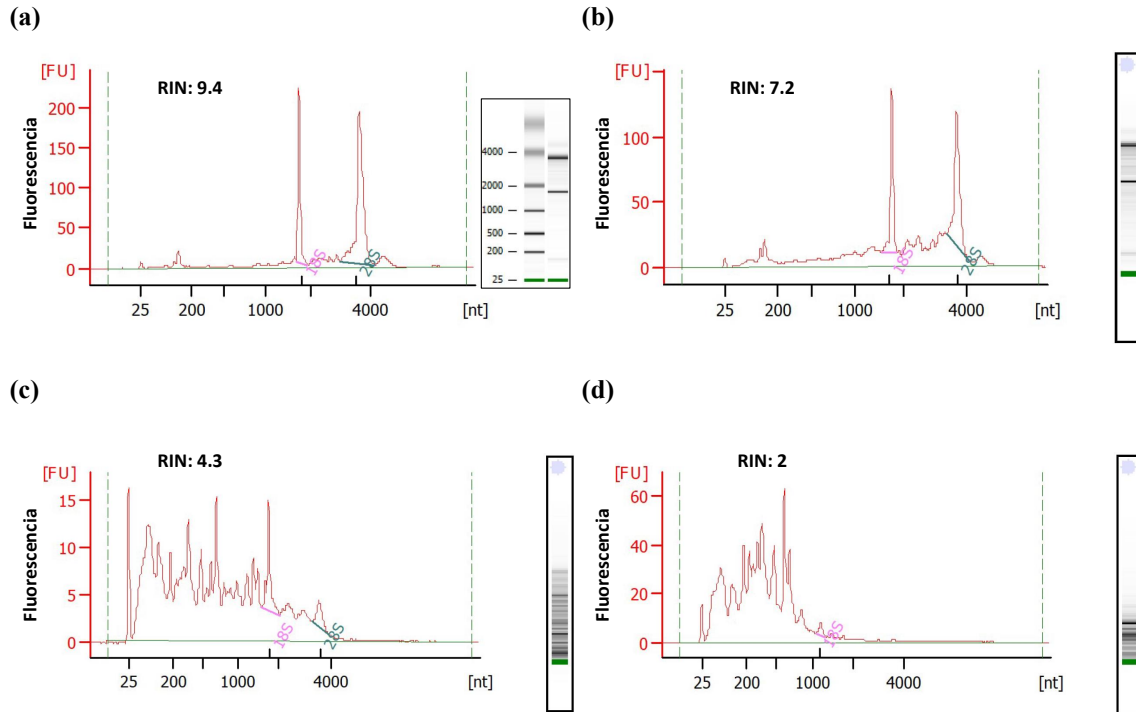


Figura 6.4: Control de calidad basado en el parámetro RIN. En los gráficos se muestra el control de calidad basado en el parámetro RIN y a la derecha de cada uno la migración electroforética correspondiente. Además, en la figura **a)** se muestra el patrón correspondiente a la migración electroforética. A grandes rasgos, uno de los puntos en los que fijarse es que la relación entre las bandas 28S:18S sea 2:1. Los gráficos de control de calidad mostrados en esta figura han sido extraídos de datos generados por el propio grupo de investigación.

6.4.2. Extracción de ARN a partir de tejido

La extracción de ARN de las muestras de tejido humano de cáncer de colon y tejido de colon normal se realiza de forma pareada; es decir, que se extrae a la vez el ARN de las muestras tanto normal como tumoral del mismo paciente. Esta extracción se realiza mediante el método fenol-cloroformo. Dado que las muestras de tejido tumoral están en criomoldes, embebidas en *OCT* (*Optimal Cutting Temperature Compound*), estas deben descongelarse y eliminar el OCT. Así, el protocolo seguido es el siguiente que, por ser muchos pasos, se enumeran de forma vertical:

1. En primer lugar, se prepara una solución de descongelación que consiste en PBS 1x e inhibidor de proteasas (X30 Pierce Protease Inhibitor Mini Tablet EDTA-free, Thermo scientific Pierce).
2. Se coloca la muestra sobre una placa de 10 cm y, con ayuda de una cuchilla, se perfila la muestra y, así, se elimina el OCT libre de ella.
3. A continuación, se coloca la muestra en un tubo eppendorf con la solución de descongelación, anteriormente preparada, a 4 °C.

4. Se lava el tejido dos veces más y se transfiere la muestra a otro tubo eppendorf con solución de descongelación fresca.
5. Una vez eliminado el OCT, se pesa el tejido en una balanza para determinar los volúmenes de reactivos a utilizar.
6. En una placa de 10 cm de diámetro se trocea el tejido y se transfiere a un tubo de homogenización (kit CK mix50), se añade TRizol a 4 °C (1 mL TRizol/100 mg de tejido) y se lleva el tubo con la muestra a un homogeneizador de tejido (PRECELLYS24, programa 6500 rpm, 3 ciclosx20s).
7. Se transfiere el sobrenadante a un tubo eppendorf nuevo y se introduce en un baño de ultrasonidos (2 ciclosx30s).
8. Después, se incuba la solución de TRizol 30 min a 4 °C y, a continuación, se centrifuga la muestra a 12000 g durante 5 min a 4 °C.
9. Se transfiere el sobrenadante en el cual se encuentra el ARN y las proteínas a un tubo eppendorf y se incuba 5 min a temperatura ambiente.
10. Se separa el ARN de las proteínas mediante la adición de cloroformo (200 µL cloroformo/1 mL TRizol) y se incuba 5 min.
11. Se centrifuga a 12000 g durante 15 min, lo que da lugar a tres fases: una fase acuosa en la parte superior que contiene el ARN, una fase orgánica fenol-cloroformo inferior que contiene las proteínas, y una interfase que contiene el ADN; así, se transfiere la fase acuosa con el ARN a un tubo nuevo, se precipita con isopropanol (500 µL isopropanol/1 mL TRizol) y se incuba 10 min a temperatura ambiente. En cuanto a la fase orgánica con proteínas y la interfase se reservan para aislar posteriormente las proteínas.
12. Se centrifuga a 12000 g durante 10 min a 4 °C y el precipitado blanco que se obtiene se lava con etanol al 76 % (1 mL etanol/1 mL TRizol).
13. Se centrifuga a 7500 g durante 5 min y, a continuación, se retira el sobrenadante y el pellet resultante se deja secar durante 5-10 min.
14. Por último, se solubiliza el ARN en H₂O DEPC (50 µL) y se incuba 10 min a 55-60 °C.

Al igual que sucedía en la extracción de ARN en líneas celulares, aquí, tras extraer el ARN, se determina su calidad mediante el espectrofotómetro NanoDrop (Thermo Scientific, Waltham, MA) a través de la medida de la DO a 230, 260 y 280 nm. Además, mediante la medida de la DO a 260 nm se cuantifica la cantidad de ARN. En particular, para los microarrays empleados, Clariom D Human, los parámetros de calidad mínimos han de ser $A_{260}/A_{280} \geq 1.8$, $A_{260}/A_{230} \geq 1.5$ y entre 100 y 250 ng de ARN total (*Guía de Usuario de Microarray Affymetrix (EXPRESION) (2021)*). Además, también es importante comprobar los parámetros de integridad $28S/18S \geq 1.2$ y $RIN \geq 7$.

En caso de que la calidad no sea aceptable -las muestras de tejido humano pareadas cumplieron todos los requisitos excepto el RIN, posiblemente a una degradación de la muestra mayor a la esperada- se recurre al kit *GeneChip® WT Pico Kit (WT Pico Kit)* de Affymetrix el cual, entre otras ventajas, resulta ser muy flexible para diferentes tipos de muestras -cultivos celulares, congeladas, frescas, embebidos en parafina, con elevada degradación...-, y permite preparar las muestras para microarrays de transcriptoma completo con tan solo 10 células o 100 pg de ARN.

6.5. Técnicas transcriptómicas de alto rendimiento

En este apartado se presentarán las técnicas de alto rendimiento comenzando con una breve introducción de los fundamentos teóricos de aquellas empleadas en la presente tesis. Posteriormente, se describirán los procedimientos y diseños experimentales seguidos. Por último, se explicará brevemente el análisis de datos realizado.

6.5.1. Fundamentos teóricos

El dogma central de la biología molecular afirma que un gen -porción de ADN que contiene información para que la célula genere un producto final- se transcribe a una secuencia de ARNm y este, a su vez, a una secuencia de proteína. Todo este proceso es sumamente complejo, pues el ARNm sufre un proceso de maduración hasta llegar a la macromolécula final capaz de ser traducida, proceso en el que se eliminan porciones de ARN no codificante (intrones), se unen aquellos que sí codifican para la proteína en cuestión (exones) y sufre una serie de modificaciones postranscripcionales. Es más, un mismo gen puede codificar para diferentes proteínas o isoformas de una proteína combinando de diferentes formas sus exones (*splicing alternativo*). Así, a una secuencia de ARNm generada a partir de un gen se le conoce como transcrito y al conjunto y cantidad de todas y cada una de ellas que posee una célula en un momento dado bajo unas condiciones fisiológicas determinadas se le conoce como transcriptoma (Z. Wang, Gerstein, and Snyder (2009)).

No obstante, durante este proceso pueden sucederse numerosos problemas, como por ejemplo la aparición de mutaciones que den lugar a procesos alternativos de maduración que conlleve que algún intrón no se elimine algún o que algún exón se elimine... Además, también es posible que no haya una clara correlación entre los niveles de expresión de ARNm con los de la proteína que codifica, lo que puede tener diferentes explicaciones: en condiciones normales, existen mecanismos de proteólisis y de modificaciones postraduccionales que degradan o modifican dichas proteínas; alguna mutación puede dar lugar a proteínas aberrantes que acaben por degradarse; moléculas de ARN más cortas, como micro-ARN, son capaces de regular la traducción de dichas proteínas...

Afortunadamente, a pesar de estos inconvenientes, el análisis de los niveles de expresión de diferentes genes puede proporcionar una información sumamente valiosa en el campo de la biomedicina, como por ejemplo en el estudio de enfermedades como el cáncer. Es más, a lo largo de la historia se han desarrollado diferentes tecnologías para medir estos nivel de expresión, desde técnicas más antiguas como *Southern blot* o *Northern blot*, que detectan la presencia de una secuencia de ADN o de ARN, respectivamente, a partir de una muestra compleja, a otras más recientes como *qPCR* capaz de cuantificar la expresión de diferentes genes, y, más recientemente, otras técnicas de alto rendimiento capaces de medir cientos o miles de variables a la vez e incluso el transcriptoma completo, como son los *microarrays* o *RNA-seq*. No obstante, el estudio de cientos y miles de genes de forma simultánea frente al estudio gen a gen tiene la desventaja de una menor resolución. En concreto, en la presente tesis se han empleado estas dos últimas y, por ello, a continuación se van a desarrollar brevemente cada una de ellas.

6.5.1.1. Microarrays de ARN

El microarray es un dispositivo con forma de *chip* que mide la expresión de cientos y miles de moléculas de interés el cual se basa en la hibridación de una sonda con una molécula diana. En particular, la expresión se mide con un escáner que registra los niveles de fluorescencia de cada sonda en función de su mayor o menor hibridación con la molécula diana. Existen diferentes tipos de microarrays en función de la molécula diana -ADN, ARN, proteínas...-; sin embargo, sólo se van a tratar aquellos capaces de medir los niveles de expresión de genes a nivel de ARNm, conocidos como *microarrays de ADN o de expresión*.

De forma general, el microarray de ADN consta de una base o soporte sólido (vidrio, silicona, plástico...) sobre el cual se imprimen, con un orden lógico y conocido, diferentes fragmentos de ADN, sondas, complementario a toda o parte de la secuencia de ADNc correspondiente a cada uno de los transcritos de interés, moléculas diana. Cabe destacar que la secuencia de cada sonda, la cual no siempre se conoce, se asocia con un sólo gen puesto que tiene elevada especificidad por este; sin embargo, en algún caso aislado alguna sonda puede hibridar con dos genes, como sucede con los genes TRPM2 y TRPC7 en algún microarray de Affymetrix. Como resultado, se obtiene un chip que recuerda a una matriz bidimensional de gran dimensionalidad en el que cada punto o *spot* de esta contiene millones de clones de una misma secuencia o sonda (Schena et al. (1995), Haben Fesseha (2020), Patirupanusara and Suthakorn (2012)).

6.5.1.1.1. Fabricación de los microarrays

El proceso de fabricación de microarrays de ADN es un proceso que engloba numerosos pasos: i) obtener de las secuencias de ADN del genoma del organismo de interés de bases de datos como CenBank ii) decidir el tipo de sondas, que pueden ser clones de ADNc, amplificaciones por PCR de oligonucleótidos iii) seleccionar el soporte, que ha de cumplir una serie de requisitos tales como estabilidad química o compatibilidad con los sistemas de detección, por lo que suelen usarse membranas de nylon, de nitrocelulosa, vidrio o plástico iv) el tipo de marcaje, siendo el más usado el fluorimétrico y, en concreto, los fluoróforos Cy3 -absorción a 550 nm y emisión a 570 nm- y Cy5 -600 nm y 670 nm- v) el tipo de unión sonda-soporte, que pueden ser físicas -tratamiento con un polímero, adsorción...- o químicas -unión electrostática, enlace covalente...- vi) la técnica para unir o inmovilizar de las sondas en la superficie del soporte (Schena et al. (1995), Miller and Tang (2009), Patirupanusara and Suthakorn (2012), Haben Fesseha (2020)).

En cuanto a las técnicas de fabricación e inmovilización de las sondas en el soporte, existen dos grandes metodologías: la *síntesis in situ* (figura 6.5) y *delivery* (figuras 6.6 y 6.7). La fabricación por *síntesis in situ* se caracteriza porque las diferentes sondas -oligonucleótidos- se sintetizan sobre el propio soporte en localizaciones conocidas (se sabe qué sonda hay en cada *spot*). Esta tecnología de fabricación, cuyo origen se remonta a la industria de los semiconductores, consiste en la adición cíclica de los cuatro nucleótidos sobre la superficie tratada con grupos protectores fotolábiles que

se activan selectivamente para la síntesis de ADN mediante ciclos de luz y oscuridad. Los soportes más comúnmente usados en fotolitografía son de plástico o de vidrio cuya superficie, vidrio o silica fusionada al soporte, es modificada covalentemente con un reactivo -silano- para generar en la superficie grupos aminos reactivos los cuales, a su vez, se modifican con los grupos protectores fotolábiles (se eliminan con la luz) nitro-piperonil-oxi-carbonilo (Schena et al. (1995), Miller and Tang (2009), Patirupanusara and Suthakorn (2012), Haben Fesseha (2020)).

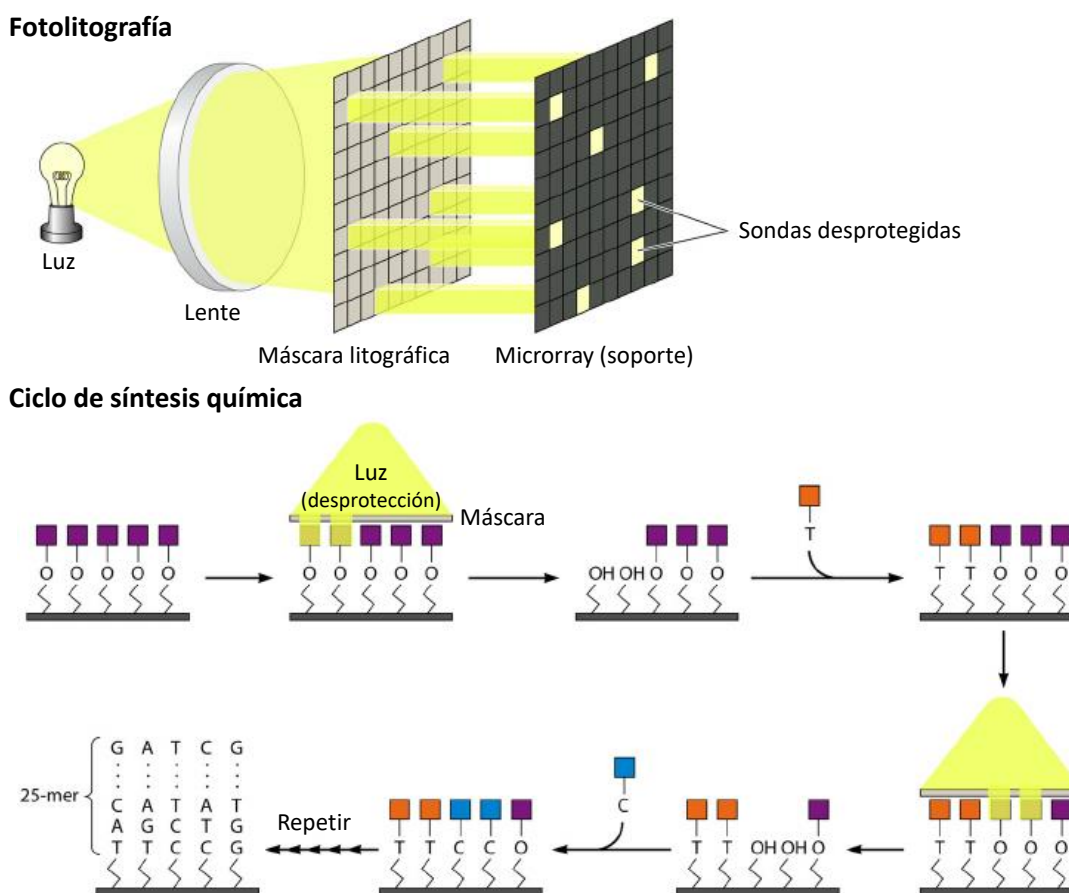


Figura 6.5: Proceso de fabricación de microarrays mediante la técnica de fotolitografía. En la parte superior se muestra la técnica de fotolitografía. Esta técnica consiste en hacer pasar una radiación UV a través de una máscara litográfica (podría pensarse como una placa perforada de los ordenadores antiguos) la cual permite que la radiación afecte únicamente a los grupos químicamente protegidos de aquellas posiciones de la superficie del microarray sobre las que sí ha de incidir, a la par que protege el resto de posiciones. Así, con diferentes máscaras es posible desproteger diferentes posiciones del microarray, posiciones sobre las cuales se realiza el proceso de síntesis. Por lo tanto, aplicando se una forma secuencial diferentes máscaras es posible controlar el orden de introducción de nucleótidos en las sondas sintetizadas y conocer la posición de estas dentro del microarray. En la parte inferior se muestra el ciclo de síntesis de las sondas aplicando la técnica de fotolitografía. En primer lugar, la radiación UV se hace incidir sobre la superficie de un microarray de tal manera que entre dicha superficie y la radiación hay una máscara litográfica, por lo que sólo ciertas localizaciones del microarray serán desprotegidas en esta ronda. A continuación, se añade un nucleótido protegido. Así, este proceso de desprotección-adición se repite el número de veces necesario hasta conseguir la longitud de las sondas deseada que, en Affymetrics es 25 pb. Obviamente, en cada ronda se aplica una máscara litográfica determinada (modificado de Miller and Tang (2009)).

En cada paso, tras la exposición a la luz de ciertas localizaciones de la superficie (determinadas por una máscara que deja o no pasar la luz a su través) sus grupos fotolábiles expuestos a la luz son eliminados y dichas localizaciones quedan activadas para reaccionar con un nucleósido de una base de nitrogenada concreta -la cual tiene un grupo fotosensible- y que se una empleando el método de síntesis de ADN con fosoramiditos de nucleósidos estándar. Posteriormente, se aplica una nueva máscara y se añade otro nucleósido con una base determinada y se repite este paso hasta alcanzar la longitud deseada de las sondas la cual suele oscilar entre 20-25 mer (bases) (figura 6.5). Cabe destacar que la eficacia de síntesis de estos oligonucleótidos oscila entre el 90 y 95 %; sin embargo, dada la elevada cantidad de sondas correctamente sintetizadas que contiene el microarray, el efecto en la hibridación es despreciable (Schena et al. (1995), Miller and Tang (2009), Patirupanusara and Suthakorn (2012), Haben Fesseha (2020)).

A diferencia de la anterior, en la tecnología *delivery* o *spotting* se fabrican las sondas, como pueden ser pequeños ADNc (300-3000 bases), que se amplifican con PCR y se purifican, y posteriormente se deposita la solución que las contiene sobre el soporte en localizaciones conocidas mediante un robot (Schena et al. (1995), Haben Fesseha (2020), Patirupanusara and Suthakorn (2012)). Existen diferentes versiones de esta metodología, pero las más comunes son la *mechanical microspotting* (figura 6.6) y la *inkjets* (figura 6.7). La *mechanical microspotting*, que es la más comúnmente usada en la fabricación de microarrays, emplea un robot de alta velocidad en cuyo extremo tiene una punta hueca que usa para depositar las sondas sobre la superficie sólida. Esta punta hueca entra en contacto con la solución que contiene las sondas para coger una pequeña cantidad de esta y posteriormente entra en contacto con la superficie del soporte para transferir parte de esta solución. Tras el primer ciclo, se lava la punta hueca y se carga una segunda muestra la cual se deposita en una localización adyacente a la anterior. Cabe destacar que los sistemas de control robótico y los cabezales de impresión multiplexados permiten la fabricación automatizada de microarrays (figura 6.6) (Schena et al. (1995), Haben Fesseha (2020), Patirupanusara and Suthakorn (2012)).

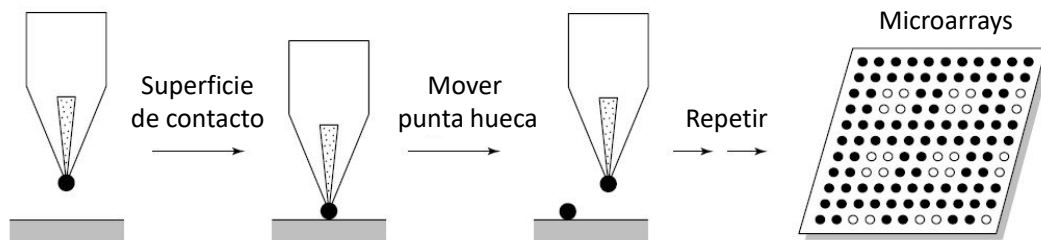


Figura 6.6: Proceso de fabricación de microarrays mediante la técnica *microspotting*. Una punta hueca robotizada entra en contacto con la solución que contiene las sondas y toma una pequeña cantidad para, posteriormente, depositarla por contacto sobre la superficie del soporte. Tras el primer ciclo, la punta hueca se lava para cargar una segunda muestra la cual se deposita en una posición adyacente a la anterior. Este proceso se repite hasta finalizar el microarray (modificado de Schena et al. (1995)).

La otra técnica *delivery* más usada es la *inkjets* (figura 6.7). En esta técnica un cabezal de impresión minúsculo acoplado con un sistema de propulsión (bomba piezoeléctrica o jeringa solenoide) toma parte de la muestra con las sondas sintetizadas y expulsa una gota sobre la superficie sólida del soporte sin entrar en contacto con este (mediante una corriente eléctrica en caso de que sea una bomba piezoeléctrica). Después del primer paso de inyección, el cabezal se lava y se carga una segunda muestra la cual se deposita en una posición adyacente a la anterior. Esto se repite hasta finalizar el microarray y, de hecho, una serie repetida de ciclos con varios cabezales permite una rápida producción de microarrays mediante esta técnica. Al igual que la técnica de *microspotting*, esta permite añadir sondas de casi cualquier tipo (ADNc, anticuerpos...) (Schena et al. (1995), Haben Fesseha (2020), Patirupanusara and Suthakorn (2012)).

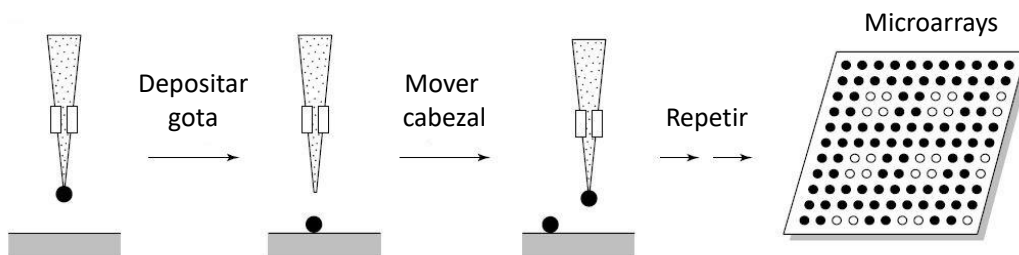


Figura 6.7: Proceso de fabricación de microarrays mediante *inkjetting*. Un cabezal de impresión minúsculo acoplado con un sistema de propulsión toma parte de la muestra con las sondas sintetizadas y expulsa una gota sobre el soporte y sin entrar en contacto con este. Posteriormente, el cabezal se lava y se carga una segunda muestra la cual se deposita en una posición adyacente a la anterior. Este proceso se repite hasta finalizar el microarray (modificado de Schena et al. (1995)).

Partiendo del mismo fundamento de la técnica anterior *injects*, otra opción que hace uso de esta, pero aplicada a la síntesis *in situ* en lugar de *delivery*, sería la empleada por Agilent, en la que un cabezal de impresión toma una pequeña cantidad de un nucleótido determinado (picolitros) y los añade sobre la primera capa de la superficie del microarray en las localizaciones deseadas. Así, este proceso se repite durante diferentes rondas añadiendo en cada una de ellas una base hasta alcanzar la longitud deseada de las sondas (60 mer).

6.5.1.1.2. Tipos de microarrays

Los microarrays también se clasifican en dos tipos según el número de muestras diferentes que soportan: *microarrays de dos colores* o “*spotted arrays*” y *microarrays de un color* o *arrays de oligonucleótidos* (figuras 6.8 y 6.8). Los *microarrays de dos colores* o “*spotted arrays*” soportan dos muestras, puesto que el ARNm de cada una de ellas, tras su extracción y transcripción reversa, el cADN resultante es etiquetado con marcador fluorescente diferente -normalmente Cy3 (verde) y Cy5 (rojo)-. De esta manera, es posible mezclar el ARNm etiquetado de ambas muestras para que hibriden de forma competitiva sobre el mismo microarray. Tras un periodo de incubación para favorecer la hibridación entre las secuencias de cADN etiquetado de las muestras y lavar el microarray para eliminar la muestra sobrante no hibridada, se ilumina en un escáner con un láser de tal manera que el ADNc etiquetado emitirá fluorescencia proporcional al ARNm en cada una de las muestras. El resultado es un valor de fluorescencia por sonda -*spot*- y muestra y, en consecuencia, medidas de expresión relativa; es decir, cuánto se expresa un gen en una muestra, rojo, respecto de la otra, verde (figura 6.8) (Schena et al. (1995), Schlingemann (2005), Mousseis (2011)).

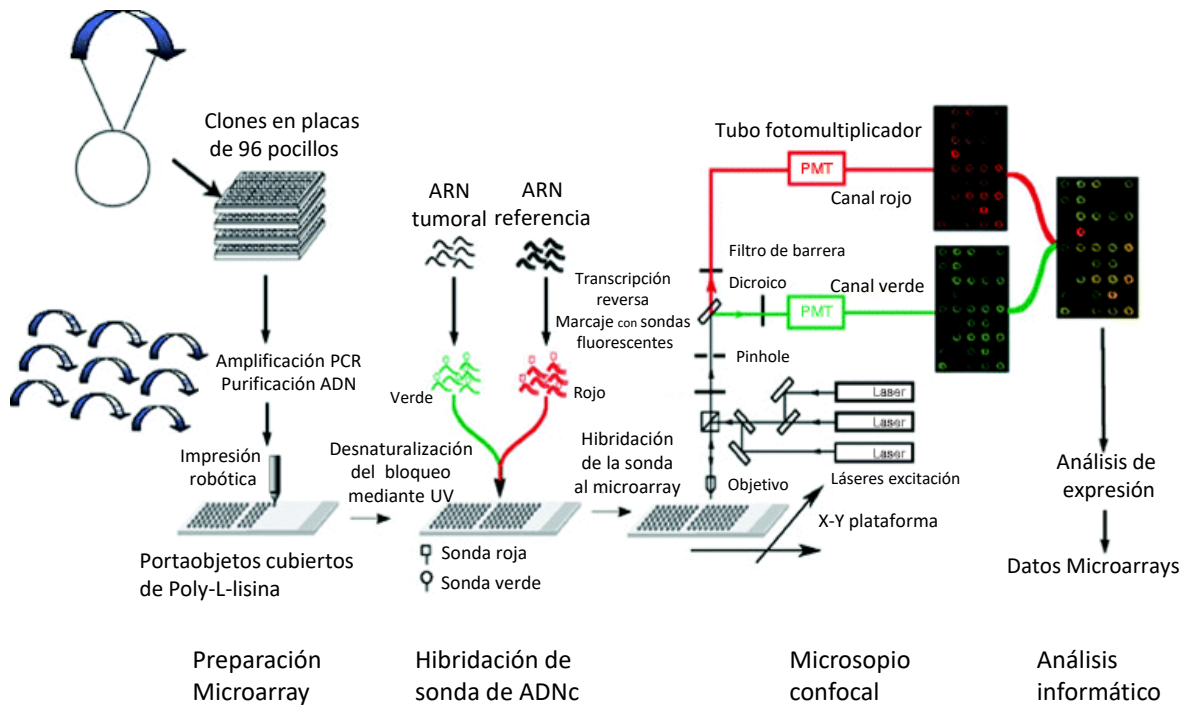


Figura 6.8: Microarrays de dos colores. En la figura se muestra un esquema del proceso de fabricación y realización de microarrays de dos colores. En primer lugar, se muestra el proceso de fabricación mediante la técnica de *microspotting*, en el que las sondas son amplificadas mediante PCR y se depositan en una superficie de vidrio. A continuación, se extrae el ARN de dos muestras diferentes, se genera su cADN de tal manera que el cADN de cada muestra está marcado con sondas diferentes: por lo general, una muestra con Cy3 (verde) y la otra con Cy5 (rojo)-. Posteriormente, este cADN etiquetado de ambas muestras, mezclado, se hibrida de forma competitiva con las sondas del microarray. Finalmente, tras finalizar el proceso de hibridación, se escanea el microarray y se registra la fluorescencia emitida tras su excitación con un láser con el fin de traducirlo a datos de expresión (modificado de Mousseis (2011)).

Por el contrario, los *microarrays de un color* o *arrays de oligonucleótidos* sólo admiten una muestra, puesto que todas las muestras se marcan con el mismo marcador fluorescente. Así, tras iluminar en un escáner el microarray con un láser, en cada *spot* se obtiene un valor de fluorescencia proporcional a la expresión del gen que representa pero, a diferencia de los microarrays de dos colores en los que al comparar las dos fluorescencias se tiene una escala de expresión relativa entre ambas muestras, en estos microarrays la escala de medida carece de sentido biológico y se denomina expresión absoluta (figura 6.9) (Schena et al. (1995), Staal et al. (2003), Schlingemann (2005), Daudén (2007)).

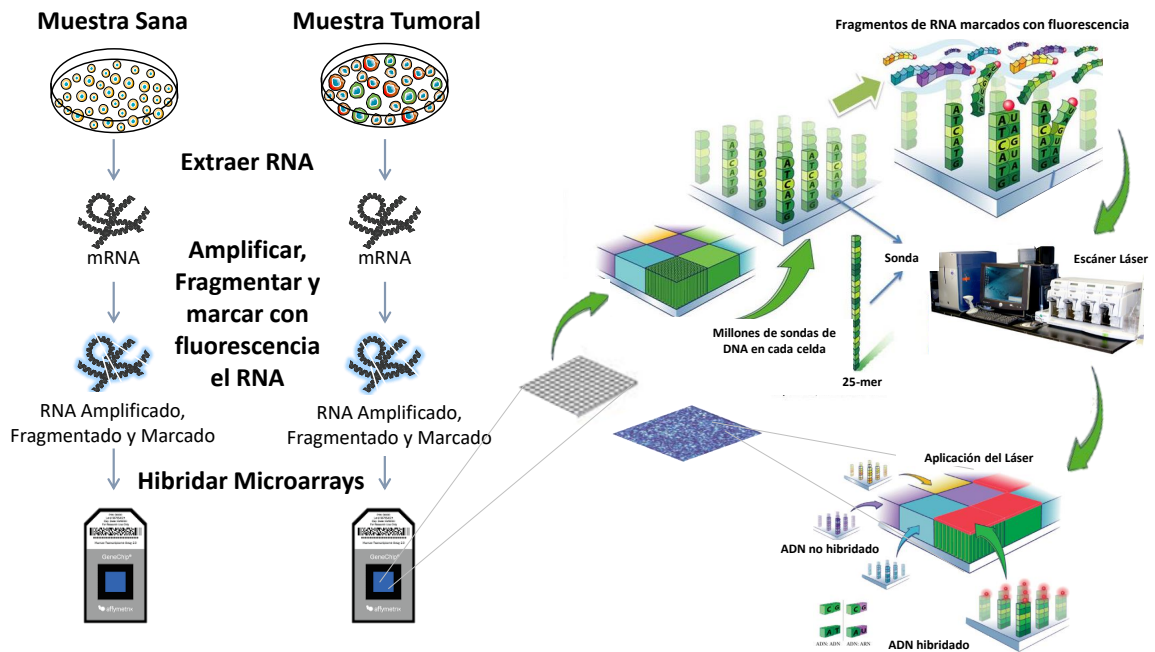


Figura 6.9: Microarrays de expresión de un color. Las fases de un experimento transcriptómico típico basado en microarrays de expresión de un color son las siguientes: i) se extrae el ARN de la muestra ii) se amplifica el ARN mediante PCR y este se fragmenta y se le añade biotina para favorecer la unión a los fluoróforos iii) se deposita el ARN fragmentado en el microarray iv) se deja entre 14-16 horas para permitir la hibridación de los fragmentos de ARN marcado con las sondas del microarray v) se aclara para eliminar el ARN que no ha hibridado vi) en un escáner se ilumina el microarray con un láser y se detecta la fluorescencia emitida que será proporcional a las sondas hibridadas en cada celda de sondas (modificado de Daudén (2007)).

6.5.1.1.3. Microarrays de expresión de Affymetrix

En la presente tesis se ha optado por los *Affymetrix GeneChip* o micorarrays de expresión de un color de Affymetrix puesto que es la empresa líder en el sector en este tipo de microarrays, y se va a explicar brevemente en qué consisten estos. Esta casa comercial sigue un método de fabricación de microarrays basado en *síntesis in situ* de oligonucleótidos de 25 bases (25 mer) mediante *fotolitografía* sobre una superficie de cuarzo. En la figura figura 6.10 se muestra un esquema de la estructura de un microarray de Affymetrix de 1.28 cm. Como puede observarse, la unidad más pequeña del microarray es la sonda o *probe* de longitud 25 mer, de tal manera que alrededor de 10 millones de *probes* con la misma secuencia se agrupan para formar celda de sondas o *probe cell* ($50 \mu\text{m}^2$) (Staal et al. (2003), Schlingemann (2005), Daudén (2007)).

En plataformas de Affymetrix más antiguas las *probe cells* forman parejas (*probe pairs*) donde las sondas de un miembro de la pareja presenta una coincidencia perfecta con el gen que hibrida -*perfect match* (PM)- y las sondas del otro miembro son idénticas a la secuencia que hibridan salvo en un nucleótido central -*mismatch* (MM)-. No obstante, en los *microarrays* más modernos MM ya no se incluye puesto que fueron diseñados en un intento fallido de eliminar el ruido. Cabe destacar que a cada gen no le corresponde un único *probe cell*, sino que suele haber entre 9 y 22 que que abarcan regiones diferentes del gen diana. Así, todos los *probe cell* que tienen el mismo gen como diana forman un *probe set*, los cuales se distribuyen a lo largo del microarray y, por ello, de la combinación de los resultados de las *probe cell* de un mismo *probe set* se obtiene la expresión del gen diana (Staal et al. (2003), Schlingemann (2005), Daudén (2007)).

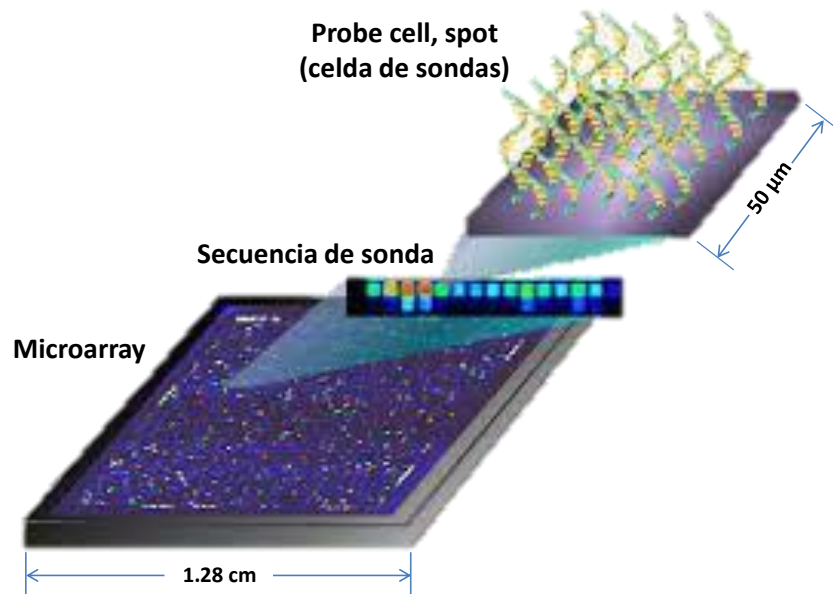


Figura 6.10: Microarrays de un color de Affymetrix. En la figura se muestra un esquema de la estructura típica de un microarray de un color de Affymetrix (modificado de González Maestre (2010)).

La tecnología de microarrays de Affymetrix ofrece cuatro generaciones de microarrays de expresión (figura 6.11):

- *3'IVT*: son los microarrays más clásicos los cuales tienen, de media, 11 sondas de cada transcrito complementarias con su región 3' desde la cadena de PolyA y presenta *bias 3'*. Se recomienda para el análisis de expresión de secuencias predichas y bien anotadas. Ejemplos de este tipo son el ampliamente usado *Human plus 133 2.0* o el *Prime View*.
- *Gene arrays*: estos microarrays tienen 26 sondas por gen y, a diferencia del 3', repartidas a lo largo del gen de forma uniforme, por lo que pueden emplearse para el estudio de la expresión del transcriptoma completo incluyendo transcritos sin extremos PolyA -lncARN, lincARN...-. Ejemplos de este tipo son el *Gene 1.0 ST Array* o la versión mejorada *Gene 2.0 ST Array*.
- *Exon arrays*: similar a gene array pero contienen 4 sondas por exón cuyas sondas han sido diseñadas a partir de un muestreo del gen entero. Estos permiten estudiar la expresión del genoma completo y variantes de *splicing*. Ejemplos de este es el *Exon 1.0 ST Array*.
- *HTA*: estos contienen una media de 10 sondas por exón y de 4 sondas para cada unión exón-exón. Permiten estudiar tanto el genoma completo como variantes de *splicing* (con mayor precisión que con los *exon arrays*). Ejemplos de este es el *Clariom D Human*.

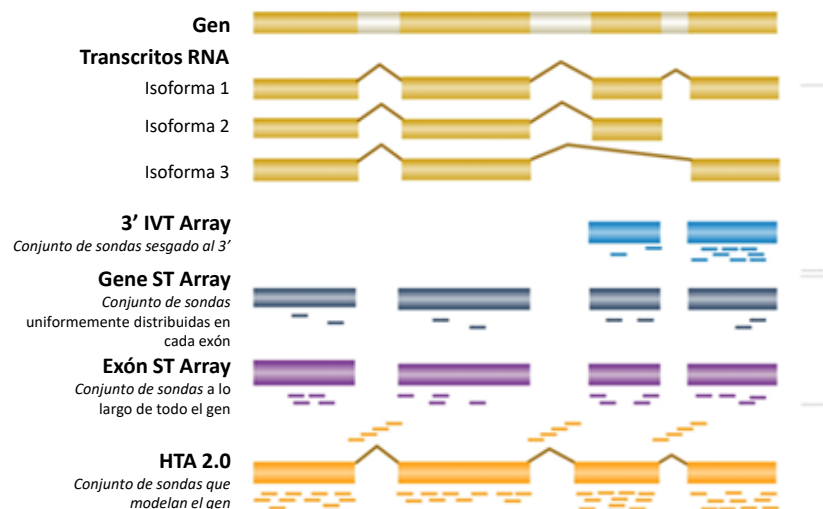


Figura 6.11: Microarrays de expresión de Affymetrix. En la figura se muestra un esquema de los diferentes tipos de microarrays de Affymetrix más comunes y las diferencias entre ellos en función de las posiciones en las que hibridan los diferentes conjuntos de genes (modificado de <https://www.cabimer.es>)

En la presente tesis se ha escogido el *Clariom D Human*, el cual contiene más de 6 765 500 sondas y proporciona información para más de 134700 genes -anotados hasta abril de 2016-, más de 542 500 transcritos y 948 300 exones. El ARN extraído se envió al Centro Andaluz de Biología Molecular y Medicina Regenerativa (CABIMER) para la realización de los microarrays.

6.5.1.2. RNA-seq

El término *RNA-seq* engloba diferentes métodos experimentales y computacionales desarrollados para identificar tanto qué transcritos hay como la abundancia de cada uno de ellos. Esta técnica ha cambiado considerablemente la forma de abordar muchas de las preguntas que se plantean en los diferentes grupos de investigación y en la clínica relacionadas con el transcriptoma, puesto que ha supuesto un cambio de paradigma el hecho de pasar de plantear experimentos para estudiar un solo transcrito a experimentos en los que se estudie la expresión de decenas de miles de genes simultáneamente en cada una de las muestras.

La secuenciación comenzó en 1977 con la tecnología desarrollada por el dos veces Premio Nobel Sanger, en la que a partir de una mezcla de nucleótidos de dedioxinucleótidos (ddNTP) se produce la síntesis de ADN a partir de una hebra molde hasta que en dicha síntesis entra el ddNTP complementario que por carecer de grupo hidroxilo en su posición 3' se detiene la reacción y es a partir de esta interrupción de la síntesis que puede estudiarse la secuencia.

Esta técnica ha estado a la cabeza de la secuenciación del ADN durante 30 años, hasta 2007, un periodo demasiado extenso sobretodo en un campo tan en constante evolución como la biología molecular, por lo que es evidente tanto su robustez como la complejidad en introducir mejoras en este campo. Obviamente, en esos 30 años han surgido mejoras en la tecnología, desde geles electroforéticos hasta secuenciadores capilares que permiten realizar 96 reacciones simultáneas -una por capilar- obteniendo cada uno una secuencia de entre 600 a 1000 bases, lo que se traduce en 100000 bases de secuencia en cuestión de dos horas.

No obstante, desde el año 2001, cuando secuenciar un genoma costaba 100 millones de dólares, esta técnica se ha ido abaratando considerablemente siguiendo la *ley de Moore* hasta que en el año 2007 se produjo un abaratamiento superior al esperable (“National Human Genome Research Institute (NHGRI) Web Site Genome.gov Is in the Public Domain.” (2014), figura 6.12). Este hecho puede explicarse por la aparición de las técnicas de secuenciación de segunda generación, que son las primeras desarrolladas de las conocidas como tecnologías de secuenciación de siguiente generación o *Next-Generation-Sequencing (NGS)*. En particular, en estas NGS de segunda generación se encuentran 454 de Roche (454 en honor a la calle en la que estaban) y IonTorrent (tecnología similar a 454 Roche), Illumina (que inicialmente se denominó Solexa), SOLID Life/APG (que durante un tiempo fueron los líderes en este campo).

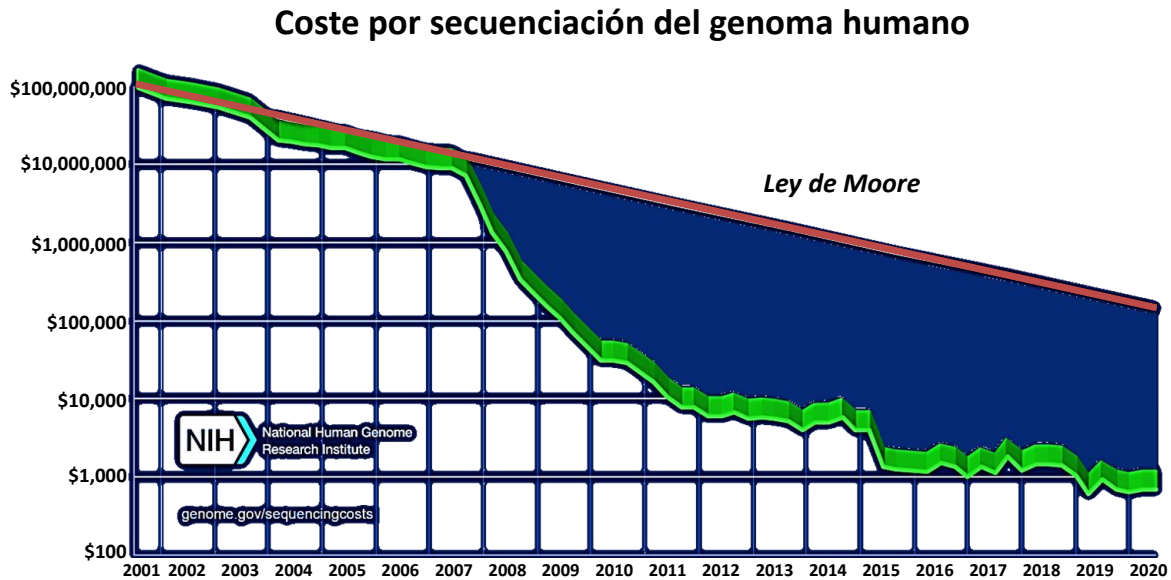


Figura 6.12: Coste por secuenciación del genoma humano. En el gráfico se muestra la reducción del coste por secuenciar un genoma desde el 2001 y cómo este disminuye considerablemente (curva verde), pues se desvía de la Ley de Moore (línea roja), a lo largo del 2007 consecuencia de la revolución en las técnicas de secuenciación, conocidas como tecnologías de secuenciación de siguiente generación o *next-generation DNA sequencing* (modificado de *National Human Genome Research Institute (NHGRI) Web Site Genome.gov Is in the Public Domain. (2014)*)

Estas tecnologías también se basan en la síntesis química de ADN a partir de nucleótidos individuales similar a Sanger, salvo por el hecho de que en estas la síntesis se miniaturiza (más que los capilares usados en Sanger), se realiza de forma masiva en un formato paralelo alcanzando millones de reacciones de secuenciación en un único ensayo (frente a 96 que realiza la secuenciación Sanger capilar) y se automatiza. Por ejemplo, si en un experimento de NGS habitual es posible realizar 6000 millones de reacciones de secuenciación, cada una de ellas de una longitud de 100 nucleótidos, es posible obtener $6 \cdot 10^{11}$ de bases de secuencia. No obstante, ha surgido también la secuenciación de tercera generación, también considerada una tecnología NGS, en la que a diferencia de la de segunda generación, en esta los moldes son moléculas individuales de ADN o ARN. Además, por un lado, tienen un menor número de reacciones por ensayo (pocos millones) pero, por otro lado, la longitud de secuencia alcanzada por reacción es unos 1500 nucleótidos.

De todas las tecnologías existentes, en la presente tesis se han empleado dos tecnologías de segunda generación: Illumina y IonTorrent. Por ello, se van a explicar brevemente cada una de ellas. En primer lugar, se pueden distinguir cuatro pasos: aislamiento del ARN y control de calidad (similar a microarrays), preparación de genotecas, secuenciación y análisis bioinformático. A continuación, se desarrollan brevemente cada una de ellas.

6.5.1.2.1. Preparación de las genotecas

Previamente a la secuenciación, el ARN extraído ha de convertirse en ADNc por varios motivos: su mayor estabilidad o es más adecuado para las reacciones de secuenciación. No obstante, el ADNc debe representar la verdadera cantidad de cada uno de los transcritos de la muestra correspondiente. Para la preparación de genotecas hay dos procedimientos generales para aislar el ARNm a partir del ARN total extraído (la cantidad depende de la plataforma): por un lado, seleccionar la cola de PoliA del ARNm y, por otro lado, eliminar el ARNr del ARN total extraído. Sin embargo, cada plataforma sigue su propio protocolo de preparación de genotecas y, por ello, existen kits comerciales correspondientes a cada una de ellas. Es más, estas genotecas pueden ser direccionales (*stranded*) o no (*unstranded*); es decir, que el protocolo empleado permita crear una genoteca que diferencie entre la cadena molde y la codificante o, por el contrario, que la genoteca sea bicatenaria y no diferencie entre ellas.

De forma muy general, tras haber aislado el ARNm, este se fragmenta -pueden emplearse métodos físicos como los ultrasonidos, incubando con reactivos de fragmentación...- y se sintetiza el ADNc para crear la genoteca. A continuación, se añaden *primers* hexaméricos aleatorios y mediante la acción de una transcriptasa reversa se sintetizan los ADNc de cada fragmento y, posteriormente, a partir del ADNc, se sintetiza la hebra complementaria a dicho ADNc y se elimina el ARN, por lo que se obtiene ADNc bicatenario. Posteriormente, se purifica el ADNc -por ejemplo, con esferas de Inmovilización Reversible de Fase Sólida (SPRI)-. Finalmente, se enriquece la librería mediante PCR.

6.5.1.2.2. Amplificación

El siguiente paso es la amplificación del ADNc de la genoteca. Este paso varía en función de las diferentes tecnologías. En particular, tanto para IonTorrent como para Illumina, se emplea el método de *moldes amplificados clonalmente*, donde cada una de las dos tecnologías tiene su propia estrategia (figura 6.13):

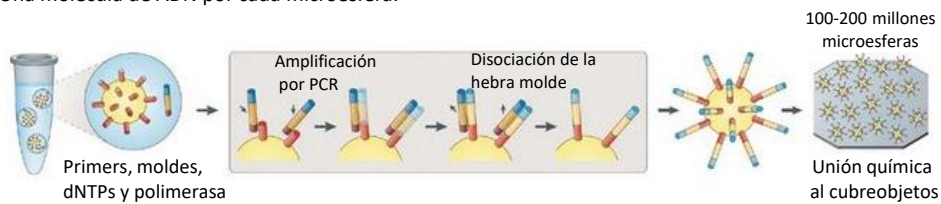
- IonTorrent emplea la técnica de PCR en emulsión: desarrollada por la tecnología 454 Roche -adaptación de otra ya existente llamada pirosecuenciación-. Para la amplificación, a las terminaciones de los ADNc de la genoteca se les ligan adaptadores que contienen regiones *primers* universales. Posteriormente, estos fragmentos ligados a adaptadores (separados en una sola hebra en caso de ARN) entran en contacto con unas microesferas las cuales poseen pequeños adaptadores que son complementarios a los adaptadores ligados a estos, de tal manera que la concentración de las microesferas y las condiciones sean tales que se favorezca que haya un fragmento por microesfera. En particular, se realiza una emulsión de manera que se obtiene una microesfera en la que alrededor de ella hay un adaptador y una cierta cantidad de reactivos necesarios para realizar la amplificación por PCR.

Tras la amplificación y creación de un nuevo molde que queda unido a la microesfera, el fragmento usado como molde en la anterior amplificación se disocia de la microesfera. A continuación, estas microesferas (de 100 a 200 millones con miles de moldes) se pueden inmovilizar en un gel de poliacrilamida sobre un portaobjetos de microscopio estándar, unidos químicamente sobre una superficie de vidrio cubierta con grupos amino, depositados en pocillos individuales en *PicoTiterPlate* o *PTP* -similar a un pocillo de 96 pero con más pocillos y muy pequeños en los que sólo cabe una microesfera- (figura 6.13a).

- Amplificación en fase sólida -es la que usa Illumina y es la más eficiente-. En un soporte físico, similar a un portaobjetos, hay una serie de *primers* directos e inversos de alta densidad que se han sintetizado por fotopirólisis (igual que en los microarrays) que están unidos covalentemente a él los cuales son complementarios a los adaptadores que se van a unir a los fragmentos. En cuanto a los fragmentos, ADNc, se les liga adaptadores a ambos extremos, se añaden al soporte sólido de tal manera que se consiga una determinada concentración que asegure que en cada microespacio del soporte haya uno de los fragmentos añadidos y que cada uno de ellos quede espacialmente separado del resto. Esta amplificación por PCR crea varias copias idénticas de cada secuencia, de tal manera que cada conjunto de secuencias producidas a partir de un mismo fragmento se le denomina clúster, cada uno de los cuales contiene alrededor de 1 millón de copias. Así, el ratio entre los *primers* y los fragmentos de ADNc sobre el soporte determinan la densidad superficial de los clúster amplificados, por lo que hay que tener especial cuidado en el control de la carga de la muestra y evitar sobrecargas que puedan dar lugar a clústers superpuestos. La amplificación que tiene lugar se conoce como amplificación puente debido a la forma en la que se elonga la secuencia, pues la hebra molde inmovilizada en el soporte se hibrida, a su vez, con el primer inmediatamente adyacente complementario al adaptador del extremo libre. De esta forma, esta técnica es capaz de producir entre 100 y 200 millones de hebras molde espacialmente separadas en los que los extremos quedan libres para facilitar la unión de *primers* universales de secuenciación para iniciar la reacción NGS (figura 6.13b).

(a) IonTorrent (PCR en emulsión)

Amplificación clonal: miles de copias en cada microreactor en emulsión.
Una molécula de ADN por cada microesfera.



(b) Illumina (amplificación en fase sólida)

Amplificación en fase sólida

Una molécula de ADN por clúster

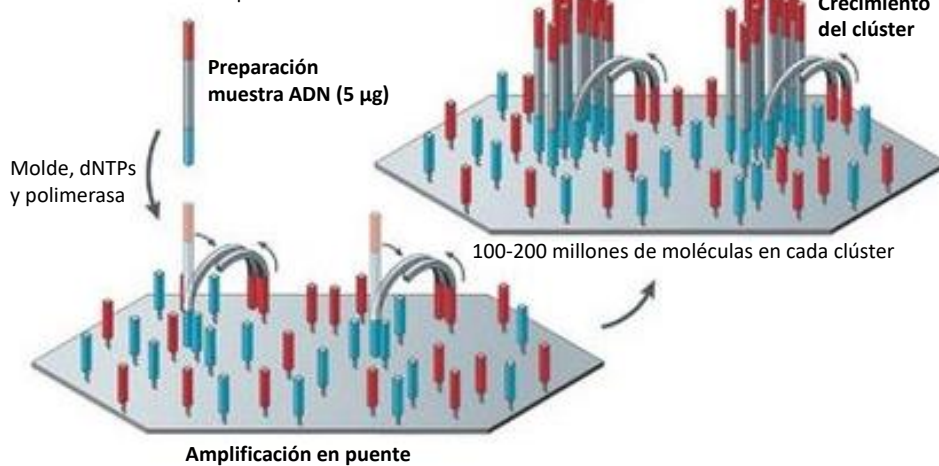


Figura 6.13: Estrategias de inmovilización y amplificación de los fragmentos. En la figura se muestra un esquema de las estrategias de las técnicas de inmovilización y amplificación de los fragmentos de cADN para las tecnologías **a) IonTorrent**, PCR en emulsión, e **b) Illumina**, amplificación en fase sólida. La **a) PCR en emulsión** consiste en crear una mezcla de reacción en la que una emulsión acuosa-lipídica encapsula complejos esfera-ADNc en gotas acuosas individuales, de tal manera que se realiza una amplificación por PCR en cada una de dichas gotas. De esta manera, en cada una de las esferas hay miles de copias de la misma secuencia molde. Por último, estas esferas pueden ser inmovilizadas sobre una superficie de vidrio mediante reacción química o depositadas en un *PicoTiterPlate* (PTP). En cuanto a la técnica de **inmovilización en fase sólida (b)**, se pueden distinguir dos pasos: cebado inicial con adaptadores e inmovilización de una única hebra molde monocatenaria en cada clúster, y amplificación en puente de los moldes inmovilizados con el primer inmediatamente adyacente para formar clústers (Modificado de Metzker (2010)).

6.5.1.2.3. Secuenciación

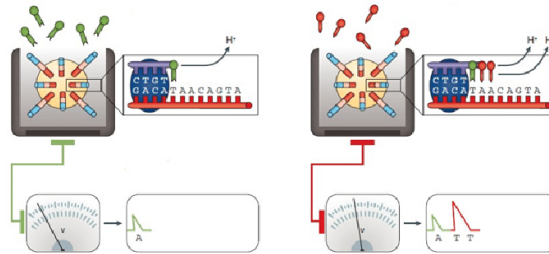
Para la secuenciación en NGS hay dos opciones principales, por síntesis o por ligación y, en particular, las tecnologías de *IonTorrent* e *Illumina* son de síntesis. A continuación, se explican brevemente las principales características y peculiaridades de ambas tecnologías (figura 6.14):

- **IonTorrent:** tras la amplificación mediante *emulsión de PCR*, es posible depositar las 100-200 millones de esferas sobre PTP de tal manera que en cada uno de los pocillos será posible realizar las reacciones de secuenciación y medir la dinámica de dicha amplificación en la que se registra qué nucleótido se incorpora en cada una de las microesferas y, por ende, se conocerá la secuencia de cada uno de esos fragmentos de ADNc de la genoteca. En particular, Life Technologies comercializó la tecnología de secuenciación por semiconductor de IonTorrent en 2010, similar a la de pirosecuenciación de 454 Rocher y, a diferencia de otras de segunda generación, no usa nucleótidos marcados con fluoróforos. En concreto, usa un chip que contiene un conjunto de micropocillos en cada uno de los cuales hay una gota con varios fragmentos idénticos en los que se va a llevar a cabo la secuenciación. En la secuenciación, tras la unión de un dNTP se va a formar un enlace fosfodiéster y se va liberar un protón, y la diferencia de potencial o de pH generada por el protón o protones liberados en el caso de que haya varios nucleótidos iguales seguidos en la secuencia, lo que se conoce como homopolímeros. Este cambio de pH es registrado por un microelectrodo sensor (CMOS-ISFET) localizado en el fondo del micropozo sensor que mide la cantidad de protones liberados o el cambio de pH que es proporcional al número de nucleótidos incorporados. Estos secuenciadores son capaces de producir lecturas de entre 200 a 600 pb (figura 6.14a).

Para controlar esta secuenciación lo que se hace es añadir, de forma cíclica, los cuatro nucleótidos, por ejemplo: primero adenina y donde se libere luz quiere decir que en la cadena complementaria (molde) hay una timina; se limpia y se añade timina; se limpia y se añade cisteína; se limpia y, por último, se añade guanina de tal manera que el ciclo se repite sucesivamente. Así, puesto que en cada paso del ciclo sólo se añade un dNTP, en el caso de que en la cadena molde haya más de uno complementario seguido, se añaden todos ellos. No obstante, cuando hay regiones que tienen muchos nucleótidos iguales seguidos (más de 6) no es capaz de discernir si se han añadido, por ejemplo, 12 o 20, y falla en estas regiones puesto que da lugar a un error indel con una tasa del 1 %, lo que supone la gran desventaja de esta técnica (Metzker (2010)). En contraposición, como ventaja sobre otras tecnologías NGS es que la mayor longitud de sus lecturas y el corto tiempo que tarda en llevar a cabo la secuenciación, entre 2 y 8 horas (Kchouk, Gibrat, and Elloumi (2017), Metzker (2010)).

- **Illumina:** para determinar la secuencia de nucleótidos de cada fragmento de ADNc amplificado, emplea la secuenciación por síntesis con terminadores reversibles de tal manera que en una misma mezcla de reacción se añaden los cuatro dNTP modificados con un fluoróforo unido (diferente para cada nucleótido), los *primers* y la ADN polimerasa. Así, estos *primers* se unen a las secuencias de cada clúster o *polonies* y, a continuación, la ADN polimerasa elonga la hebra de ADN a partir de los *primers* usando los dNTP modificados. Estos dNTP modificados tienen su -OH 3' inactivado para asegurar que sólo un nucleótido se añade cada vez. Así, un láser excita a los clúster para registrar el color del dNTP incorporado y será detectado por una cámara con dispositivo de carga acoplada (CCD) y, a través de un equipo informático, la señal registrada se traducirá a las correspondientes secuencias. Posteriormente, el terminador con el marcaje fluorescente se elimina, se regenera el -OH 3' y comienza de nuevo otro ciclo. Posteriormente se libera la molécula fluorescente, se regenera el grupo -OH y se realiza nuevamente el ciclo. Cabe destacar que esta es la tecnología más utilizada y que se han alcanzado lecturas de alrededor de 125 pb y un total de 600 Gpb (figura 6.14b). No obstante, está el problema de que la sobrecarga de muestra puede dar lugar a clústers superpuestos que disminuye considerablemente la calidad de la secuenciación. Además, la tasa de error de secuenciación es de 1 % (similar a IonTorrent) que, por lo general, se debe a la sustitución de los dNTP y a la mala identificación del nucleótido incorporado (Kchouk, Gibrat, and Elloumi (2017), Metzker (2010)).

(a) IonTorrent (secuenciación con semiconductor)



(b) Illumina (Secuenciación por 4 colores)

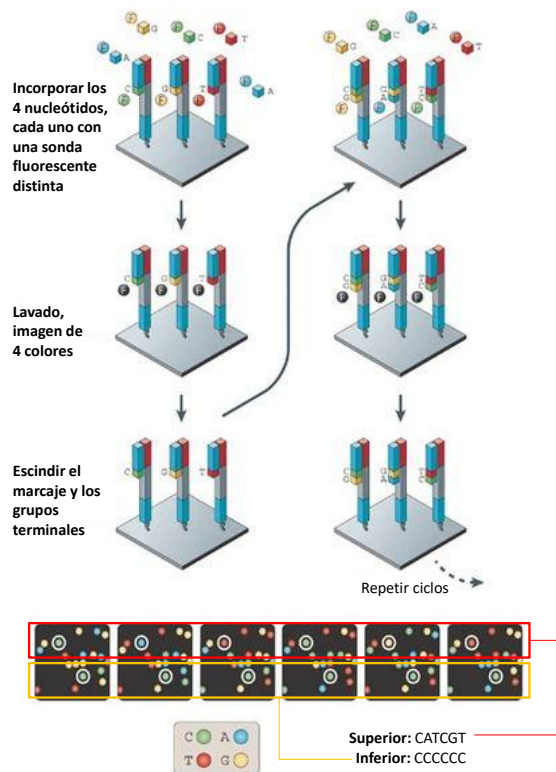


Figura 6.14: Técnicas de secuenciación. La **a) secuenciación con semiconductor** se basa en realizar varios ciclos de 4 rondas en cada una de las cuales se emplea un único nucleótido. Así, puesto que tras la amplificación e inmovilización por PCR en emulsión se tienen diferentes micropocillos en cada uno de los cuales hay un chip y una gota con varios fragmentos de cADN idénticos, en cada ronda se añade a todos los pocillos un único dNTP y en función de la secuencia de cADN de cada micropocillo, se unirá o no el dNTP añadido, de tal manera que se registrará con un microelectrodo un cambio de pH consecuencia del protón/es liberado/s del enlace/s fosfodiéster, por lo que la cantidad de protones liberados será proporcional a la cantidad de dNTP iguales añadidos en una sola ronda. La **b) secuenciación por cuatro colores** consiste en usar una mezcla de los cuatro cNTP modificados en su extremo 3'-OH y, además, cada uno de ellos tiene un fluoróforo diferente unido. En esta mezcla también hay presente *primers* ADN polimeras, con el fin de que dichos *primers* se unan a las hebras amplificadas de los clústers y, a continuación, la ADN polimerasa elonga la hebra a partir del primer en función de la hebra molde. La modificación de los dNTP hace que sólo se pueda unir un único dNTP en cada ronda, por lo que tras la unión de uno de ellos un láser lo excita y se registra la fluorescencia. A continuación, se regenera el 3'-OH y se realiza otra ronda. Así, tras todas las rondas, se conocerá la secuencia de cada clúster, pues en función de la longitud de onda registrada en cada ronda y de la posición se sabrá la secuencia de la cadena que ha crecido en cada clúster (Modificado de Kchouk, Gibrat, and Elloumi (2017) y Metzker (2010)).

6.5.1.2.4. Ventajas, desventajas y aplicaciones de RNA-seq

Tras lo anterior uno puede pensar que el tamaño de las lecturas es pequeño, pero lo que se consigue con las tecnologías NGS son muchos millones de pequeñas secuencias de ADN (o ADNc), de tal manera que al unir las adecuadamente es posible obtener la secuencia de un genoma o del transcriptoma completo de una célula. Esto se debe a que se tienen tantos millones de lecturas que se asegura que, como mínimo, cada una de las bases de la genoteca se vaya a leer una vez (profundidad o redundancia). Puede verse como que de forma individual se tiene poca capacidad de lectura (100 bases) pero de forma global se tiene una gran capacidad de hacer lecturas de forma simultánea. Por ejemplo, en el contexto del estudio del transcriptoma, si el genoma de un mamífero tiene unos 3000 millones de dNTP de los cuales el 1/30 se corresponde con los 20000 genes codificantes, si se realizan 1 millón de lecturas de 100 pb de bases se tendrá una redundancia de $1 \times$; es decir, que cada base debería de leerse una vez. Ahora bien, si en lugar de un millón se tienen 30 millones (que es lo más común en un experimento de IonTorrent o Illumina), entonces la redundancia será del $30 \times$ y, por ende, se espera tener una buena cobertura del transcriptoma completo.

La principal ventaja que tiene esta tecnología frente a otras más clásicas es su alto rendimiento, la sensibilidad y la capacidad de descubrir nuevos transcritos, modelos de genes y ARN no codificante. Además, frente a los microarrays tiene la ventaja de que no necesita conocer previamente los transcritos; es más, puede identificar nuevos transcritos. No obstante, también hay que destacar el gran volumen de datos que se generan (300 Gb por genoma) y que hay que procesar, lo que hace necesario el uso de supercomputación en muchas ocasiones. Sin embargo, en la presente tesis se ha partido de la matriz de expresión de los experimentos de *RNA-seq* y, por lo tanto, no ha sido necesario realizar el preprocesado de los datos. Por lo tanto, dicho preprocesado no se explicará dada la gran extensión que ello conllevaría.

En cuanto a sus aplicaciones, a pesar de que en la presente tesis la tecnología *RNA-seq* se emplea para el estudio de la expresión del transcriptoma completo, existen muchas otras, tales como la secuenciación y ensamblado de genomas *de novo*, análisis de la actividad funcional del genoma y epigenética, análisis de la cromatina, estudios de metagenómica...¹

6.5.2. Procedimientos y diseños experimentales

A continuación, se describen los diferentes experimentos realizados en la presente tesis basados en técnicas transcriptómicas.

¹ estudio del conjunto de genomas de todos los microorganismos que habitan un ecosistema concreto el cual se obtiene a partir de la secuenciación de muestras sin la necesidad de aislar ni cultivar cada una de las especies.

6.5.2.1. Análisis transcriptómico de las líneas celulares HT29 y NCM460

Para comenzar el estudio de las diferencias a nivel transcriptómico entre el fenotipo de cáncer de colon y normal en humanos, el estudio partió del análisis de las líneas celulares HT29 (cáncer de colon) y NCM460 (células epiteliales de colon normales). A partir del ARN extraído de estas, se realizaron tres experimentos basados en tres tecnologías transcriptómicas diferentes: *RNA-seq* de IonTorrent, *RNA-seq* de Illumina y *microarrays Clariom D Human* de Affymetrix. Cabe destacar que cada uno de los experimentos se llevó a cabo en periodos diferentes a lo largo de 5 años: el primero fue el de IonTorrent en 2015, el segundo el de Illumina en 2016 y el tercero el de microarrays en 2020.

El diseño experimental en cada uno de los tres casos fue el mismo: el factor fijo *Línea Celular* a dos niveles, HT29 y NCM460. No obstante, el número de muestras en cada uno de los tres experimentos es diferente, pues para Illumina y los microarrays se han empleado 3 muestras para HT29 y otras 3 para NCM460, mientras que para IonTorrent se han empleado 4 de cada una.

Además, el experimento basado en microarrays procede de un diseño experimental más complejo el cual se explicará en detalle en el apartado “6.2 Tratamiento de células NCM460 y HT29 con DFMO.” En este, no sólo se han analizado las células HT29 y NCM460 en condiciones basales, sino que también se han analizado estas mismas líneas celulares tratadas con DFMO. Por lo tanto, en dicho experimento hay 4 condiciones experimentales diferentes y, entre ellas, las condiciones HT29.Control y NCM460. Control, que son las que han sido usadas para el análisis de expresión diferencial entre ambas líneas celulares.

6.5.2.2. Análisis transcriptómico del efecto del DFMO en los modelos celulares HT29 y NCM460

Tras realizar el tratamiento con DFMO tal y como se explicó en el apartado “6.2 Tratamiento de células NCM460 y HT29 con DFMO,” y extraer el ARN como se indicó en el apartado “6.4.1 Extracción de ARN a partir de líneas celulares,” se realizaron los microarrays Clariom D Human de Affymetrix. Para ello, el ARN extraído de cada una de las muestras se envió a la empresa encargada de realizar los microarrays que, en este caso, fue el Centro Andaluz de Biología Molecular y Medicina Regenerativa (<https://www.cabimer.es/web3/>). Posteriormente, dicha empresa devolvió los datos correspondientes en formato *.CEL* para su posterior análisis bioinformático y estadístico por nuestro grupo, tal y como se explica en el apartado “6.5.3.1 Análisis de datos de microarrays.”

En cuanto al diseño experimental, se tienen dos factores:

- *Línea Celular*: este factor se corresponde con las diferentes líneas celulares evaluadas, por lo que tiene dos niveles: *HT29* y *NCM460*.
- *Tratamiento*: este factor se corresponde con el tratamiento o no con DFMO, por lo que tiene dos niveles: *Control* y *DFMO* (tratadas con DFMO).

En consecuencia, combinando ambos factores se obtienen cuatro condiciones experimentales diferentes: *NCM460.Control*, *NCM460.DFMO*, *HT29.Control* y *HT29.DFMO*. Cabe destacar que, para evitar factores de confusión debidos al día de realización de cada réplica, en un mismo día se realiza una réplica (de un total de tres) de cada una de las cuatro condiciones experimentales. En cuanto a los factores aleatorios que pueden afectar al experimento y que, por ende, hay que controlar, son:

- *Día*: correspondiente al día de procesamiento de la muestra que, como se ha dicho antes, en un mismo día se procesa una réplica de cada condición experimental de un total de tres réplicas para cada una.

Es importante recordar que de forma simultánea a la preparación de cada una de las muestras de las que se ha extraído el ARN para estudiar el efecto del DFMO también se han preparado las muestras equivalentes para realizar el experimento de imagen de Ca^{2+} para estudiar este mismo efecto a nivel funcional.

6.5.2.3. Análisis transcriptómico de muestras de tejido humano de cáncer de colon

A partir de las muestras pareadas de tejido humano de cáncer de colon proporcionadas por el Banco de Tumores de Castilla y León (ver apartado “5.1.2 5.1.2”), se extrajo su ARN mediante el procedimiento descrito en el apartado “6.4.2 Extracción de ARN a partir de tejido” para, posteriormente, realizar el experimento transcriptómico mediante los microarrays Clariom D Human de Affymetrix. Estos microarrays fueron realizados por (CABIMER). Posteriormente, dicha empresa devolvió los datos correspondientes en formato *.CEL* para su posterior análisis bioinformático y estadístico por nuestro grupo, tal y como se explica en el apartado “6.5.3.1 Análisis de datos de microarrays.” En cuanto al diseño experimental, se tiene un factor fijo y uno aleatorio:

- *Tipo de tejido*: este es el factor fijo y se corresponde con el tipo de tejido, por lo que tiene dos niveles: *tumoral* y *normal*.
- *Paciente*: este es el factor aleatorio y se corresponde con el donante de las muestras, de tal manera que para cada individuo o paciente se dispone tanto de tejido tumoral como normal, por lo que es necesario controlar la correlación existente entre ellas.

Cabe destacar que de estas mismas muestras se extrajo también, de forma simultánea, su contenido proteico con el fin de analizar la expresión de diferentes proteínas, tal y como se explica en el apartado “6.6. Extracción de proteínas” .

6.5.3. Análisis de datos transcriptómicos de alto rendimiento

A continuación, se presenta el protocolo seguido para el análisis de datos transcriptómicos, tanto para microarrays de Affymetrix como *RNA-seq*. Además, en el apéndice “B.2 Caso práctico: análisis de microarrays de las muestras pareadas de tejido humano de cáncer de colon” se muestra un caso práctico de análisis de datos originales de microarrays y la interpretación de los resultados, extrapolable, en gran parte, a los datos de *RNA-seq*. Para más detalles sobre algunos de los conceptos más generales relacionados con este tipo de análisis, se recomienda leer el apéndice “A Fundamentos básicos para el análisis de datos.”

6.5.3.1. Análisis de datos de microarrays

En la presente tesis, todos los experimentos basados en microarrays han sido obtenidos a partir de *Affymetrix GeneChips*, tanto los realizados por nuestro grupo de investigación -líneas celulares y muestras de tejido- como los experimentos extraídos de repositorios públicos (NCBI GEO). Si bien es cierto que en el apéndice “A Fundamentos básicos para el análisis de datos” se hace un breve repaso de algunos de los conceptos necesarios para comprender estos análisis, se recomienda los textos Irizarry et al. (2003), B. Bolstad (2008) y Smyth (2004).

En primer lugar, se leen los archivos originales *.CEL*, que guardan los datos de fluorescencia de cada una de las sondas del microarray. A continuación, se realiza un control de calidad de cada uno de los microarrays, para lo que se emplean métodos generales y específicos de microarrays de Affymetrix. Por un lado, los métodos generales empleados son: estimadores de la densidad *kernel* de la distribución de los valores de fluorescencia originales; gráficos de cajas y bigotes de los valores originales de fluorescencia; análisis de componentes principales (ACP) combinados con la detección de *outliers* multivariante mediante la distancia robusta de Mahalanobis, y análisis de clúster jerárquico con el método de Ward. Por otro lado, los métodos específicos para microarrays de Affymetrix son métodos a nivel de sonda: gráficos de degradación del ARN y modelización del valor de fluorescencia de los microarrays y, de estos modelos, se analizan sus valores *NUSE* y *RLE*.

Tras el control de calidad anterior, se eliminan las muestras que no superen dicho control y, a continuación, se preprocesan los datos mediante el método *robust multi-array average (RMA)* el cual corrige el ruido de fondo, normaliza los microarrays para hacerlos comparables entre sí y resume los valores de todas las sondas del mismo conjunto de sondas en un único valor (Irizarry et al. (2003), B. Bolstad (2008)). Una vez preprocesados los datos, debido a la gran dimensión de este tipo de datos (decenas de miles de sondas), es necesario realizar un filtrado no específico para eliminar aquellas sondas que no se estén expresando, que no varíen de una forma evidente su expresión a lo largo de los diferentes microarrays -sin tener en cuenta ninguna covariable o factor- y/o que no disponían de una correspondencia con la anotaciones usadas en los análisis posteriores de expresión diferencial -ALIAS- o de enriquecimiento -UNIGENE-. Además, puesto que los diferentes análisis se van a centrar, principalmente, en los genes relacionados con el transporte de Ca^{2+} intracelular, las sondas correspondientes a estos se mantendrán tras el filtrado independientemente de su expresión.

Posteriormente, se realiza un análisis de patrones. Para ello, se emplea el ACP y, mediante el algoritmo de árboles de clasificación C5.0, se selecciona/n la/s componentes principales (CP) que mejor clasifica/n las muestras según el interés del estudio (entre muestras sanas y tumorales). Además, con el fin de encontrar qué genes son más importantes respecto de la variable de interés, se analiza la correlación de cada uno de los genes con dichas CP -que mejor clasifican-. En concreto, se contrasta la hipótesis nula $H_0 = \rho_{jk} = 0$ (no hay correlación) frente a la alternativa $H_1 = \rho_{jk} \neq 0$ (sí hay correlación) suponiendo que el estimador de ρ_{jk} , r_{jk} , sigue una *t de Students* con $n-2$ grados de libertad y empleando el método de la *transformación Z de Fisher* para estimar los intervalos de confianza.

A continuación, se realiza el análisis de expresión diferencial mediante el método limma (Ritchie et al. (2015)) el cual realiza un modelo lineal sobre cada conjunto de sondas a lo largo de todas las muestras y, posteriormente, para mejorar las estimaciones y, por ende, conseguir unos contrastes de hipótesis más fiables, se emplea un enfoque bayesiano en el que se tiene en cuenta la información de todos los genes de forma simultánea para mejorar la estimación de los parámetros (Smyth (2004)). Posteriormente, para afrontar el problema de las comparaciones múltiples, se controla el *False Discovery Rate* mediante el método propuesto por Benjamini and Hochberg (1995); en otras palabras, se corrige el p valor. Cabe destacar que, en ocasiones, en función del tipo de microarray, pueden existir conjuntos de sondas redundantes (un mismo gen puede estar representado por más de un conjunto de sondas en el microarray), por lo que, en función del interés del análisis, ha de tratarse este “problema” de un modo u otro. Con el fin de no entorpecer la lectura, en el apartado “6.5.3.1.1. Sondas redundantes” se explica más detalladamente el procedimiento llevado a cabo para tal fin.

Tras el análisis de expresión diferencial y obtener una lista de genes diferencialmente expresados (DEG) se busca el significado biológico de esta. En este caso, se realiza un análisis de enriquecimiento de genes y un análisis de los perfiles de expresión tanto de los genes como de las muestras. Por un lado, el análisis de enriquecimiento permite inferir qué conjuntos de genes podrían estar asociados con el estudio (por ejemplo, con las diferencias entre tejido tumoral y sano, o entre células tratadas con DFMO y no tratadas). Para ello, es posible utilizar información previamente generada por la comunidad científica sobre diferentes grupos de genes con algún significado biológico (rutas de señalización, genes implicados con el cáncer...). Para tal fin, se han creado diferentes bases de datos, como por ejemplo KEGG PATHWAY, GO y MSigDB, cada una de las cuales puede estar dividida en diferentes colecciones de conjuntos con ciertos significados. Por ejemplo, dos de las colecciones que hay dentro de MSigDB están directamente relacionadas con el cáncer: *Oncogenic Signature Gene Sets* y *Computational Gene Sets*. Por lo tanto, cruzando los datos de la lista de DEG con dichas bases de datos es posible, por ejemplo, inferir qué rutas de señalización están alteradas en el fenotipo tumoral de cáncer de colon respecto del sano.

En cuanto al contraste de hipótesis que se realiza, se contrastan la hipótesis nula competitiva y la autocontenida (L. Tian et al. (2005)): hipótesis nula competitiva, $H_{0,Comp}$ que afirma que los genes en un grupo dado S están como mucho tan frecuentemente expresados de un modo diferencial como los genes en el universo de genes, y la hipótesis nula autocontenida, $H_{0,A}$, por la que ningún gen en

Se está diferencialmente expresado. Por lo tanto, rechazar alguna de estas hipótesis nulas significará que el conjunto de genes está asociado con el objeto del estudio (por ejemplo, con el fenotipo tumoral si la lista de DEG proviene e comparar tejido tumoral con tejido sano). Para contrastar estas hipótesis, se han empleado dos métodos: el análisis de sobrerrepresentación (ORA), para contrastar $H_{0,Comp}$, y CAMERA (Di Wu and Smyth (2012)), para contrastar $H_{0,A}$. En la presente tesis se han consultado las bases de datos KEGG PATHWAY, GO y MSigDB, con alrededor de 30000 conjuntos de genes en total. Puesto que hay un número demasiado elevado, el análisis de estos datos se ha centrado en obtener un resumen numérico de cuántos conjuntos en total rechazan alguna de las dos hipótesis nulas y, de estos, cuántos pertenecen a bases de datos relacionadas con el cáncer (*Oncogenic Signature Gene Sets* o *Computational Gene Sets*) y cuántos contienen en su nombre alguna de las siguientes palabras clave de interés para la temática de la tesis: *calcium*, *mitochondria*, *reticulum*, *cancer*, *wnt* y *poliamines*.

En cuanto al análisis de los perfiles de expresión, se realizan un análisis de clúster jerárquico con el método de Ward tanto sobre las muestras como sobre los genes con el fin de encontrar agrupaciones entre las muestras o entre los genes, respectivamente, pero teniendo en cuenta sólo aquellos genes que hayan mostrado expresión diferencial en alguno de los contrastes de hipótesis llevados a cabo. Posteriormente, se representan los valores de expresión centrados mediante un mapa de calor y se agrupan tanto las muestras como los genes en función del análisis clúster anterior.

6.5.3.1.1. Sondas redundantes

En ocasiones, los microarrays presentan diferentes conjuntos de sondas correspondientes a un mismo gen. Además, puede suceder que algunos conjuntos de sondas de un gen mapeen frente a regiones comunes a una familia de genes, mientras que otros sean específicos de regiones de ciertas isoformas. En particular, en los microarrays considerados en la presente tesis, tanto los realizados en los experimentos en las líneas celulares y en las muestras de tejido humano pareadas como aquellos extraídos de repositorios públicos, se da tal situación. No obstante, en los empleados en los experimentos realizados por nuestro propio grupo, dadas las características del tipo de microarray empleado, pues es del tipo HTA -*clariom D human de Affymetrix*- (ver apartado), la redundancia de los conjuntos de sondas es menos preocupante que en el caso de otros como, por ejemplo, el microarray *Human Genome U133 Plus 2.0* el cual es del tipo 3'.

Para el primero, el Clariom D human, los únicos conjuntos de sondas redundantes encontrados (entre los genes de interés) son para el *TRPM2* y el *TRPC7* que, además, ambos genes comparten una misma sonda, la *TC2100007359.hg.1*. Por el contrario, para el segundo, el microarray *Human Genome U133 Plus 2.0*, existen numerosos genes que poseen varios conjuntos de sondas redundantes. Es más, en este segundo tipo de microarrays existe una nomenclatura para los conjuntos de sondas (figura 6.15):

- *Sin sufijo*: las sondas del conjunto de sondas hibridan únicamente con parte de la secuencia consenso de un único transcrito y no presentan hibridación cruzada con la secuencia de otros transcritos. Se denominan *Conjuntos de Sondas Únicos* o *Unique Probe Set*.
- *Sufijo _at*: indica que las sondas del conjunto de sondas hibridan con la hebra antisentido² del gen de interés.
- *Sufijo _a*: las sondas del conjunto de sondas presentan hibridación cruzada con secuencias consenso consideradas de la misma familia (indicado como tal en el atributo *geneCluster*³ de los archivos de anotación); es decir, que si hay varios transcritos que comparten cierta secuencia consenso, entonces las sondas del conjunto de sondas presentan hibridación cruzada con dichos transcritos.
- *Sufijo _s*: las sondas del conjunto de sondas presentan hibridación cruzada con secuencias consenso consideradas de distinta familia (puede que no se disponga de la información sobre la familia *-geneCluster*⁴- o que hibride con secuencias pertenecientes a diferentes familias).
- *Sufijo _x*: el conjunto de sondas contiene al menos una sonda que presenta hibridación cruzada con otra secuencia. Otras sondas del mismo conjunto de sondas pueden ser únicas para la secuencia consenso de un solo transcrito.

²En el contexto de la biología molecular y genética, la hebra ADN de sentido, también conocida como hebra codificante, es aquella que contiene la secuencia 5' → 3' complementaria a la cadena antisentido de ADN, conocida como hebra molde 3' → 5'. La primera tiene la misma secuencia (en el lenguaje de ADN) que el ARNm correspondiente (en el lenguaje del ARN) y, en consecuencia, la molde tiene la secuencia complementaria al ARNm.

³El concepto *clúster de genes* se refiere a aquellos genes de un mismo organismo que codifican para proteínas similares (homólogas, de la misma familia génica), que comparten una función generalizada. Para cada uno de los genes del mismo clúster, a pesar de que la porción de la secuencia de ADN que comparten es idéntica, la proteína resultante codificada por cada uno de los genes es diferente.

⁴El concepto *clúster de genes* se refiere a aquellos genes de un mismo organismo que codifican para proteínas similares (homólogas, de la misma familia génica), que comparten una función generalizada. Para cada uno de los genes del mismo clúster, a pesar de que la porción de la secuencia de ADN que comparten es idéntica, la proteína resultante codificada por cada uno de los genes es diferente.

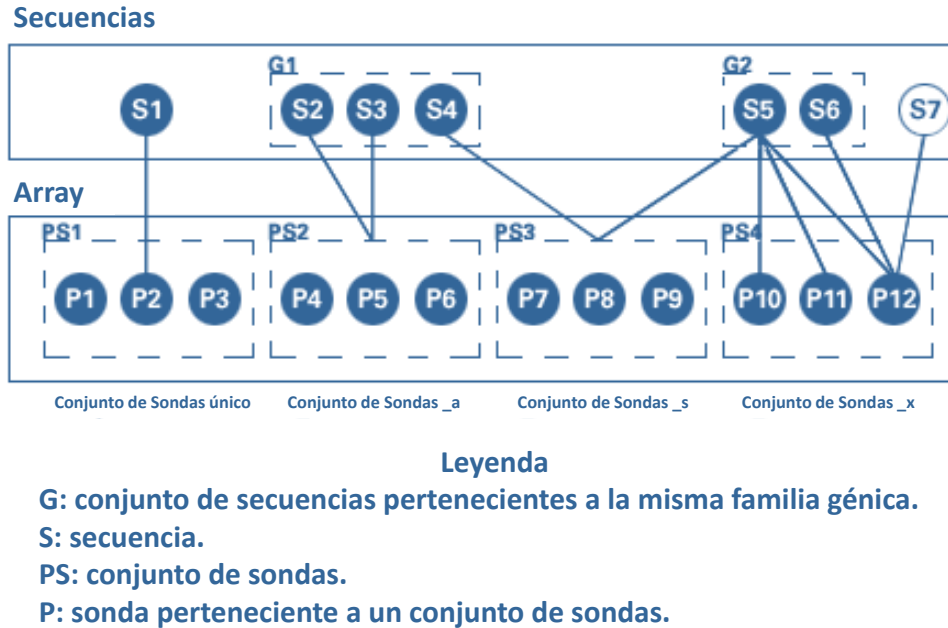


Figura 6.15: Ejemplo de sondas redundantes. En la imagen anterior se muestra aun ejemplo de los tipos de sondas redundantes que puede haber en un chip de expresión 3' de Affymetrix; en este caso, para el chip *Mouse Expression Set 430*. Las secuencias diana (S), que pueden pertenecer a un clúster de genes dado (G) pueden ser reconocidas o están representadas por diferentes sondas (P) pertenecientes a un conjunto de sondas (PS) del microarray, donde la relación anterior se indica por las líneas sólidas que conectan los nodos del grafo anterior. Además, para indicar que todas las P de un mismo PS se relacionan con una o varias S, la relación se muestra entre la “caja” que engloba a las P del mismo PS. Un ejemplo de conjuntos de sondas **sin sufijo** se muestra con la relación S1 – PS1, pues indica que todas las sondas del conjunto de sondas PS1 representan únicamente a la secuencia S1. Un ejemplo de conjuntos de sondas con sufijo **_a** se muestra mediante las relaciones S2 – PS2 y S3 – PS2, donde se observa como las secuencias tanto S2 como S3, que pertenecen al mismo clúster de genes (G1), están representadas en el microarray por el mismo conjunto de sondas, PS2; sin embargo, a pesar de que S4 pertenece a la misma familia, no está representado por este mismo conjunto de sondas. Un ejemplo de conjuntos de sondas con sufijo **_s** se puede observar en las relaciones establecidas con PS3, S4 – PS3 y S5 – PS3, que muestran como todas las sondas del conjunto de sondas PS3 representa a secuencias de diferentes familias de genes. Por último, para ejemplificar un conjunto de sondas con sufijo **_x** se tienen las relaciones con PS4, pues a pesar de que todas las sondas del conjunto de sondas representan a la secuencia S5 perteneciente a la familia genética G2, la sonda P12 de PS4 muestra hibridación cruzada, pues no solo representa a S5, sino también a S6, de la misma familia génica de S5 secuencia -G2-, y a S7, que no pertenece al clúster G2. Modificado de Affymetrix (ver enlace).

Esto da lugar a diferentes escenarios: todos los conjuntos de sondas frente al mismo gen produzcan un valor de logFC similar y, además, que todos sean significativamente diferentes de 0; que únicamente algunas sean significativamente diferentes de 0, o que haya valores de logFC tan diferentes que en ocasiones sean significativamente diferentes de 0 pero en sentido contrario. Ante esta situación, es lógico que surjan diferentes cuestiones en relación a cuál es la mejor forma de evaluar la expresión del gen de interés, sobretudo en aquellos casos en los que se obtienen estimaciones contradictorias: ¿Cómo evaluar adecuadamente la expresión del gen de interés? ¿Qué sonda es más fiable? ¿Se puede saber si alguna de las estimaciones está equivocada? ¿Es adecuado promediar los valores de expresión obtenido por los diferentes conjuntos de sondas para un mismo gen previo al análisis de expresión diferencial?...

Tras realizar una búsqueda en diferentes foros especializados en el análisis de datos de datos transcriptómicos, como es el caso de aquellos hilos existentes en la plataforma bioconductor, llama la atención los muy diferentes enfoques que existen para abordar este problema y que, a día de hoy, sigue siendo una pregunta abierta. No obstante, a pesar de las diferentes estrategias que se plantean, el uso de unas u otras dependerá, en cierto modo, del fin último del análisis. Por un lado, para el análisis de expresión diferencial o crear modelos de clasificación, entre otros, se pueden emplear los valores a nivel de sonda, sin necesidad de resumir en un único valor de expresión para cada gen, si bien es cierto que al realizar, por ejemplo, metaanálisis con diferentes plataformas, puesto que cada plataforma puede tener sus propias sondas, sí que es necesario resumir la expresión de cada gen en un único valor de expresión. Además, puesto que los diferentes conjuntos de sondas para un mismo gen pueden proporcionar información sobre la expresión de diferentes formas de *splicing* alternativo, también existen métodos de análisis de expresión diferencial en las que se evalúa la expresión diferencial a nivel de isoformas de *splicing* alternativo. No obstante, en la presente tesis no se han tenido en cuenta las diferentes isoformas de *splicing* para un mismo gen, sino la expresión diferencial del gen *per se*. Por otro lado, para el análisis de enriquecimiento, puesto que también es necesario un único valor de expresión por gen, también se hace necesario combinar de algún modo los valores de los diferentes conjuntos de sondas correspondientes a un mismo gen en un único valor. Algunas de las estrategias más habituales para combinar y resumir los valores de los diferentes conjuntos de sondas en un único valor son:

- Bajo el supuesto de que los conjuntos de sondas del mismo gen tienen un patrón de expresión similar, pero con diferente media, se selecciona aquel conjunto de sondas que presente la segunda expresión más alta en la mayoría de las muestras. Así, se propone que, por un lado, aquellos conjuntos de sondas con mayores niveles de expresión evitarían estar “por debajo” de la detección en los tejidos donde realmente se expresan y, por otro lado, al evitar seleccionar la de mayor expresión se evita, en parte, escoger un conjunto de sonda que pudiera corresponderse con valores extremos u *outliers* (ver enlace).
- Seleccionar aquel conjunto de sondas que tenga la mayor medida de dispersión (varianza, MAD...) a través de todos los microarrays (enlace)
- Si al menos uno de los conjuntos de sondas resulta ser significativa, se considera que el gen correspondiente está diferencialmente expresado. (ver enlace).
- Tras obtener los valores de logFC, considerar aquel conjunto de sondas que proporcione el valor de logFC más extremo (ver enlace).

Las estrategias anteriores conllevan una serie de problemas. Por un lado, se pierde toda la información correspondiente al resto de sondas que no se consideran. Es más, puesto que la hibridación o la fluorescencia registrada puede afectar a la sensibilidad (cambio en la señal a un cambio dado de ARNm), hay que tener cautela con seleccionar conjuntos de sondas con valores extremos tanto aquellos conjuntos de sondas con una baja pueden fallar a la hora de detectar un

cambio verdadero entre condiciones, y aquellos conjuntos con una elevada hibridación o intensidad de fluorescencia pueden saturar la señal y, en consecuencia, infraestimar los cambios que realmente tienen lugar entre condiciones para un gen dado (gen correspondiente al conjunto de sondas en cuestión). Por lo tanto, seleccionar aquellos conjuntos de sondas con una mayor intensidad, mayor varianza... puede dar lugar a sesgos. Por otro lado, cuando diferentes conjuntos de sondas estiman valores de $\log FC$ en sentidos opuestos y significativamente diferentes de 0, seleccionar un único conjunto de sondas de entre todos ellos puede ser un tanto arriesgado, pues no tiene sentido creer que se expresa diferencialmente en un sentido y no en el otro simplemente porque sea más o menos significativo (dentro de que sí que lo sean), o porque el valor de $\log FC$ sea mayor en valor absoluto.

Así, surgen otras estrategias que, en vez de escoger aquella sonda que más significativa sea o mayor $\log FC$ en valor absoluto presente tras el análisis de expresión diferencial, lo que proponen es, previo al análisis de expresión diferencial, realizar la media (mediana...), dentro de cada microarray (muestra), de los valores de los diferentes conjuntos de sondas para un mismo gen. En particular, el paquete `limma` implementa una función que realiza la media de los valores de los diferentes conjuntos de sondas para un mismo gen dentro de cada uno de los microarrays, `avereps()`. No obstante, resumir en un único valor de intensidad de fluorescencia para el gen g dentro de cada microarray a partir del valor de la intensidad de fluorescencia de las diferentes sondas redundantes dentro de dicho microarray mediante la media aritmética de dichas intensidades de fluorescencia para, posteriormente, estudiar la expresión diferencial para dicho gen a partir del valor resumido no tiene por qué ser la mejor opción, pues no tiene por qué ser la media aritmética el mejor estimador del parámetro poblacional *Intensidad de Fluorescencia* para el gen g . Es decir, que podría ser la mediana, la media geométrica... Así, la opción que se propone en la presente tesis, en lugar de resumir en un único valor de intensidad de fluorescencia para el gen g en el microarray m a partir de la intensidad de fluorescencia de las diferentes sondas redundantes para ese mismo gen g y microarray m previo al análisis de expresión diferencial, es combinar los valores de $\log (FC)$ de las diferentes sondas redundantes s para el gen g , $\log (FC)_{sg}$, en un único valor de $\log (FC)$ global para el gen g , $\log (FC)_g$. Es decir, realizar el análisis de expresión diferencial a nivel de sonda, $\log (FC)_{sg}$, y, a continuación, combinar los valores $\log (FC)_{sg}$ en un único valor $\log (FC)_g$ de expresión diferencial para el gen g , de tal manera que, por un lado, se tiene la información a nivel de sonda y, por otro lado, el resumen del análisis de expresión diferencial para cada gen.

No obstante, es lógico plantearse cuál es la mejor forma de combinar los N valores estimados individualmente de cada una de las sondas redundantes s correspondientes al gen g . Es decir, cómo obtener un estimador global centrado $\widehat{\log (FC)}_g$ del parámetro verdadero $\log (FC)_g$ de tal manera que este sea el más preciso posible. Además, antes de plantearse cómo realizar dicha combinación también hay que tener en cuenta que el análisis de expresión diferencial es un análisis marginal, es decir, que se hace la suposición de que los genes son independientes unos de otros y, por extensión, que las sondas redundantes para un mismo gen también lo son, pues debido a la gran dimensión de los datos no es viable considerar la correlación que podría existir entre todos los genes (maldición de la dimensión). Por lo tanto, la forma más adecuada de obtener un estimador global centrado y de máxima precisión (menor varianza) de $\log (FC)_g$ es a través de la combinación lineal de los

N estimadores centrados del parámetro de interés $\log(\text{FC})_g$; osea, se realiza una combinación lineal de los N valores de cada una de las sondas redundantes del gen g , $\widehat{\log(\text{FC})}_{s,g}$. En cuanto a los coeficientes que multiplica a cada estimador, se puede demostrar que ha de ser directamente proporcional a su precisión.

Para ejemplificarlo, se va a presentar un caso más sencillo en el que se mide la expresión diferencial de un gen g mediante dos sondas diferentes, 1 y 2, de tal manera que cualquier estimador definido como en la ecuación 6.1 será centrado:

$$\widehat{\log(\text{FC})}_g = c \cdot \widehat{\log(\text{FC})}_{1,g} + (1 - c) \cdot \widehat{\log(\text{FC})}_{2,g} \quad (6.1)$$

donde c es el coeficiente de la combinación lineal el cual puede interpretarse como un peso ponderación que se otorga a cada estimador centrado de cada sonda redundante. A continuación, hay que encontrar los valores de c que hagan que el estimador centrado $\widehat{\log(\text{FC})}_g$ sea el más preciso o el de menor varianza, por lo que sabiendo que la varianza del estimador global en función de los estimadores de cada sonda redundante es (ecuación 6.2):

$$\text{Var}(\widehat{\log(\text{FC})}_g) = c^2 \cdot \text{Var}(\widehat{\log(\text{FC})}_{1,g}) + (1 - c)^2 \cdot \text{Var}(\widehat{\log(\text{FC})}_{2,g}) \quad (6.2)$$

se tiene que el estimador de máxima verosimilitud de $\log(\text{FC})_g$, que se obtiene derivando la ecuación ecuación 6.2 respecto de los coeficientes c e igualando a 0, es (ecuación 6.3):

$$\frac{d\text{Var}(\widehat{\log(\text{FC})}_g)}{dc} = 2 \cdot c \cdot \text{Var}(\widehat{\log(\text{FC})}_{1,g}) - 2 \cdot (1 - c) \cdot \text{Var}(\widehat{\log(\text{FC})}_{2,g}) \quad (6.3)$$

y, despejando c (ecuación 6.4):

$$c = \frac{\text{Var}(\widehat{\log(\text{FC})}_{2,g})}{\text{Var}(\widehat{\log(\text{FC})}_{1,g}) + \text{Var}(\widehat{\log(\text{FC})}_{2,g})} = \frac{1/\text{Var}(\widehat{\log(\text{FC})}_{1,g})}{1/\text{Var}(\widehat{\log(\text{FC})}_{1,g}) + 1/\text{Var}(\widehat{\log(\text{FC})}_{2,g})} \quad (6.4)$$

Por lo tanto, de las ecuaciones 6.2 y 6.4 se tiene que, de forma general, el estimador centrado del efecto global más preciso (de menor varianza) se obtiene mediante la combinación lineal de los estimadores independientes del parámetro de interés (expresión diferencial del gen g , por ejemplo) de cada una de las sondas redundantes, tal que los coeficientes de cada término son directamente proporcionales a la precisión (inverso de la varianza) de cada estimador, $w_s = 1/\text{Var}(\widehat{\log(\text{FC})}_{s,g})$, o sea (6.5):

$$\widehat{\log(\text{FC})}_g = \frac{\sum_{s=1}^N w_s \cdot \widehat{\log(\text{FC})}_{s,g}}{\sum_{s=1}^N w_s} \quad (6.5)$$

Por ello, este método se conoce como el *método de la inversa de la varianza*. En cuanto a la varianza del estimador global centrado más preciso $\widehat{\log(\text{FC})}_g$, $\text{Var}(\widehat{\log(\text{FC})}_g)$, será la inversa del sumatorio de los pesos o coeficientes de la combinación lineal anterior (ecuación 6.6):

$$\widehat{\text{Var}}(\widehat{\log(\text{FC})}_g) = \frac{1}{\sum_{s=1}^N w_s} \quad (6.6)$$

Finalmente, tras realizar la combinación de los logFC de las sondas redundantes, y suponiendo que, al igual que los estimadores individuales, la combinación también sigue una distribución t con los mismos grados de libertad (con el fin de que estos estimadores combinados estén acorde al resto de genes que no presentan sondas redundantes), se estima el valor de t (logFC/S) y se obtiene el p valor asociado a dicho estadístico t para esos mismos grados de libertad. Por último, todos los p valores, tanto los obtenidos mediante la combinación de los logFC de las sondas redundantes como de aquellos genes que no presentaban sondas redundantes, se corrigen a través del control de FDR. Además, para los valores de las sondas redundantes *per se*, previo a la combinación de estas, se mantendrá el valor de FDR obtenido previamente a la combinación; es decir, que los valores de FDR para las sondas redundantes se obtiene tras el análisis de expresión diferencial realizado a nivel de sonda. Posteriormente, se obtienen los valores de $\widehat{\log(\text{FC})}_g$ (combinaciones de los valores de las sondas redundantes), se eliminan las sondas redundantes de la base de datos y, a partir, únicamente, de los $\widehat{\log(\text{FC})}_g$ que resultan de la combinación de las sondas redundantes y de los $\widehat{\log(\text{FC})}_g$ de aquellos genes que sólo poseen una sonda en el microarray se corrigen nuevamente los p valores obtenidos para cada uno de los genes y se obtienen los FDR para cada gen. Por lo tanto, por un lado, se tienen los valores de expresión diferencial a nivel de sonda y, por otro lado, los valores de expresión diferencial a nivel de gen en términos de la anotación escogida para identificar cada gen (ALIAS, ENTREZ, RefSeq...)

A continuación, se muestra el código que se ha empleado para $\widehat{\log(\text{FC})}_g$ a partir de las estimaciones de las diferentes sondas redundantes $\widehat{\log(\text{FC})}_g$:

```
# Objetos obtenidos previamente al análisis de expresión diferencial
## ExpressionSet tras el filtrado
esetFilt_A
## Matriz de diseño
matdis
## Matriz de contrastes
matcont
## Paquete de anotación correspondiente al microarray
annotation.db

# Análisis de expresión diferencial
```

```

## En caso de que se consideren factores aleatorios en el modelo
correlacion_modelo <- duplicateCorrelation(esetFilt_A, matdis,
                                           block = pData(esetFilt_A)[, "RandomFactors"])
modelo <- lmFit(esetFilt_A, matdis,
               block = pData(esetFilt_A)[, "RandomFactors"],
               correlation = correlacion_modelo$consensus)
contraste <- contrasts.fit(modelo, matcont)
contrasteB <- eBayes(contraste)

# Extraer datos necesarios para realizar la combinación de logFCsg para cada uno
# de los contrastes de interés
topTablesNombres <- NULL
for(i in colnames(contrasteB$p.value)){
  # Contraste de interés 'i'
  DEGs <- topTable(contrasteB,
                  number = nrow(contrasteB),
                  coef = i,
                  adjust.method = "BH",
                  confint = TRUE)

  # Información de interés
  informacion_adicional_DEGs <- data.frame(logFC = contrasteB$coefficients[, i],
                                           dfttotal = contrasteB$df.residual + contrasteB$df.prior,
                                           n = nrow(contrasteB$coefficients),
                                           S = contrasteB$stdev.unscaled[, i] * sqrt(contrasteB$s2.post),
                                           tCI = qt(0.975, (contrasteB$df.residual + contrasteB$df.prior)),
                                           CI.L = contrasteB$coefficients[, i] -
                                             qt(0.975, (contrasteB$df.residual + contrasteB$df.prior)) *
                                             contrasteB$stdev.unscaled[, i] * sqrt(contrasteB$s2.post),
                                           CI.R = contrasteB$coefficients[, i] +
                                             qt(0.975, (contrasteB$df.residual + contrasteB$df.prior)) *
                                             contrasteB$stdev.unscaled[, i] * sqrt(contrasteB$s2.post))
  informacion_adicional_DEGs <- informacion_adicional_DEGs[rownames(DEGs), ]
  colnames(informacion_adicional_DEGs) <- c("logFC", "dfttotal", "n", "S",
                                           "tCI", "CI.L", "CI.R")
  DEGs <- cbind(DEGs, informacion_adicional_DEGs)

  # Asignar nombre a la información anterior para cada contraste de interés
  DEG_Nombre <- paste("DEGs_", i, sep = "")
  assign(paste(DEG_Nombre), DEGs)
  # Definir un vector cuyos elementos sean los nombres de los objetos que

```

```

# contienen la información anterior para cada contraste
topTablesNombres <- c(topTablesNombres, DEG_Nombre)
}

# Iterar sobre cada contraste de interés
for(i in topTablesNombres){
  nombre <- gsub("DEGs_", i, replacement = "")
  # Extraer la información necesaria del contraste de interés 'i'
  genes <- get(i)[, - 1]
  # Mapear entre diferentes anotaciones
  IDs <- AnnotationDbi::select(annotation.db,
                              keys = rownames(genes),
                              columns = "ALIAS",
                              keytype = "PROBEID")
  indices <- match(IDs$PROBEID, rownames(genes))
  tablaTotal <- cbind(PROBE = IDs$PROBEID, ALIAS = IDs$ALIAS, genes[indices, ])
  tablaTotal <- tablaTotal[!is.na(tablaTotal$ALIAS), ]

  # Combinar los valores de expresión de las sondas redundantes (PROBEID)
  # que pertenezcan a un mismo ID ALIAS
  combinacion_exprs <- (by(tablaTotal[, c("logFC", "S", "tCI", "PROBE",
                                         "dfttotal")], tablaTotal$ALIAS,
                          function(x){
                            # Se combinan los logFC ponderando por su Varianza ( $S^2$ )
                            varianzas <- x[, 2] ^2
                            wt <- 1 / (varianzas) # Pesos
                            logFCp <- x[, 1] * wt # Sumandos ponderados
                            # logFC combinado
                            logFC <- sum(x[, 1] * wt)/sum(wt,na.rm=TRUE)
                            # Varianza del estimador combinado
                            S2 <- 1/sum(wt,na.rm=TRUE)
                            S <- sqrt(S2)
                            # Intervalos de confianza (donde t, c[1, 3], es el mismo que para cada sonda
                            # individual: 'qt(0.975, (contrasteB$df.residual + contrasteB$df.prior)) )'
                            CI.L <- logFC - x[1, 3] * S
                            CI.R <- logFC + x[1, 3] * S
                            # Se calcula el estadístico t de contraste de expresión diferencial
                            tstat <- logFC / S
                            # Se obtiene el p valor asociado al estadístico t anterior para unos grados

```

```

# de libertad iguales a los que se tienen para las sondas individuales
P.Value <- pt(abs(tstat),
              lower.tail = FALSE, df = tablaTotal$dftotal[1]) * 2
# Se añade "combined" a aquellos registros correspondientes a la
# combinación de las sondas redundantes
if(nrow(x) > 1){
  PROBE <- paste0(x[1, 4], "combined")
}else{
  PROBE <- x[1, 4]
}
dftotal <- x[1, 6]
return(c(PROBE, logFC, S, CI.L, CI.R, P.Value, tstat, dftotal))
}))

# Almacenar 'combinacion_exprs_2' en una matriz de 8 columnas
combinacion_exprs_2 <- data.frame(matrix(unlist(combinacion_exprs),
                                         byrow = TRUE, ncol = 8))
# Convertir de tipo 'character' a tipo 'numeric' aquellos valores del dataframe
# anterior que deban de ser numéricos
combinacion_exprs_2[, -c(1)] <- apply(combinacion_exprs_2[, -c(1)], 2,
                                     as.numeric)
rownames(combinacion_exprs_2) <- names(combinacion_exprs)
colnames(combinacion_exprs_2) <- c("PROBE", "logFC", "S", "CI.L", "CI.R",
                                   "P.Value", "t", "dftotal")
# Corregir los p valores (se hace sobre los nombres ID únicos)
# Para evitar que una misma sonda combinada corresponda a varios genes a la
# vez, se concatena con el logFC (es como si, en términos de bases de datos
# relacionales, se considerase como clave primaria al subconjunto de atributos
# {'PROBE', 'logFC'})
IDgenes_unico <- unique(paste0(combinacion_exprs_2$PROBE,
                               combinacion_exprs_2$logFC))
indices_ID_unico <- match(IDgenes_unico,
                        paste0(combinacion_exprs_2$PROBE,
                               combinacion_exprs_2$logFC))
# Se ajustan los p valores considerando únicamente un valor por gen; es decir,
# aquellas tuplas correspondientes a las combinaciones de las sondas
# redundantes (una por gen) y a aquellas que únicamente tienen una sonda por
# gen
pvalores_ajustados <- p.adjust(combinacion_exprs_2$P.Value[indices_ID_unico],
                               method = "BH")

```

```

# Se guardan los valores FDR anteriores en un nuevo atributo 'adj.P.Val', pero
# sólo en aquellas tuplas en las que se haya realizado dicha corrección tras
# la combinación de las sondas redundantes
indices_ID_unico_iv <- match(paste0(combinacion_exprs_2$PROBE,
                                   combinacion_exprs_2$logFC),
                             IDgenes_unico) # de único a BD original
combinacion_exprs_2$adj.P.Val <- pvalores_ajustados[indices_ID_unico_iv]

# Buscar los ALIAS que tienen más de un ID (genes con sondas redundantes)
varios_alias <- which(table(tablaTotal$ALIAS) > 1)
# Crear un objeto que contenga únicamente aquellas tuplas correspondientes
# a las sondas redundantes
tablaTotal_alias_varID <- tablaTotal[tablaTotal$ALIAS %in% names(varios_alias), ]
tablaTotal_alias_varID <- tablaTotal_alias_varID[sort(tablaTotal_alias_varID$ALIAS,
                                                    index.return = TRUE)$ix, ]

# Unir tanto las combinaciones de los logFC como las expresiones del mismo
# ALIAS con diferentes ID
combinacion_exprs_3 <- rbind(combinacion_exprs_2,
                             tablaTotal_alias_varID[, colnames(combinacion_exprs_2)])

# Ordenarlos según el ALIAS
combinacion_exprs_3 <- combinacion_exprs_3[sort(combinacion_exprs_3$ALIAS,
                                                    index.return = TRUE)$ix, ]

tablaTotal <- combinacion_exprs_3
# Guardar tabla expresion con ALIAS
# Adecuar el nombre del contraste de interés al nombre del csv exportado
# (no puede contener * ni /, entre otros)
DEG_Nombre_adequado_csv <- gsub(pattern = "[/*]", "_", i)
write.csv2(tablaTotal, file = file.path(PATH,
                                       paste(DEG_Nombre_adequado_csv,
                                             "_ALIAS.csv", sep = "")))
}

```

6.5.3.2. Análisis de datos de RNA-seq

Para los análisis de datos de las matrices de expresión de los experimentos basados en *RNA-seq* se sigue un procedimiento similar al usado para los microarrays (apartado 6.5.3.1). No obstante, hay ligeras diferencias. En primer lugar, no se aplican los métodos específicos para los microarrays de Affymetrix, pues ya se parte, en este caso, de la matriz de expresión de los experimentos de *RNA-seq* y, por ende, no hay ni que leerlos de ningún archivo especial ni que preprocesarlos. Otra diferencia es que el control de calidad se lleva a cabo sólo mediante métodos generales. Por último, puesto que los datos en *RNA-seq* son discretos (lecturas), en lugar de continuos (valores de fluorescencia), han de ser transformados previamente a los diferentes análisis. No obstante, hay que distinguir entre dos tipos de transformaciones diferentes que pueden hacerse; por un lado, para realizar análisis exploratorios de datos de RNA-seq se puede emplear la normalización *rlog* (Michael I. Love, Huber, and Anders (2014b)), implementado por el paquete *DESeq2* (Michael I. Love, Huber, and Anders (2014a)), que realiza una transformación logarítmica en base 2 de los datos de expresión y minimiza las diferencias entre las muestras para aquellos genes que tengan un número bajo de lecturas (ruido de Poisson) y la normalización frente al tamaño de la librería (figura B.2).

Por otro lado, a la hora de realizar análisis de expresión diferencial y de significación biológica, pueden usarse diferentes métodos, tales como *DESeq2* (Michael I. Love, Huber, and Anders (2014a)) o *edgeR* (Robinson, McCarthy, and Smyth (2010), McCarthy, Chen, and Smyth (2012)); sin embargo, en este caso se ha usado el método *limma* (Ritchie et al. (2015)), para el cual pues es necesario, en primer lugar, realizar una normalización de los datos mediante el método TMM (Robinson and Oshlack (2010)) para obtener los factores de escala que permiten convertir los datos de conteo originales en aquellos que tienen unos tamaños de librería efectivos. En segundo lugar, una transformación de estos datos normalizados empleando el método *voom* (Law et al. (2014)) -diferente al *rlog* (Michael I. Love, Huber, and Anders (2014b)) empleado para el análisis exploratorio- el cual lleva a cabo una transformación logarítmica de los datos de conteo en base 2 (logCMP) y una estimación de la relación entre la media y la varianza que se usará posteriormente para calcular los pesos o ponderaciones a nivel observacional para poder aplicar sobre estos datos el modelo lineal a cada gen mediante el método *limma*.

6.6. Extracción de proteínas

De las mismas muestras pareadas de tejido humano de cáncer de colon y normal de las que se extrae el ARN también se extraen, de forma simultánea, las proteínas. En particular, hasta el paso 11, ambas extracciones comparten la primera parte del protocolo presentado en el apartado “6.4.2 Extracción de ARN a partir de tejido” . Por lo tanto, el protocolo específico de la extracción de proteínas se enumera a partir del paso 11:

12. Se recupera la fase orgánica (fenol-etanol) y la interfase que durante la extracción de ARN se reserva a 4 °C.
13. Se precipita el ADN mediante la adición de etanol absoluto (300 µL/1 mL TRizol), se mezcla y se incuba a temperatura ambiente durante 3 min.
14. Tras la centrifugación (2000 g, 5 min, 4 °C), se obtiene un pellet en el cual se encuentra el ADN y el sobrenadante (fase orgánica fenol-etanol) con las proteínas.
15. Se transfiere la fase orgánica a un tubo eppendorf y se precipitan las proteínas. Para ello, se añade isopropanol (1,5 mL/1 mL TRizol) y se incuba 10 min, para después centrifugar.
16. Tras la centrifugación (12000 g, 10 min, 4 °C), se retira el sobrenadante y el pellet resultante se resuspende con una solución de cloruro de guanidinio 0.3 M (en etanol al 96 %) y se incuba durante 20 min.
17. De nuevo, tras centrifugar (7500 g, 5 min, 4 °C), se retira el sobrenadante y pellet resultante, en el cual están las proteínas, se deja secar al aire entre 5-10 min.
18. Por último, se solubilizan las proteínas mediante la adición de SDS 1 %. Con el fin de eliminar posibles contaminantes insolubles, la muestra resultante se centrifuga durante 10 min a 10000 g y 4 °C, de tal manera que en el sobrenadante se encuentran las proteínas.
19. Posteriormente, estas se cuantifican por el método de Lowry.

6.7. Análisis de la expresión de proteínas mediante *Western Blot*

A continuación, se va a presentar el protocolo de análisis de expresión de proteínas realizado mediante la técnica *Western Blot* en el cual se detallan los pasos seguidos y se argumenta brevemente cada uno de ellos. Posteriormente, se presenta el procedimiento y diseño experimental seguido. Por último, se explica el análisis de datos correspondiente.

6.7.1. Fundamentos teóricos y protocolo

El *Western Blot* fue desarrollado en el laboratorio de George Stark en la Universidad de Stanford, es una técnica analítica que permite la determinación de una proteína específica. A continuación, se explica brevemente cada uno de los pasos de esta técnica y su fundamento así como el procedimiento concreto llevado a cabo en el laboratorio:

- *Electroforesis*: en primer lugar, se utiliza la electroforesis en un gel de poliacrilamida, uni o bidimensional, con el fin de separar las proteínas en función del punto isoeléctrico, peso molecular y carga eléctrica. Así, la separación electroforética estará determinada por las características del gel utilizado y por el pretratamiento de las muestras como la homogenización del tejido y lisado celular, para lo que pueden emplearse tanto métodos físicos (sonicación, congelación/descongelación...) como químicos (enzimas, detergentes...). Debido a que las proteínas pueden tener hasta 4 niveles diferentes de estructuración, es necesario llevar a cabo un pretratamiento de las mismas para que la estructura tridimensional no influya en la separación. Para ello, las muestras de proteínas se mezclan con un tampón de carga 2x en relación 1:1 el cual contiene agentes desnaturizantes (β -mercaptoetanol, que rompe los puentes disulfuro, y urea) y se incuban a 95 °C durante 4 min. En consecuencia, se consigue la desnaturalización de las proteínas. No obstante, es importante tener sumo cuidado de no degradar la proteína durante este preprocesado, por lo que suelen emplearse bajas temperaturas (5 °C) e inhibidores de proteasas en el tampón empleado en la homogeneización. Las electroforesis pueden ser unidimensionales o bidimensionales, estas últimas capaces de resolver más de 1000 proteínas sobre un mismo gel (Alberts et al. (2010)). De los diferentes tipos de electroforesis, en la presente tesis se han empleado geles de poliacrilamida en presencia del tampón dodecilsulfato al 10 %, que interacciona con las proteínas e induce que estas adquieran densidad de carga negativa (SDS-PAGE, del inglés, *sodium dodecyl sulfate polyacrylamide gel electrophoresis*), en el cual se cargan las muestras de proteínas y el marcador de peso molecular (Precision Plus Protein Kaleidoscope Standard).

Los geles empleados, por lo tanto, actúan como tamiz molecular y se caracterizan por ser prácticamente inertes a la mayoría de las moléculas biológicas, resistentes a agentes desnaturizantes, estables mecánicamente y, además, pueden ser fácilmente almacenados tras su deshidratación. Su obtención se basa en reacciones de polimerización de la acrilamida

por acción de un agente entrecruzante, la bisacrilamida, en presencia de un iniciador, TEMED (N,N,N',N'-tetrametilendiamina) y un catalizador, el persulfato amónico $-(\text{NH}_4)_2\text{S}_2\text{O}_8^-$, de tal manera que la velocidad de polimerización está determinada por estos dos últimos. En cuanto a la porosidad del gel que, evidentemente, influirá en la separación entre las diferentes proteínas, viene determinada por las proporciones relativas de acrilamida/bisacrilamida. En particular, el tamaño del poro será menor cuanto mayor sea esta relación. La elección del tamaño del poro vendrá determinada por el rango de pesos moleculares de las proteínas a separar. En consecuencia, puesto que en la presente tesis se desea estudiar la expresión de las proteínas STIM1-2, ORAI1-2 y TRPC1, el porcentaje de acrilamida/bisacrilamida óptimo es del 10 %, ya que permite separar proteínas entre 15-100 KDa.

- *Transferencia*: tras haber realizado la electroforesis, se realiza una transferencia de las proteínas previamente separadas desde el gel a una membrana, como por ejemplo de nitrocelulosa. En este caso, se han empleado membranas de difluoruro de polivinilideno (PVDF) (Millipore, Bedford, MA). Cabe destacar que este proceso puede estar afectado por diferentes factores de los que dependerá el éxito de dicha etapa, como por ejemplo el hecho de que las proteínas de alto peso molecular se transfieren peor que las de menor peso.
- *Bloqueo*: tras la separación de las proteínas en función de su peso molecular, es hora de detectar aquellas de interés. De entre los diferentes métodos que hay, el más habitual es el *immunoblotting* el cual emplea anticuerpos contra la proteína de interés. Por ello, previo al uso de anticuerpos primarios, con el fin de que la unión antígeno-anticuerpo sea lo más específica posible, se realiza una etapa previa de bloqueo. Para ello, sobre la membrana que contiene las proteínas adheridas a ella se aplica un tampón de bloqueo, como por ejemplo suero de leche u ovoalbúmina con una pequeña cantidad de detergente. Con esta etapa de bloqueo se reduce la posibilidad de que un anticuerpo se una de forma inespecífica a una proteína e, incluso, que se una a los lugares que han quedado libres en la membrana tras la transferencia de las proteínas. Este paso es sumamente importante, pues protege frente a los errores tipo I o de cometer falsos positivos. En este caso, como solución de bloqueo se ha empleado 5 % de leche en polvo desnatada en solución de lavado TTBS y, a continuación, la membrana se incuba durante 90 min.
- *Detección*: posteriormente, la membrana se lava tres veces con TTBS y después se incuba con la solución de anticuerpo primario específico para la proteína de interés (diluido en el tampón de bloqueo) durante 12 horas a 4 °C. Esto permite poder detectar la proteína de interés de entre todas las presentes en la membrana. A continuación, la membrana se lava tres veces con TTBS para así eliminar todo el anticuerpo que no se haya unido a su antígeno y, seguidamente, se incuba en agitación con la solución de anticuerpo secundario conjugado con la enzima peroxidasa de Rábano Picante (HRP) el cual reconoce de forma específica al anticuerpo primario. Por último, la membrana se lava de nuevo con TTBS.

Finalmente, para poner de manifiesto las bandas de las proteínas de interés, la membrana se incuba con una mezcla de dos sustratos del Kit PIERCE ECL (Western Blotting Substrate

Thermo Scientific) en relación 1:1 durante 3 min. En este paso la enzima peroxidasa de rábano cataliza la reacción de oxidación del ILUMINOL ECL, de tal manera que emite luz mediante quimioluminiscencia. Por lo tanto, esta luz emitida es registrada a través del revelado de la membrana en un equipo VersaDoc 4000 (BioRad, Munich, Alemania). Finalmente, tras el revelado, el propio *software* del equipo digitaliza la imagen y esta se procesa para realizar un análisis de densitometría de las bandas con el *software* Quantity One (BioRad, Munich, Alemania). En este caso, como control de carga se ha usado la proteína β -actina, por lo que los resultados de expresión de las proteínas estudiadas se normalizan respecto a esta.

6.7.2. Procedimiento y diseño experimental

De las mismas muestras pareadas de tejido humano de cáncer de colon y normal empleadas en el experimento transcriptómico explicado en el apartado “6.5.2.3 Análisis transcriptómico de muestras de tejido humano de cáncer de colon,” se extrae la proteína, tal y como se explica en el apartado “6.6 Extracción de proteínas.” Tras obtener la proteína de cada muestra, se siguió el protocolo explicado en el apartado “6.7.1 Fundamentos teóricos y protocolo.” Las proteínas medidas mediante esta técnica fueron cinco relacionadas directamente con SOCE: ORAI1, ORAI2, TRPC1, STIM1 y STIM2. En cuanto al diseño experimental, se tienen dos factores fijos y dos aleatorios. Por un lado, los factores fijos son:

- *Tejido*: este es el factor fijo y se corresponde con el tipo de tejido, por lo que tiene dos niveles: *tumoral* y *normal*.
- *Sexo*: se disponen de muestras tanto de hombres como de mujeres, por lo que este factor tendrá dos niveles, *hombre* y *mujer*.

Por otro lado, los factores aleatorios son inherentes al tipo de muestra, pues son pareadas, y al tipo de técnica y que muy probablemente tengan efecto sobre la variable respuesta:

- *Gel*: correspondiente a un único ensayo de *Western Blot* que genera una única membrana. En este puede haber tanto intercepto aleatorio como pendientes aleatorias debido al factor *Sexo* como *Tejido*.
- *Paciente*: correspondiente al donante de cada par de muestras tumoral y normal. Es decir, que para un mismo paciente se dispone de dos muestras, una tumoral y otra normal y, de tal manera que siempre se procesan de forma simultánea una réplica de cada paciente. Por lo tanto, en este puede haber tanto intercepto aleatorio como pendiente aleatorias debido al factor *Tejido*.

Respecto a estos efectos aleatorios, es importante aclarar cómo se relacionan entre sí. Se hacen entre dos y tres réplicas de cada una de las proteínas, y las diferentes réplicas se analizan al menos en

dos geles diferentes, de tal manera que en un mismo gel siempre esté la muestra normal y tumoral de un mismo paciente. Por lo tanto, en cada gel hay muestras pareadas de dos o tres pacientes diferentes y, a su vez, las réplicas pareadas de un mismo paciente están también en diferentes geles. A este tipo de factores aleatorios se les conoce como factores aleatorios cruzados. Hay que tener en cuenta que es necesario controlar la correlación que generan estos factores aleatorios; es decir, que, por un lado, las muestras de un mismo gel estarán correlacionadas entre sí, y, además, las muestras pareadas (tumoral y sana) de un mismo paciente también estarán correlacionadas entre sí.

6.7.3. Análisis de datos de *Western Blot*

A continuación, se muestra el protocolo seguido resumido del análisis de los experimentos de expresión de proteínas mediante la técnica *Western Blot*. Además, en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos” se presenta un caso práctico de cómo analizar datos empleando los mismos fundamentos teóricos que se usarán en estos análisis, donde se muestra el procedimiento estándar llevado a cabo y la interpretación de los diferentes gráficos, tests y resultados. Más detalles en el apéndice “A. Fundamentos básicos para el análisis de datos.”

Tras revelar las membranas con el equipo VersaDoc 4000 (BioRad, Munich, Alemania) y obtener las imágenes de cada membrana mediante el *software* Quantity One (BioRad, Munich, Alemania), se extraen los valores de densidad óptica de cada muestra con el *software* *ImageJ*, un programa de dominio público del Instituto Nacional de Salud que permite procesar imágenes. Para ello, en primer lugar, se selecciona la región de interés de la imagen de la membrana analizada, de tal manera que a la imagen resultante se la considera como una matriz $n \times m$, $\mathbf{A}_{n \times m}$, donde n son los píxeles correspondientes a su altura y m los píxeles correspondientes a su anchura. A continuación, se crea un nuevo vector $\mathbf{b}_{1 \times m}$ donde el elemento i del vector $\mathbf{b}_{1 \times m}$ será la suma de los n elementos de la columna i de la matriz $\mathbf{A}_{n \times m}$; es decir, que es la suma de todos los vectores filas que componen la matriz $\mathbf{A}_{n \times m}$. Así, considerando al vector $\mathbf{b}_{1 \times m}$ como la variable respuesta y a la posición de cada elemento en el vector $\mathbf{b}_{1 \times m}$ como la variable independiente, se representan los m pares ordenados (i, b_i) , donde para cada muestra que contenga la proteína de interés habrá un incremento de la señal, por lo que, tras estimar el ruido de fondo y eliminarlo, el área bajo la curva de cada señal se corresponderá con la densidad óptica de cada muestra (figura 6.16). Para más información, consultar ejemplos en la guía del *software* *ImageJ* (ver enlace1 -guía- o enlace2 -video-). Finalmente, como variable respuesta, para cada muestra, se considera el ratio entre su densidad óptica para la proteína estudiada y la densidad óptica de la β -actina. En particular, se han estudiado cinco proteínas relacionadas con SOCE: ORAI1, ORAI2, TRPC1, STIM1 y STIM2.

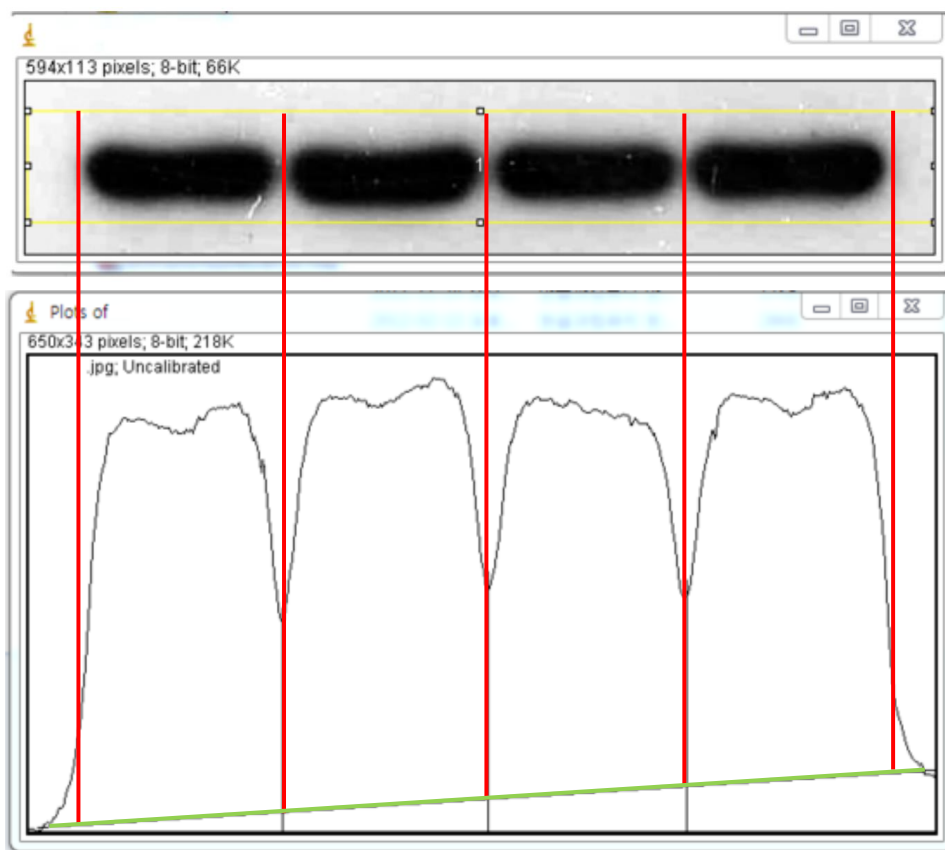


Figura 6.16: Obtención de los valores de densidad óptica. Tras escanear la membrana, la imagen generada se analiza con el *software ImageJ*. En primer lugar, se selecciona la región de interés (recuadro en amarillo). Posteriormente, se representan los valores de dicha región de izquierda a derecha y se corrige el ruido de fondo (se resta el área bajo la línea verde). Finalmente, se obtiene el área bajo la curva de cada una de las señales correspondiente a cada una de las muestras tras corregir el ruido de fondo (señales entre las líneas verticales rojas).

A continuación, para cada una de estas proteínas se han empleado modelos lineales para evaluar el efecto de los factores estudiados en estos experimentos (*Tumor* y *Sexo*) y, en caso de ser necesario, modelos lineales mixtos para controlar los factores aleatorios, *Gel* y *Paciente*, y estimar sus correspondientes efectos aleatorios (interceptos). Así, en primer lugar, se realiza un proceso de selección del modelo (las variables explicativas y efectos aleatorios que contendrá). Para ello, se parte del modelo sobreóptimo, que es aquel que contiene todas las posibles variables explicativas y sus interacciones así como todos los factores aleatorios posibles, y se selecciona la estructura aleatoria más adecuada. No obstante, en caso de que se concluya que no es necesario considerar la parte aleatoria entonces el modelo será de efectos fijos (sin ningún factor aleatorio) y se empleará la estimación por mínimos cuadrados para estimar el modelo multivariante de efectos fijos resultante (el habitual). Como criterio de selección se emplea la información de Akaike condicional (cAIC), de tal manera que el mejor modelo (comparando las estructuras aleatorias) será el que menor cAIC tenga. No obstante, hay ocasiones en las que el criterio cAIC puede indicar que la mejor opción es un modelo con parte aleatoria y, sin embargo, a la hora de estimar los componentes de la varianza

de algún efecto aleatorio correspondiente a algún factor aleatorio resulta ser cero, por lo que este se elimina del modelo a pesar del criterio de cAIC, pues en realidad significa que el modelo no tiene tal efecto aleatorio. En consecuencia, en función de la parte aleatoria el análisis posterior se llevará a cabo a partir de un modelo lineal mixto o uno de efectos fijos. Además, en caso de ser necesario, se recurrirá a modelos lineales generalizados (mixtos o no) e, incluso, en caso de violarse de forma preocupante los supuestos del modelo, a técnicas *bootstrapping*. Cabe destacar que no han sido necesarios estas últimas técnicas ya que se cumplen los supuestos del modelo para cada una de las proteínas.

Posteriormente, tras haber seleccionado la parte aleatoria más adecuada, se selecciona la parte fija mediante los criterios AIC y BIC. En cierto modo, esto indica si un factor es o no significativo, pues hay que tener presente que en los modelos lineales mixtos los contrastes F y t realizados, por ejemplo, en el ANOVA, hay que tomarlos con precaución, pues puede que un p valor esté ligeramente por encima de 0.05 (del nivel de significación) y ser realmente significativo y viceversa. Por lo tanto, la información obtenida por AIC y por la tabla de ANOVA sobre cada uno de los factores fijos (*Tumor* y *Sexo*) será complementaria. Además, puesto que se está interesado en ambos factores, sólo se eliminará el efecto de la interacción entre ambos en caso de no ser necesaria según estos criterios. Finalmente, en caso de que exista el término de interacción se realizan las comparaciones dos a dos entre todas las condiciones experimentales empleando el método de Tukey y, en caso de que no haya tal término, se realizarán las comparaciones dos a dos dentro de cada factor; es decir, entre normal y tumoral o entre hombre y mujer. Además, a partir de aplicar la teoría de grafos sobre estas diferencias dos a dos se analizó qué condiciones experimentales pueden ser consideradas iguales y cuales no a través de agruparlas en diferentes grupos. En concreto, se aplicó el algoritmo de Bron-Kerbosch⁵, de tal manera que dos condiciones representadas con una misma letra se consideran iguales y, además, en caso de tener más de una letra por ejemplo (*ab*) se considerarán similares a más de una condición (las condiciones del grupo *a* y del *b*), a pesar de que las condiciones a las que esta se parezca sean diferentes entre sí (las condiciones de *a* son diferentes de las condiciones de *b*).

Para evaluar los supuestos del modelo se realizó el diagnóstico de los residuos mediante métodos gráficos y basados en contrastes de hipótesis. Por un lado, se emplearon los gráficos Q-Q, histogramas de los residuos de Pearson junto con la estimación kernel de su densidad, gráficos de cajas y bigotes de los residuos de Pearson en función de las condiciones experimentales y diagramas de dispersión de los residuos de Pearson en función de los valores ajustados. Por otro lado, los test empleados para evaluar su normalidad fueron los de Shapiro-Wilks, Jarke Bera y Lilliefors, mientras que para su homocedasticidad se usó el método de Breush Pagan y el test de Bartlett. En cuanto a los efectos aleatorios, en el caso de que el modelo seleccionado sea mixto, también se evalúa su normalidad mediante gráficos Q-Q.

⁵Este algoritmo busca los cliques maximales. En un grafo no dirigido, un clique es un subgrafo en el cual todo vértice está conectado por una arista a todos los demás vértices pertenecientes a dicho subgrafo.

6.8. Metaanálisis

En este apartado se va a realizar una breve introducción al metaanálisis y, en particular, al metaanálisis de datos transcriptómicos. Posteriormente, se describirá el metaanálisis llevado a cabo en este trabajo.

6.8.1. Fundamentos teóricos

En la actualidad existen un sin fin de publicaciones sobre un mismo tema e, incluso, sobre experimentos y ensayos prácticamente idénticos, pero realizados por diferentes grupos de investigación y/o en distintos momentos. Es más, aunque las condiciones sean, en principio, exactamente las mismas, se observa que en numerosas ocasiones dichas publicaciones se contradicen unas a otras. Por todo ello, parece necesario emplear metodologías de análisis que permitan extraer conclusiones globales a partir de toda la evidencia disponible sobre un mismo tema, desde ensayos clínicos hasta análisis transcriptómicos realizados en un laboratorio para evaluar la expresión diferencial entre diferentes condiciones. Así, en este punto es donde cobra importancia el *metaanálisis*. A diferencia de una revisión sobre una temática en particular, en la que simplemente se recopila información y se describe la evidencia existente y el estado del arte, el metaanálisis va más allá de decir cuantos estudios coinciden y cuantos discrepan entre ellos. En concreto, permite dar una conclusión global a partir de todos los datos disponibles que, en principio, deberá ser más fiable y cercana a la realidad al basarse en varios estudios y experimentos realizados por diferentes grupos de investigación en diferentes épocas y en condiciones, muy posiblemente, heterogéneas.

Por ejemplo, en el caso de la presente tesis se evalúa la expresión diferencial entre el fenotipo tumoral del cáncer de colon y el sano que, evidentemente, ha sido previamente estudiado por diferentes grupos de investigación. Por ello, actualmente existe un gran volumen de datos correspondientes a varios estudios transcriptómicos sobre este tema. En un primer vistazo, es obvio que muchas veces unos genes aparecerán diferencialmente expresados en unos estudios mientras que en otros no e, incluso, que algunos de ellos muestren sobreexpresión significativa en unos estudios e infraexpresión significativa en otros. Esta situación, que es más habitual de lo que cabría esperar, puede resultar realmente confusa y, en ocasiones, llevar a desconfiar de los resultados publicados.

En el caso de datos correspondientes a experimentos de microarrays, lo que se pretende en cada uno de ellos, de forma independiente, es obtener un efecto del cambio en la expresión de los genes. Así, mediante un metaanálisis será posible obtener un *efecto global* θ de la comparación de interés estimado a partir de los resultados independientes, denominados *ensayos identificados*. En el caso del análisis de expresión diferencial obtenido a partir del método *limma*, para cada gen el efecto se proporciona como *fold change*, que es el logaritmo en base dos del cociente de las expresiones de los tratamientos de las condiciones experimentales de interés. En este análisis, como

estadístico de contraste se proporciona el *t moderado*.⁶ A grandes rasgos, podría decirse que hay dos tipos de metaanálisis; por un lado, aquel que combina los efectos de cada ensayo identificado para obtener un efecto global, y otro que parte de los p valores de cada ensayo identificado, los combina y proporciona un p valor global. En concreto, Marot et al. (2009) comparó diferentes métodos de metaanálisis para analizar datos de microarrays, tanto de combinación del tamaño del efecto como de p valores. Por un lado, observó que los métodos de *combinación del tamaño del efecto moderado* da mejores resultados respecto de otros métodos existentes para estimar también el tamaño del efecto. Por otro lado, también observó que el método basado en la *combinación del p valor* proporciona una mejor sensibilidad y una mejor clasificación de los genes respecto de los otros métodos basados en el tamaño del efecto, puesto que estos últimos son más conservativos⁷.

En definitiva, el problema que se presenta en un metaanálisis se puede dividir en dos partes. En primer lugar, obtener un estimador centrado del mismo efecto de interés en cada uno de los N_s ensayos identificados, $\hat{\theta}_s$, independientes unos de otros. En segundo lugar, combinar los diferentes $\hat{\theta}_s$, un total de N_s , en un estimador global $\hat{\theta}$ o, en este caso, efecto global, con el fin de obtener un mejor estimador para el contraste de interés que, en cierto modo, resumirá o contendrá toda la información disponible en el metaanálisis. Sin embargo, cabe preguntarse cuál es la mejor manera de combinar los θ_s de cada uno de los ensayos identificados S para obtener el efecto global θ . La respuesta a esta pregunta es: realizar *una combinación lineal de los estimadores centrados del parámetro de interés de los diferentes ensayos identificativos*.

Para ejemplificarlo, se va a presentar un caso donde se mide la expresión diferencial de un gen g mediante microarrays en dos ensayos identificativos diferentes, 1 y 2, de tal manera que cualquier estimador definido como en la ecuación 6.7 será centrado:

$$\hat{\theta} = c \cdot \hat{\theta}_1 + (1 - c) \cdot \hat{\theta}_2 \quad (6.7)$$

donde c es el coeficiente de la combinación lineal el cual puede interpretarse como un peso ponderación que se otorga a cada estimador centrado de cada ensayo identificativo. A continuación, hay que encontrar los valores de c que hagan que el estimador centrado $\hat{\theta}$ sea el más preciso o el de menor varianza, por lo que sabiendo que la varianza del estimador global en función de los estimadores de cada ensayo identificativo es (ecuación 6.8):

⁶La interpretación del estadístico *t moderado* es igual que el proporcionado por un *test de t* salvo que el *error estandar* se ha “moderado” a través de todos los genes mediante un modelo *Bayesiano*, lo que proporciona una estimación conjunta de este error al tomar prestada información de todos los genes para ayudar en las inferencias y estimaciones sobre cada gen individual. Además, los grados de libertad empleados son más grandes, lo que evidencia la mayor fiabilidad de estas estimaciones. Este tipo de métodos, en los que se estudia, en este caso, un gen usando información de todos los genes evaluados, se conocen como *métodos de contracción*, y resultan más potentes al tener que estimar un menor número de parámetros y, por ende, incrementa la sensibilidad.

⁷En el contexto del contraste de hipótesis, un test es más conservativo que otro cuando le cuesta más rechazar la hipótesis nula; es decir, que si se estudia la expresión diferencial de un gen, cuanto más conservativo sea el test más le costará “decir” que dicho gen esta diferencialmente expresado. Por un lado, se cometerán menos falsos positivos pero, por otro lado, se identificarán menos genes como diferencialmente expresados de los que, en realidad, puedan existir.

$$\text{Var}(\hat{\theta}) = c^2 \cdot \text{Var}(\hat{\theta}_1) + (1 - c)^2 \cdot \text{Var}(\hat{\theta}_2) \quad (6.8)$$

se tiene que el estimador de máxima verosimilitud de θ , que se obtiene derivando la ecuación 6.8 respecto de los coeficientes c e igualando a 0, es (ecuación 6.9):

$$\frac{d\text{Var}(\hat{\theta})}{dc} = 2 \cdot c \cdot \text{Var}(\hat{\theta}_1) - 2 \cdot (1 - c) \cdot \text{Var}(\hat{\theta}_2) \quad (6.9)$$

y, despejando c (ecuación 6.10):

$$c = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2)} = \frac{1/\text{Var}(\hat{\theta}_1)}{1/\text{Var}(\hat{\theta}_1) + 1/\text{Var}(\hat{\theta}_2)} \quad (6.10)$$

Por lo tanto, de las ecuaciones 6.8 y 6.10 se tiene que, de forma general, el estimador centrado del efecto global más preciso (de menor varianza) se obtiene mediante la combinación lineal de los estimadores independientes del parámetro de interés (expresión diferencial del gen g , por ejemplo) de cada uno de los ensayos identificativos (publicaciones por diferentes grupos, 1 y 2 en el ejemplo), tal que los coeficientes de cada término son directamente proporcionales a la precisión (inverso de la varianza) de cada estimador, $w_s = 1/\text{Var}(\hat{\theta}_s)$, o sea (6.11):

$$\hat{\theta} = \frac{\sum_{s=1}^{N_s} w_s \cdot \hat{\theta}_s}{\sum_{s=1}^{N_s} w_s} \quad (6.11)$$

Por ello, este método se conoce como el *método de la inversa de la varianza*. En cuanto a la varianza de $\hat{\theta}$, $\text{Var}(\hat{\theta})$, se puede estimar mediante (ecuación 6.12):

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{\sum_{s=1}^{N_s} w_s} \quad (6.12)$$

Lo anterior establece cómo combinar los estimadores de los diferentes ensayos identificativos o estimadores independientes de cada estudio. Sin embargo, todavía queda hacer suposiciones sobre cómo se distribuyen estos o cómo es la población de la que han sido extraídos. Para tal fin, se definen dos modelos diferentes:

- Modelo de efectos fijos:** por un lado, puede suponerse que todos los estudios o experimentos son homogéneos en el sentido de que todos están afectados por los mismos factores y , por ende, sólo hay un tamaño del efecto verdadero θ_g . En consecuencia, todos los N_s estimadores del efecto de interés estiman el mismo efecto global θ_g a través de $\hat{\theta}_{sg}$ ya que todas las estimaciones del efecto provienen de la misma población homogénea. Por lo tanto, la única variabilidad que afecta al estimador $\hat{\theta}_{sg}$ es su varianza en el muestreo o intra-estudio,

$\text{Var}(\hat{\theta}_{sg})$, que, para abreviar, se llamará σ_{sg}^2 . Es decir, que si el tamaño muestral de cada estudio fuera infinito (situación totalmente imposible) entonces todos los estimadores $\hat{\theta}_{sg}$ serían igual al efecto verdadero θ y la varianza en el muestreo sería 0. En definitiva, el modelo de efectos fijos tendría la siguiente forma (ecuación 6.13):

$$\hat{\theta}_{sg} = \theta_g + \epsilon_{sg} \mid \epsilon_{sg} \sim N(0, \sigma_{sg}^2) \quad (6.13)$$

donde $\hat{\theta}_{sg}$ es la estimación del tamaño del efecto del gen g en el experimento s y σ_{sg}^2 la varianza de la distribución en el muestreo. En cuanto a las estimaciones del efecto global $\hat{\theta}_g$ y de su varianza $\hat{\text{Var}}(\hat{\theta}_g)$ se emplearán las ecuaciones 6.11 y 6.12, respectivamente, donde los pesos $w_{sg} = 1/\hat{\sigma}_{sg}^2$.

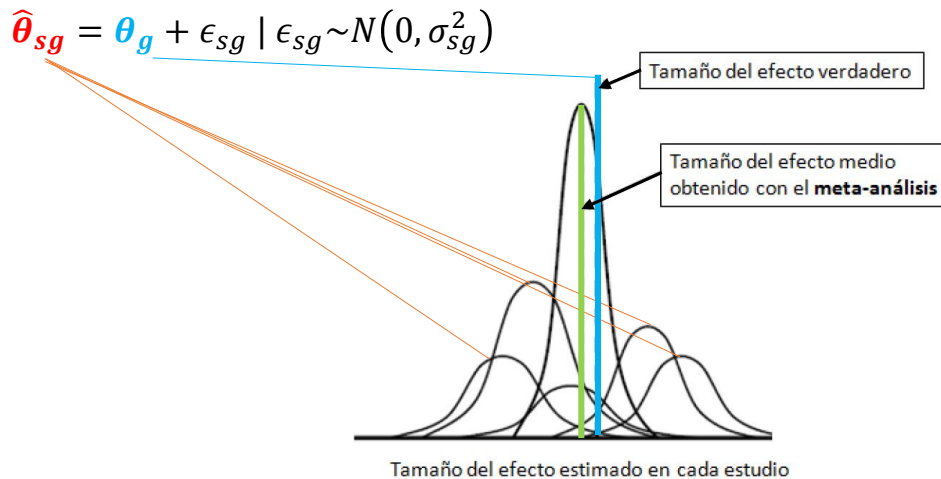


Figura 6.17: Modelo de efecto fijos. En el modelo de efectos fijos todos los estudios provienen de la misma población y el efecto estimado en cada uno de ellos estima el mismo efecto global verdadero

- Modelo de Efectos Aleatorios:** la suposición de homogeneidad de los diferentes experimentos no siempre es acertada, pues en muchas ocasiones las condiciones de dichos experimentos no son lo suficientemente similares como para que esta pueda aceptarse, como podría pasar, por ejemplo, si los pacientes de los que se extrae el ARN no tuvieran la misma edad en unos estudios que en otros, que en unos haya más hombres y en otros más mujeres, que en unos predomine una raza... En definitiva, pueden darse multitud de situaciones que hagan que los diferentes experimentos difieran lo suficiente como para que no puedan considerarse homogéneos y, por ende, el efecto verdadero estimado con cada estudio será diferente. Por lo tanto, habrá que suponer que cada uno de los N_s ensayos identificados es una muestra aleatoria simple de todos los posibles experimentos que pueden llevarse a cabo para obtener el efecto global de interés, θ_g . O sea, que cada estudio o

experimento con unas condiciones dadas estima un efecto correspondiente a un posible experimento de todos los que podrían contestar a una misma pregunta y, además, cada uno de los posibles experimentos estima un único valor verdadero θ_g .

Así, en la obtención de las estimaciones de cada uno de los experimentos podría pensarse como un proceso donde, en primer lugar, se obtiene una muestra aleatoria (efecto aleatorio del modelo) de tamaño N_s , $\{\theta_{1g}, \theta_{2g}, \dots, \theta_{sg}, \dots, \theta_{N_s g}\}$, de una población $N(\theta_g, \tau_g^2)$, donde τ_g^2 es la varianza entre los estudios, y, en segundo lugar, en cada experimento s se estima θ_{sg} a través del estimador $\hat{\theta}_{sg} \sim N(\theta_{sg}, \sigma_{sg}^2)$ donde esta segunda parte corresponde a un modelo de efectos fijos y, por ende, σ_{sg}^2 será la varianza en el muestreo del modelo de efectos fijos que, en este caso puede entenderse como la varianza dentro del estudio s (ecuación 6.14):

$$\theta_{sg} = \theta_g + v_{sg} \quad | \quad v_{sg} \sim N(0, \tau_g^2) \quad (6.14)$$

$$\hat{\theta}_{sg} = \theta_{sg} + \epsilon_{sg} \quad | \quad \epsilon_{sg} \sim N(0, \sigma_{sg}^2)$$

o (ecuación 6.15):

$$\hat{\theta}_{sg} = \theta_g + v_{sg} + \epsilon_{sg} \quad | \quad \epsilon_{sg} \sim N(0, \sigma_{sg}^2); v_{sg} \sim N(0, \tau_g^2) \quad (6.15)$$

siendo v_{sg} y ϵ_{sg} independientes y, además, v_{sg} no está ligando al experimento s , sino que en una nueva realización del experimento s el valor de v_{sg} será un nuevo valor observado de la variable aleatoria de distribución $v_{sg} \sim N(0, \tau_g^2)$. Para el cálculo de τ_g , en primer lugar se calcula Q (6.16):

$$Q_g = \sum_{s=1}^{N_s} w_s (\hat{\theta}_{sg} - \hat{\theta}_g)^2 \quad (6.16)$$

y, a continuación, se tiene que τ_g^2 es (ecuación 6.17):

$$\tau_g^2 = \max \left\{ 0, \frac{Q_g - N_s - 1}{\sum_{s=1}^{N_s} w_s - \left(\sum_{s=1}^{N_s} w_s^2 \right) / \sum_{s=1}^{N_s} w_s} \right\} \quad (6.17)$$

Finalmente, puede comprobarse que la media, o el valor esperado, de $E[\hat{\theta}_{sg}] = \theta_g$ y que la varianza $\text{Var}(\hat{\theta}_{sg}) = \sigma_{sg}^2 + \tau_g^2$; por ello, los pesos de cada $\hat{\theta}_{sg}$ pasan a ser $w_s = 1/\text{Var}(\hat{\theta}_{sg}) = 1/(\sigma_{sg}^2 + \tau_g^2)$. Cabe destacar que el modelo de efectos fijos (ecuación 6.13) podría ser un caso especial del modelo de efectos aleatorios cuando $\tau = 0$.

Otra forma de visualizar el modelo de efecto fijos es pensar que al azar (en realidad según la distribución $N(\theta_g, \tau_g^2)$ que es común a todos los N_s experimentos obtenidos), se realizase un experimento s de todos los infinitos posibles experimentos diferentes que pueden llevarse a cabo para estimar el efecto de interés, de tal manera que se ha seleccionado un experimento s cuyo efecto verdadero de interés es θ_{sg} . Así, ahora se tiene una población diferente, $N(\theta_{sg}, \sigma_{sg}^2)$, correspondiente al experimento s , que tiene una media diferente a θ_g , sobre la que se realiza un muestreo aleatorio y se estima el efecto $\hat{\theta}_{sg} \sim N(\theta_{sg}, \sigma_{sg}^2)$. En los efectos fijos, directamente la población sobre la que se hace el muestreo aleatorio es $N(\theta_g, \sigma_{sg}^2)$ ya que no hay una población de infinitos posibles experimentos diferentes.

A continuación, se van a presentar los dos métodos de metaanálisis empleados en la presente tesis: combinación de efectos y combinación de p valores. Por un lado, puesto que el metaanálisis de la expresión diferencial se realiza a partir de diferentes ensayos identificados o experimentos independientes cada uno de los cuales ha sido analizados con el método limma de forma independiente, se explica cómo realizarlo a partir de los efectos estimados mediante limma (Marot et al. (2009)). Por otro lado, puesto que también se va a realizar el metaanálisis del análisis de enriquecimiento (análisis de conjuntos de genes) a partir de los análisis obtenidos independientemente en cada experimento mediante dos métodos diferentes (ORA y CAMERA), se aplicará la combinación de los p valores, para cada gen, obtenidos de forma independiente en cada uno de los experimentos reanalizados y en el experimento *IBGM*.

6.8.1.1. Combinación de los efectos

Para la estimación del tamaño del efecto moderado de cada uno de los N_s experimentos se hace, para limma, mediante (ecuación 6.18):

$$d_{\text{limma},sg} = \frac{t_{\text{limma},sg}}{\sqrt{\tilde{n}_{sg}}} \quad (6.18)$$

donde $t_{\text{limma},sg}$ es el estadístico estimado para cada gen y cada estudio mediante el método limma y \tilde{n} (ecuación 6.19):

$$\tilde{n}_{sg} = \frac{n_{i,s}n_{j,s}}{n_{i,s} + n_{j,s}} \quad (6.19)$$

Además, para aplicar el metaanálisis, la varianza del tamaño del efecto moderado es, usando la distribución de los tamaños de los efectos proporcionada por Hedges (1981) (ecuación 6.20):

$$\text{Var}(d_{sg}) = \frac{m_{sg}}{(m_{sg} - 2)\tilde{n}_{sg}} [1 + \tilde{n}_{sg}\delta_{sg}^2] - \frac{\delta_{sg}^2}{[c(m_{sg})]^2} \quad (6.20)$$

donde m_{sg} es el número de grados de libertad, que en limma es la suma de $g.d.l.$ a priori más los de los residuales del modelo lineal del gen g , y δ es el tamaño del efecto (ecuación 6.21):

$$\delta_{sg} = d_{limma,sg} \quad (6.21)$$

y

$$c(m_{sg}) = \frac{\Gamma\left(\frac{m_{sg}}{2}\right)}{\sqrt{\frac{m_{sg}}{2}} \Gamma\left(\frac{m_{sg}-1}{2}\right)} \quad (6.22)$$

De nuevo, basándose en la distribución de los tamaños de los efectos proporcionada por Hedges (1981), el estimador no sesgado será (ecuación 6.23):

$$d'_{limma,sg} = c(m_{sg})d_{limma,sg} \quad (6.23)$$

y su varianza (ecuación 6.24)

$$\text{Var}(d'_{sg}) = c(m_{sg})^2 \left(\frac{m_{sg}}{(m_{sg}-2)\tilde{n}_{sg}} [1 + \tilde{n}_{sg}(d')^2] - \frac{(d')^2}{[c(m_{sg})]^2} \right) \quad (6.24)$$

siendo ahora la varianza de los tamaños de los efectos insesgados $c(m_{sg})^2$ veces la varianza de los estimadores sesgados, y sustituyendo d_{sg} por d'_{sg} . Por lo tanto, se tiene que $\hat{\theta}_{sg}$ los d'_{sg} , como σ_{sg}^2 a $\text{Var}(d'_{sg})$. No obstante, puesto que se aplica un modelo de efectos aleatorios, τ_g^2 se calcula tal y como se muestra en las ecuaciones ecuación 6.16 y ecuación 6.17, por lo que la varianza final del estimador del efecto para el gen g en el estudio s es (ecuación 6.25):

$$\text{Var}(\hat{\theta}_{sg}) = \text{Var}(d'_{sg}) + \tau_g^2 \quad (6.25)$$

6.8.1.2. Combinación de p valores

Para combinar las probabilidades (p valores) de los diferentes estudios suele emplearse el método de Fisher. No obstante, aplicar este método tiene como consecuencia la pérdida de la direccionalidad del efecto o del cambio de expresión y habría que separar genes sobreexpresados e infraexpresados para aplicarlo por separado sobre cada uno de ellos. Por ello, una estrategia seguida que evita este problema es usar el método de la normal inversa que consiste en obtener el promedio de los p valores previamente transformados a *puntajes normales* y para evitar los conflictos direccionales se emplean los p valores de una cola de cada estudio $\tilde{p}_g(s)$ (ecuación 6.26).

$$S_g = \frac{1}{\sqrt{N_s}} \sum_{s=1}^{N_s} \phi^{-1}(\tilde{p}_g(s)) \quad (6.26)$$

Así, bajo la hipótesis nula, S_g sigue una distribución normal estándar. El procedimiento sería obtener los valores de los estadísticos, en este caso los t asociados al método limma, y obtener los p valores de una cola, por lo que la dirección se mantiene. Es decir, que valores de t bajos y/o negativos serán p valores bajos y viceversa. Tras obtener S_g , el p valor asociado es (ecuación 6.27):

$$p_g = 2 \cdot (1 - \phi(|S_g|)) \quad (6.27)$$

Se emplea una modificación o alternativa propuesta por Marot et al. (2009), que proporciona unos pesos a cada p valor individual en función del número de réplicas del experimento correspondiente a dicho p valor (ecuación 6.28):

$$S_g = \sum_{s=1}^{N_s} w_s \phi^{-1}(1 - \tilde{p}_g(s)) \quad (6.28)$$

donde los pesos son (ecuación 6.29)

$$w_s = \sqrt{\frac{n_s}{\sum_{i=1}^{N_s} n(i)}} \quad (6.29)$$

siendo $n(s)$ el número total de réplicas en el estudio s . Tras obtener los valores de t moderada del test con limma (t_{limma}), se calculan los p valores de una cola y se realiza el procedimiento explicado en este apartado. DE igual modo, para el metaanálisis de enriquecimiento, se obtienen los p valores de una cola y se realiza este mismo experimento. Por último, en ambos casos, se controla la tasa de error FDR corrigiendo los p valores mediante el método propuesto por Benjamini and Hochberg (1995).

6.8.2. Procedimiento y diseño experimental

Como se comentó en el apartado “5.1.2 Modelos de tejido humano,” aparte de analizar, mediante microarrays, muestras procesadas por nuestro propio grupo de investigación, también se extrajeron datos de experimentos transcriptómicos del repositorio NCBI GEO. Cada uno de los experimentos transcriptómicos extraídos se reanalizaron a partir de los archivos *.CEL* de igual modo que el resto de muestras procesadas por nuestro grupo. Por lo tanto, cada uno de los experimentos son los ensayos identificados del metaanálisis. Posteriormente, los resultados obtenidos para cada uno de estos experimentos fueron empleados para llevar a cabo el metaanálisis por los dos métodos anteriormente comentados.

Parte IV

Resultados

“-Winwood Reade ha dicho cosas muy interesantes sobre el tema-, dijo Holmes-. Asegura que el hombre tomado individualmente es un enigma irresoluble pero en conjunto es una certidumbre matemática. No puedes, por ejemplo, predecir lo que hará un hombre determinado pero sí prever con precisión lo que hará la mayoría de ellos. Los individuos varían, los porcentajes permanecen constantes. Eso dice la estadística”.

Ser Arthur Conan Doyle (1890)

Capítulo 7

Resultados

En este capítulo se van a exponer los resultados obtenidos en la presente tesis. En primer lugar, se presentan los resultados correspondientes a los experimentos realizados con líneas celulares. En segundo lugar, los resultados de los experimentos transcriptómicos realizados en muestras de tejido humano de cáncer de colon y normal, y, finalmente, el metaanálisis.

7.1. Remodelado de la homeostasis del Ca^{2+} intracelular en líneas celulares de cáncer de colon

Para estudiar el remodelado de la homeostasis del Ca^{2+} intracelular en el cáncer de colon (CCR), se ha realizado una primera aproximación a través del estudio de dos líneas celulares que son bien conocidas por nuestro grupo, pues durante años se ha caracterizado y usado en diferentes experimentos, además de ser modelos celulares aceptados en el estudio del CCR: HT29 (CCR) y NCM460 (normales). Por ello, en primer lugar, se ha comprobado que el comportamiento de estas células es el esperado según la evidencia ya existente, para lo que se han realizado diferentes experimentos funcionales de imagen de Ca^{2+} . En segundo lugar, se presentan los resultados correspondientes a los análisis transcriptómicos, para lo que se han empleado tres tecnologías diferentes: IonTorrent e Illumina, ambas basadas en *RNA-seq*, y los microarrays Clariom D Human de Affymetrix. A partir de los datos generados, se estudió la expresión diferencial de todos los genes analizados¹ y, de entre todos ellos, se centró el interés en un subconjunto de genes relacionados con el transporte de Ca^{2+} intracelular. Posteriormente, a partir de la lista de todos los genes diferencialmente expresados (DEG) se investigó qué conjuntos de genes podrían estar relacionados con el fenotipo tumoral o cuales de ellos contenían DEG. Por último, se estudió el efecto del tratamiento con DFMO, que inhibe la síntesis de poliaminas, sobre cada una de estas

¹Salvo que se diseñe previamente un panel de genes de interés, las tecnologías transcriptómicas como microarrays o *RNA-seq* proporcionan información sobre decenas de miles de genes en función de las características de cada plataforma (tipo de microarray, casa comercial...).

líneas celulares. Por un lado, se evaluó el efecto a nivel funcional mediante imagen de Ca^{2+} para estudiar las posibles alteraciones en los depósitos intracelulares y SOCE. Por otro lado, se examinó el transcriptoma de las células tratadas y sin tratar de ambas líneas celulares con el fin de identificar qué genes se ven alterados y si son o no los mismos en ambos modelos. Además, también se evaluó el comportamiento en la expresión de los genes implicados en el transporte de Ca^{2+} y qué rutas de señalización están relacionadas con el tratamiento para cada una de las líneas celulares. A continuación, se describen los resultados obtenidos en cada uno de los experimentos anteriormente expuestos.

7.1.1. Estudio funcional del remodelado del Ca^{2+} intracelular mediante imagen de Ca^{2+} en líneas celulares

En primer lugar, se realizaron algunos experimentos confirmatorios de algunos de los resultados previos obtenidos por nuestro grupo de investigación con el fin de corroborar que los modelos celulares que se han empleado en la tesis (HT29 como modelo de cáncer de colon y NCM460 como modelo de colonocitos normales) se comportan según lo esperado. En concreto, se evaluó tanto el tamaño de los depósitos de Ca^{2+} como SOCE de modelos celulares de cáncer de colon respecto de normales. Para ello, se ha empleado la técnica de imagen de Ca^{2+} y la sonda de Ca^{2+} citosólico Fura2/AM. Por un lado, el tamaño de los depósitos se evaluó a partir de la liberación del Ca^{2+} desde el RE al citosol causada por el inhibidor reversible de la SERCA, CPA. Por otro lado, cuando los depósitos intracelulares de Ca^{2+} se han vaciado, se reveló SOCE mediante la perfusión de medio con Ca^{2+} 1 mM. El diseño experimental² y el análisis de datos se explican en los apartados “6.3.2. Procedimiento y diseño experimental para el análisis funcional del efecto del DFMO en los modelos celulares HT29 y NCM460” y “6.3.3. Análisis de los datos de experimentos de imagen de Ca^{2+} ”, respectivamente. En cuanto a los resultados correspondientes al análisis de datos (modelización de la señal, diagnóstico del modelo, contrastes de hipótesis, estimaciones, métodos *bootstrapping*...) se presentan en los diferentes subapartados del apéndice “C. Análisis funcional del efecto del DFMO sobre la homeostasis del Ca^{2+} intracelular” .

²Los resultados mostrados en este apartado, en realidad, provienen de un experimento en el que no sólo se estudian las diferencias entre las líneas celulares de cáncer de colon (HT29) y normales (NCM460), sino que también se estudia el efecto del DFMO sobre estas mismas líneas, los cuales se mostrarán en el apartado “7.1.3.1. Estudio del efecto del DFMO sobre el Ca^{2+} citosólico, el tamaño de los depósitos de Ca^{2+} intracelular y SOCE” . Brevemente, el diseño experimental completo consiste en cuatro condiciones experimentales generadas por los factores fijos *Línea* = {HT29, NCM460} y *Tratamiento* = {Control, DFMO}, de cada una de las cuales hay 8 réplicas realizadas en 4 días diferentes -dos cada día-. En total, hay 32 réplicas en las que se han registrado más de 1000 células.

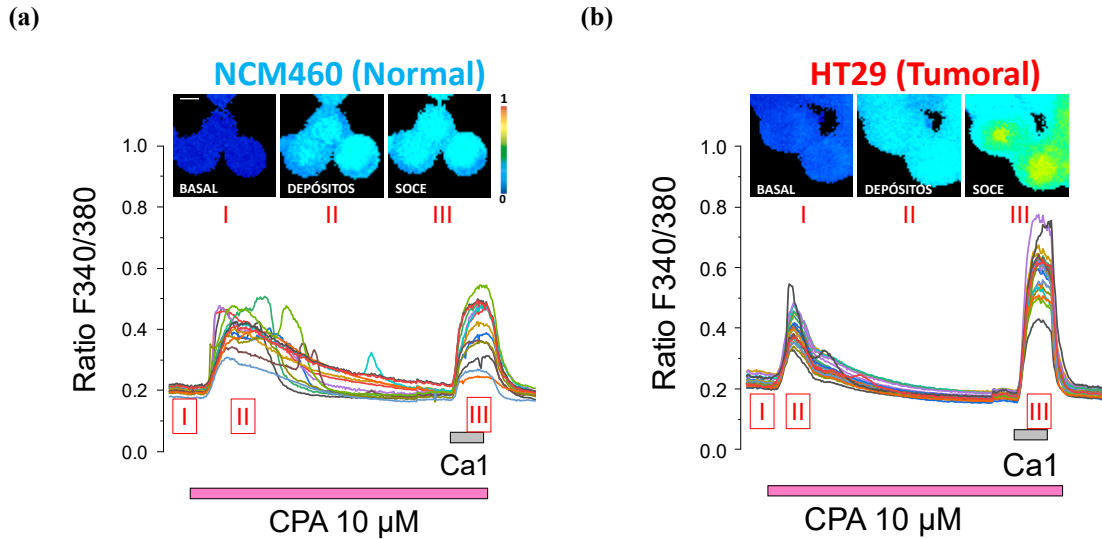


Figura 7.1: Registros en célula única del tamaño de los depósitos intracelulares de Ca^{2+} y SOCE en las líneas celulares HT29 y NCM460. Se muestran registros representativos de imagen de Ca^{2+} intracelular en célula única para cada una de las condiciones experimentales (a) células NCM460 y (b) HT29 Control. Por un lado, en la parte superior de cada figura se muestra la imagen, en escala de pseudocolor, captada por la cámara para la condición basal (I), liberación del Ca^{2+} de los depósitos (II) y SOCE (III). Por otro lado, en la parte inferior se representa el ratio de fluorescencia 340/380 en función del tiempo y los estímulos empleados. En total, se han obtenido 8 registros para cada condición experimental los cuales han sido realizados en tandas de dos réplicas de cada una de las diferentes condiciones experimentales cada día a lo largo de cuatro días diferentes espaciados en el tiempo una semana mínimo. El número total de células en los 8 registros, para cada condición experimental, ha sido 272 para NCM460 Control y 230 para HT29 Control. En estos experimentos se ha empleado la sonda de Ca^{2+} ratiométrica Fura2/AM para detectar el Ca^{2+} presente en el citosol y se ha registrado la fluorescencia a 340 y 380 nm. En particular, se estudió el tamaño de los depósitos de Ca^{2+} a través de la liberación del Ca^{2+} al citosol con CPA 10 μM , inhibidor reversible de la SERCA, en medio libre de Ca^{2+} y, posteriormente, en el mismo registro, se evaluó SOCE a través de la perfusión de medio con Ca^{2+} 1 mM. La escala de pseudocolor a la derecha de las imágenes de la primera figura muestra el código de colores para el ratio de las fluorescencias que va desde el 0, azul, al 1, rojo. La barra blanca representa 10 μm .

A partir de estos registros, se han modelizado, de forma independiente, cada una de las cinco variables respuesta de interés: $[\text{Ca}^{2+}]_{\text{cit}} \text{ basal}$ (*CaBasal*), *Área Bajo la Curva del Ratio de Fluorescencia 340/380* (*AUC*) de la señal correspondiente a los depósitos (*AreaDepositos*), *Máximo Incremento del Ratio de Fluorescencia 340/380* de la señal (ΔMax) correspondiente a los depósitos (*MaxDepositos*), *AUC* de la señal correspondiente a SOCE (*AreaSOCE*) y (ΔMax) de la señal correspondiente a SOCE (*MaxSOCE*). Para ello, tal y como se explica en el apartado “6.3.3. Análisis de los datos de experimentos de imagen de Ca^{2+} ”, se han empleado modelos lineales mixtos, puesto que permite analizar datos con estructura agrupada (en este caso por las células agrupadas dentro de un mismo cubreobjetos o cristal que, a su vez, se agrupan dentro de un mismo día en los que se realizan dos réplicas de cada condición experimental. Es decir, que no sólo se consideran las variables explicativas de interés (factor *Línea* celular, en este caso), sino también los efectos aleatorios debidos a los factores aleatorios *Cristal* y *Día*. Los resultados correspondientes al análisis de datos de este experimento (modelización de la señal, diagnóstico del modelo, contrastes de hipótesis, estimaciones, métodos *bootstrapping*...), para cada una de las señales estudiadas, se presentan en los diferentes subapartados del apéndice “C. Análisis funcional

del efecto del DFMO sobre la homeostasis del Ca^{2+} intracelular”. En el apéndice “C.1. Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos” se presenta un ejemplo simplificado de aplicación con lineales mixtos para este tipo de experimentos: seleccionar el mejor modelo, evaluar si es o no adecuado el modelo y, en caso de que no lo sean -no cumplan la normalidad, la homocedasticidad...-, usar técnicas más complejas de remuestreo paramétrico como *bootstrapping*...

Tras ajustar los diferentes modelos lineales mixtos clásicos y *bootstrap* para cada una de las 4 variables respuesta, se llevaron a cabo los contrastes de hipótesis correspondientes para evaluar cuales de las variables respuestas estudiadas son o no diferentes entre las dos líneas celulares (ver apéndice “C. Análisis funcional del efecto del DFMO sobre la homeostasis del Ca^{2+} intracelular”). Así, al igual que en los estudios los realizados por Sobradillo et al. (2014), se observó que la línea celular de cáncer de colon, respecto de las normales, presenta una mayor $[Ca^{2+}]_{cit}$ basal (figura 7.2, un menor tamaño de los depósitos (figura 7.3a,b) y una mayor SOCE (figura 7.3c,d), tanto en términos de incremento máximo como de AUC. Por lo tanto, estos resultados, aparte de corroborar la evidencia existente, contribuyen, por un lado, a apoyar la evidencia ya existente y, por otro lado, a cerciorarse de que las líneas celulares empleadas en los diferentes experimentos a lo largo de la presente tesis doctoral se comportan adecuadamente.

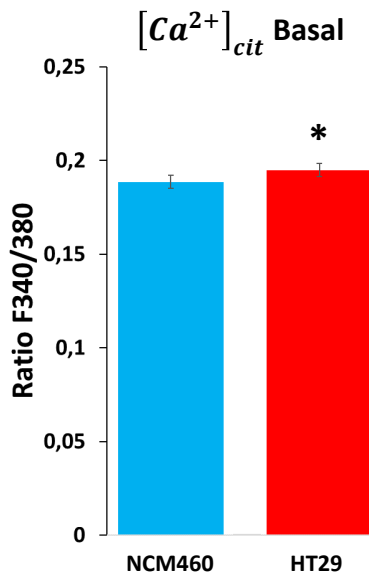


Figura 7.2: Estimaciones del efecto del DFMO sobre la $[Ca^{2+}]_{cit}$ basal. Los diagramas de barras representan los valores esperados (estimados mediante modelos lineales mixtos) de $[Ca^{2+}]_{cit}$ basal (mediana de la señal durante el primer minuto en medio Ca^{2+} 1 mM) para cada una de las dos líneas celulares y las barras su error \pm SEM (estimados mediante modelos lineales mixtos). El asterisco indica que se observan diferencias significativas entre las dos líneas celulares para un nivel de significación inferior a 0.05 en la tabla ANOVA o que el criterio de AIC sugiere que el factor *Línea Celular* ha de estar en el modelo, tal y como se detalla en el apéndice ”C. Análisis funcional del efecto del DFMO sobre la homeostasis del Ca^{2+} intracelular”. El número de células registradas, en un total de 16 unidades experimentales, han sido 272 células para NCM460 y 230 para HT29.

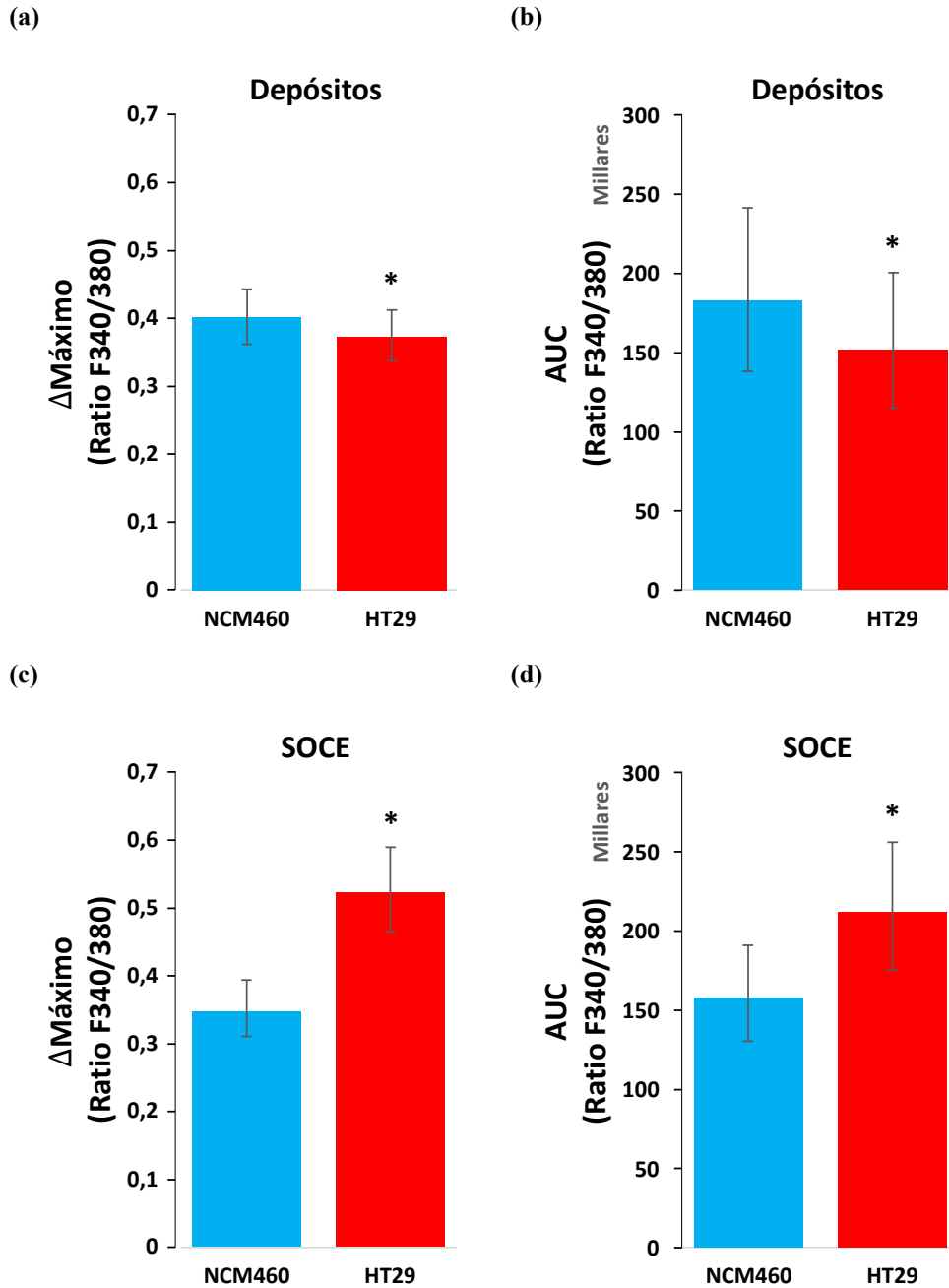


Figura 7.3: Estimación del tamaño de los depósitos de Ca^{2+} y SOCE para las líneas celulares NCM460 y HT29. Los diagramas de barras representan los valores esperados del tamaño de los depósitos y de SOCE para cada una de las condiciones experimentales evaluadas y las barras su error \pm SEM (estimados mediante modelos lineales mixtos). En la parte superior se representan el Δ máximo **b)** y el AUC **a)** de la señal correspondiente al tamaño de los depósitos de Ca^{2+} . En la parte inferior se representan el Δ máximo **d)** y el AUC **c)** de la señal correspondiente a SOCE. El asterisco indica que se observan diferencias significativas entre las dos líneas celulares para un nivel de significación inferior a 0.05 o que el criterio de AIC sugiere que el factor *Línea Celular* ha de estar en el modelo (ver apéndice C). El número de células registradas, en un total de 16 unidades experimentales, han sido 272 células para NCM460 y 230 para HT29.

7.1.2. Análisis transcriptómico de las líneas celulares de cáncer de colon HT29 y de colon normal NCM460

El primer paso en la investigación fue estudiar el transcriptoma en un modelo celular de CCR, HT29, y compararlo con el transcriptoma de un modelo celular de colon normal, NCM460 para contestar diferentes preguntas tales como ¿Cuántos genes se están expresando y cuántos de ellos están diferencialmente expresados? ¿Qué patrón de expresión muestran los genes relacionados con la homeostasis del Ca^{2+} intracelular? o ¿Qué conjuntos de genes están relacionados con el fenotipo de CCR? Para ello, la metodología empleada consistió en tres técnicas genómicas de alto rendimiento: dos basadas en *RNA-seq* -IonTorrent e Illumina- y otra en *microarrays* -Clariom D Human de Affymetrix-. Por ello, tras la extracción del ARN realizada siguiendo las pautas recomendadas por cada una de las casas comerciales; es decir, que para cada una de los experimentos correspondientes a cada una de las tecnologías se emplearon cultivos celulares diferentes y, además, se realizaron en momentos temporales muy diferentes -a lo largo de 4 años-.

A pesar de que a nivel general se comparan dos condiciones experimentales correspondiente a cada una de las líneas celulares, por lo que hay un único factor fijo en el diseño experimental, *Línea Celular*, existen ciertas diferencias entre los experimentos llevados a cabo con cada una de las técnicas genómicas empleadas y que, por ello, es necesario comentar antes de mostrar los resultados. En primer lugar, los datos generados por IonTorrent sólo contienen información sobre un panel de 80 genes de interés previamente diseñado por nuestro grupo; en segundo lugar, los experimentos basados en *microarrays*, a pesar de que en este apartado sólo se evalúan las diferencias entre las condiciones experimentales dadas por las líneas celulares HT29 y NCM460, en realidad provienen de un diseño experimental más complejo en el que además hay otras dos condiciones experimentales correspondientes a cada una de las líneas celulares tratadas con DFMO y cuyos resultados se mostrarán en el apartado “7.1.3.2. Análisis transcriptómico de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO”. Además, los tamaños muestrales también son diferentes para cada uno de los experimentos, pero todos son balanceados³, pues para el experimento realizado con IonTorrent $N_{\text{HT29}} = N_{\text{NCM460}} = 4$, mientras que tanto para Illumina como para los *microarrays* $N_{\text{HT29}} = N_{\text{NCM460}} = 3$.

En cuanto al análisis de los datos, mientras que los datos generados por los experimentos realizados con *RNA-seq* fueron analizados a partir de las matrices de expresión sin normalizar proporcionadas por la empresa encargada de secuenciar las muestras, los correspondientes al experimento basado en *microarrays* fue analizado desde cero a partir de los datos crudos sin procesar; es decir, a partir de los archivos *.CEL* proporcionados por la empresa. Por lo tanto, el tratamiento de los datos fue diferente para cada uno de los dos tipos de experimentos. Por un lado, los datos del experimento con *microarrays* fueron preprocesados, por lo que tras ser “leídos” por el *software R* se sometieron un control de calidad para identificar posibles *outliers* o *microarrays* -muestras- problemáticos y, en tal caso, eliminarlos del análisis. Para ello, se aplicaron métodos tanto generales como específicos para *microarrays* de Affymetrix. Afortunadamente, no fue

³Cada una de las condiciones experimentales tiene el mismo número de unidades experimentales -muestras-.

necesario eliminar ninguna de ellas. A continuación, se corrigió el ruido de fondo, se normalizaron y se resumieron los valores de los conjuntos de sondas mediante el método RMA (ver apéndice “B.3.1. Control de calidad y preprocesado de los datos transcriptómicos de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO”). Así, se obtuvo la matriz de expresión del experimento realizado con microarrays.

Por otro lado, los datos de expresión de los experimentos basados en *RNA-seq* también fueron sometidos a un control de calidad con métodos generales (ACP, clúster jerárquico...) y, de nuevo, ninguna muestra tuvo que ser eliminada del análisis. Posteriormente, fueron normalizados mediante el método TMM (Robinson and Oshlack (2010)). No obstante, puesto que los datos de expresión generados por *RNA-seq* son variables respuesta discretas y que se va a emplear el método *limma* para el análisis de expresión diferencial, es necesario prepararlos para ello. En concreto, se aplica el método *VOOM* (Law et al. (2014)) que realiza una transformación logarítmica en base 2 de los datos de conteo (lecturas por millón) y además, estima la relación entre la media y la varianza que será usada en el análisis de expresión diferencial posterior (ver apéndice “B.1.1. Control de calidad y preprocesado de los datos transcriptómicos de los experimentos basados en RNA-seq para la comparación entre las líneas celulares HT29 y NCM460”).

A partir de este punto, el análisis llevado a cabo para cada una de las tres matrices de expresión generadas para cada uno de los experimentos es prácticamente idéntico. Así, en primer lugar, se realizó un filtrado con el fin de eliminar de la matriz de expresión aquellos genes que o bien no se expresan o bien no muestran una variación suficiente a lo largo de las diferentes muestras -sin tener en cuenta las posibles covariables-. Posteriormente, sobre esta nueva matriz de expresión, se ajustó un modelo lineal mediante el método *limma* y se llevó a cabo un análisis de expresión diferencial sobre todos los genes analizados con cada una de las diferentes tecnologías y, posteriormente, se realizó un análisis de significación biológica. En concreto, el análisis de significación biológica consistió, por un lado, en analizar los perfiles de expresión y, por otro lado, en realizar un análisis de enriquecimiento de genes, este último sólo con los datos de Illumina y de microarrays, puesto que son los que contienen datos sobre el transcriptoma completo.

Puesto que los datos generados por las tres tecnologías diferentes son complementarios y con el fin de no abrumar al lector con la presentación de un análisis tras otro y de no tener que realizar saltos entre páginas para compararlos, a continuación se mostrarán de forma conjunta los resultados obtenidos con todas ellas. Además, todos los datos y análisis correspondientes al preprocesado de los datos se muestran en los apéndices “B.1.1. Control de calidad y preprocesado de los datos transcriptómicos de los experimentos basados en RNA-seq para la comparación entre las líneas celulares HT29 y NCM460” y “7.1.3.2. Análisis transcriptómico de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO”.

7.1.2.1. Análisis de expresión diferencial entre las líneas celulares HT29 y NCM460

A partir de las tres matrices de expresión, cada una con un número diferente de registros de partida (filas, sondas -microarrays-, identificadores ENSEMBL -RNA-seq-...), se realizó un filtrado no específico en el que se eliminaron todos aquellos innecesarios (sin correspondencia con identificador ALIAS⁴, que no muestren expresión...) y, posteriormente, un análisis de expresión diferencial en el que se consideró como DEG a aquellos genes cuyo FDR < 0.05⁵ (tabla 7.1).

Tabla 7.1: Resumen del filtrado y de los DEG para cada experimento.

| Contraste | Genes o Sondas ¹ | | |
|-----------------|-----------------------------|----------------------------|------------------|
| | Total | Post-filtrado ² | DEG ² |
| IonTorrent | 77 | 69 | 36 |
| Illumina | 65219 | 16375 | 14511 |
| Clariom D Human | 138745 | 12973 | 6623 |

Abreviaturas:

DEG: genes (sondas) diferencialmente expresados.

Notas:

¹ En este contexto, *Genes* se refiere a los registros de cada experimento; es decir genes analizados con identificador ENSEMBL para *RNA-seq* y sondas para microarrays.

² Registros únicos con correspondencia ALIAS.

No obstante, de todos los DEG identificados en cada uno de los experimentos (tabla 7.1), el análisis gen a gen se centró únicamente en los genes relacionados con la homeostasis del Ca²⁺. Así, a continuación, dividido en las diferentes familias en las que se han agrupado previamente estos genes, se muestran sus datos de expresión diferencial y si esta es significativa para un FDR < 0.05, por lo que se analizarán por separado los resultados para cada una de ellas. Cabe destacar que se ha considerado a un gen como diferencialmente expresado lo es, al menos, con dos de las tres tecnologías y en el mismo sentido para un FDR < 0.05, que para una de ellas sea al 0.05 y otra al 0.1, o que lo sea con un FDR < 0.1 para dos de ellas y que la restante muestre un logFC en el mismo sentido que las otras dos. También, si con una tecnología un gen resulta diferencialmente expresado y, además, con las otras dos tecnologías se observa que el signo del logFC es el mismo, entonces también se considerará que el gen está diferencialmente expresado. Además, también se considerará que un gen está diferencialmente expresado si lo es únicamente con los micorarrays por una serie de razones: el transcriptoma de las muestras empleadas en los experimentos de imagen de Ca²⁺ se ha analizado mediante estos microarrays; el transcriptoma de los experimentos posteriores para estudiar el efecto del tratamiento con DFMO en estas mismas líneas celulares se ha estudiado

⁴Son los diferentes nombres alternativos que se han usado y/o se usan para referirse a un gen en diferentes fuentes bibliográficas, bases de datos...

⁵En el análisis de expresión diferencial se contrasta la hipótesis nula H₀ : para el gen g no existe expresión diferencial, y se rechaza para un FDR < 0.05, donde el FDR es el p-valor ajustado para las comparaciones múltiples, en el que se controla la tasa de falsos positivos al 5%; es decir, que de todos los genes que rechazan la H₀ se espera un 5% de falsos positivos.

con estos mismos microarrays, y el transcriptoma de las muestras de tejido humano pareadas de cáncer de colon y normal también se ha analizado con estos mismos microarrays. No obstante, en aquellas ocasiones en las que se observen contradicciones se tendrá precaución con la conclusión.

En la figura 7.4 se observa el análisis transcriptómico para los 10 genes de la familia de los canales de Ca^{2+} operados por voltaje (VOCC). Por un lado, con la tecnología IonTorrent se identifica la sobreexpresión en células tumorales únicamente del *CAVI.3*, mientras que se detecta la infraexpresión de *CAVI.2*, *CAV2.1*, *CAV2.2*, *CAV2.3* y *CAV3.2*. A diferencia de lo observado con IonTorrent, con la tecnología Illumina no se detecta expresión diferencial ni de *CAV2.1* ni de *CAV2.2*, pero sí sobreexpresión de *CAV3.1*. Por otro lado, el experimento basado en microarrays únicamente detecta la sobreexpresión de *CAVI.3* y la infraexpresión de *CAVI.2*, que también han sido identificadas mediante las dos tecnologías de *RNA-seq*. Por lo tanto, se puede concluir que en HT29, respecto de las NCM460, hay una mayor expresión de *CAVI.3* y una menor de *CAVI.2*, *CAV2.3* y *CAV3.2*.

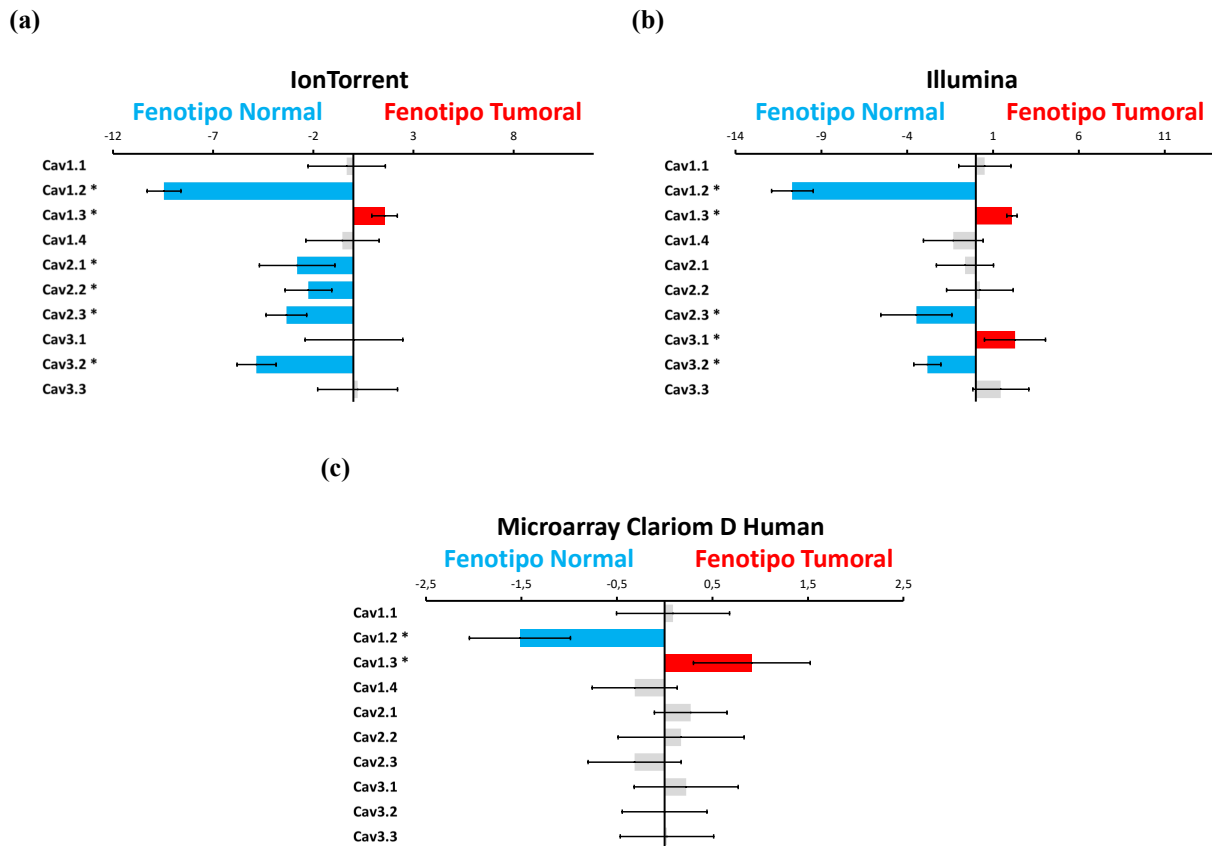


Figura 7.4: Análisis de la expresión diferencial entre las células HT29 (fenotipo tumoral) y NCM460 (fenotipo normal) para los genes de la familia VOCC. Los valores de expresión diferencial, en concreto los valores $\log_2(\text{FC})$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($\text{FDR} < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $\text{FDR} < 0.05$ y "." $\text{FDR} < 0.1$. La expresión diferencial se ha realizado con tres tecnologías transcriptómicas diferentes: (a) IonTorrent, (b) Illumina y (c) microarray Clariom D Human de Affymetrix.

En cuanto a los genes de la familia SOCE, con IonTorrent se ha identificado una mayor expresión de *STIM1* y *ORAI2* en las células HT29 respecto de las células NCM460, mientras que *CRACR2A* lo está menos. Con Illumina, además, también se ha detectado mayor expresión de *STIM2* en HT29. A diferencia de los anteriores, en el experimento basado en microarrays no se detecta una mayor expresión de *ORAI2* ni de *STIM2*, pero sí la menor expresión de *ORAI1* en HT29 respecto de las NCM460 (figura 7.5). Poniendo todos estos resultados en común, se puede concluir que en las HT29 hay una mayor expresión de *ORAI2* y *STIM1* y una menor de *ORAI1* y *CRACR2A*.

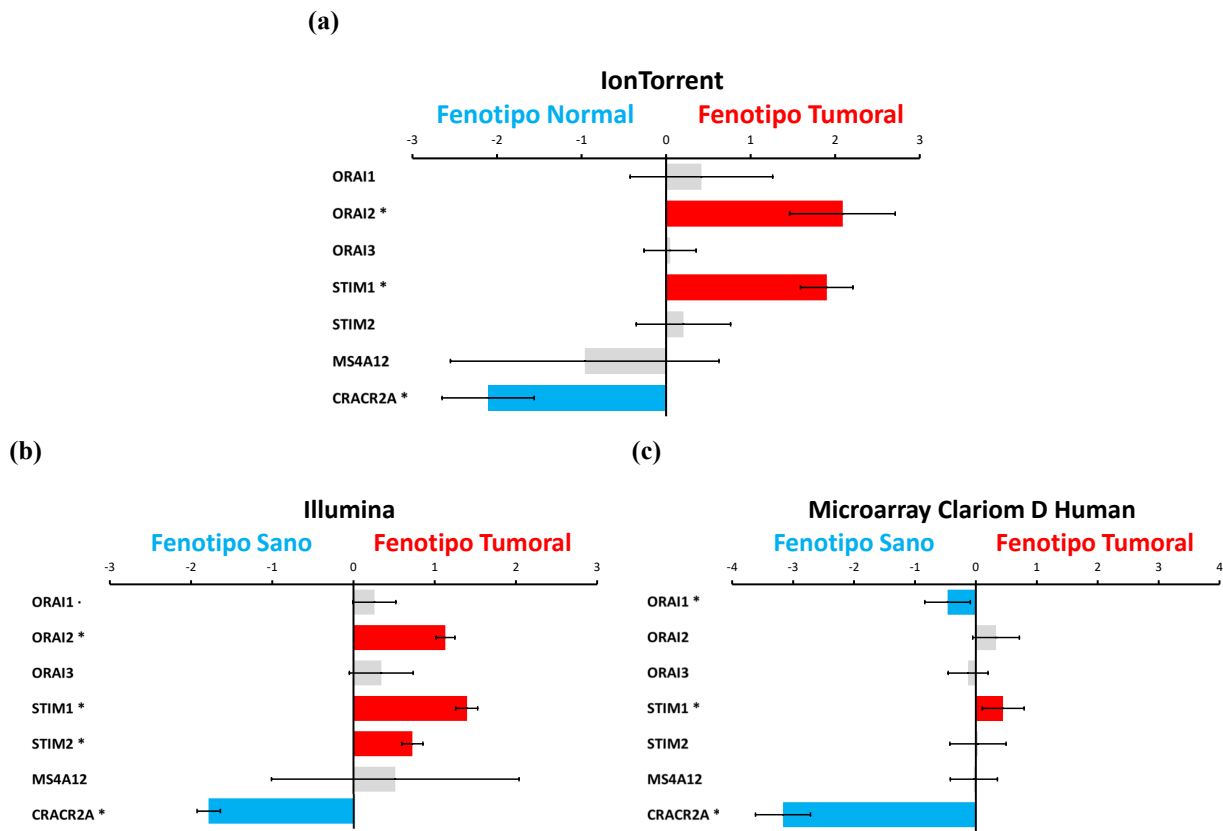
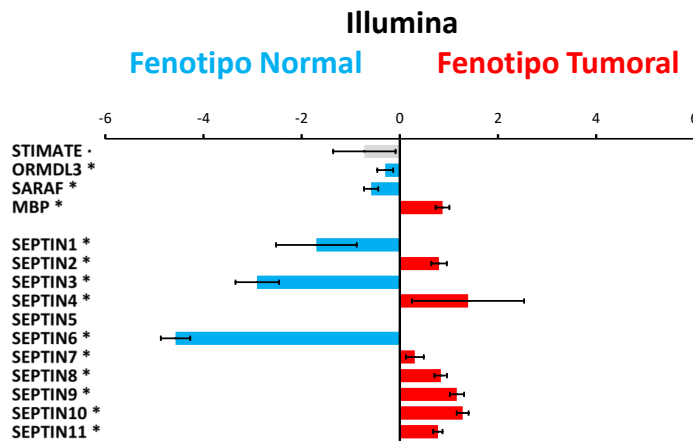


Figura 7.5: Análisis de la expresión diferencial entre las células HT29 (fenotipo tumoral) y NCM460 (fenotipo normal) para los genes de la familia SOCE. Los valores de expresión diferencial, en concreto los valores log₂(FC) estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa (FDR < 0.1) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" FDR < 0.05 y "*" FDR < 0.1. La expresión diferencial se ha realizado con tres tecnologías transcriptómicas diferentes: **(a) IonTorrent, (b) Illumina y (c) microarray Clariom D Human de Affymetrix.**

Además, con Illumina y con microarrays también se ha podido estudiar la expresión diferencial de otros reguladores de SOCE, aparte de *CRACR2A*. En este análisis, tanto con Illumina como con los microarrays se ha identificado una mayor expresión de *MBP*, *SEPTIN8* y *SEPTIN9*, y una disminución de *ORMDL3*, *SARAF*, *SEPTIN3* y *SEPTIN6* en las HT29 respecto de las NCM460.

Además, con Illumina también se detecta una mayor expresión de *SEPTIN2*, *SEPTIN4*, *SEPTIN7* y *SEPTIN11*, y una disminución de *SEPTIN1*. Curiosamente, también se observa una expresión diferencial contradictoria, pues mientras que con Illumina se observa una mayor expresión de *SEPTIN8* en HT29, con microarrays se observa una disminución (figura 7.6). Por lo tanto, se puede concluir que *MBP*, *SEPTIN2*, *SEPTIN4*, *SEPTIN7*, *SEPTIN9*, *SEPTIN10* y *SEPTIN11* están más expresados en las células HT29 respecto de las NCM460, mientras que *ORMDL3*, *SARAF*, *SEPTIN1*, *SEPTIN3* y *SEPTIN6* lo están menos.

(a)



(b)

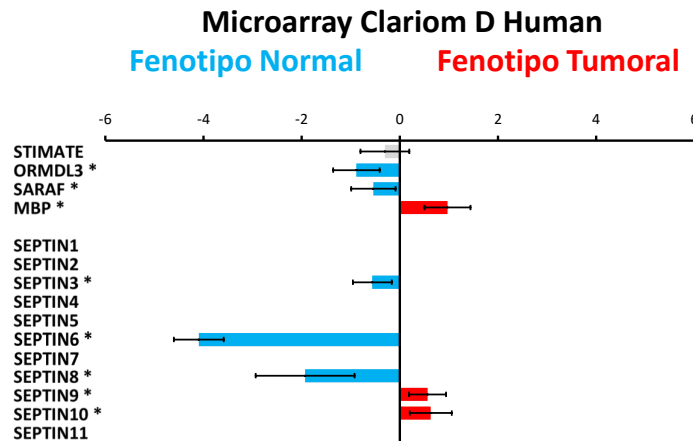


Figura 7.6: Análisis de la expresión diferencial entre las células HT29 (fenotipo tumoral) y NCM460 (fenotipo normal) para los genes que codifican para proteínas reguladoras de SOCE. Los valores de expresión diferencial, en concreto los valores log(FC) estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa (FDR <0.1) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*"FDR <0.05 y "."FDR <0.1. La expresión diferencial se ha realizado con dos tecnologías transcriptómicas diferentes: **(a) Illumina** y **(b) microarray Clariom D Human de Affymetrix**. Estos genes no fueron analizados mediante la tecnología IonTorrent

En cuanto a los genes de la superfamilia de los TRP se puede observar un patrón de infraexpresión mayoritario en las células HT29, en especial con las tecnologías de *RNA-seq*. Algunos de los genes diferencialmente expresados se detectan con las tres tecnologías transcriptómicas: infraexpresión de los genes *TRPML1-3* y sobreexpresión del *TRPV6*. No obstante, otros muestran un patrón diferente entre ellas. Para los TRPC, tanto Illumina como los microarrays muestran una menor expresión de *TRPC7* en las células HT29. Además, los microarrays también detectan la mayor expresión de HT29 de *TRPC5* y, para un FDR < 0.1, del *TRPC4*.

En cuanto a los TRPM, con IonTorrent se observa una menor expresión de *TRPM3*, *TRPM5* y *TRPM8*, mientras que con Illumina, además, también se identifica una menor expresión de *TRPM2* y *TRPM6* junto con una mayor de *TRPM7*. Como ya se ha dicho anteriormente, las tres isoformas de los *TRPML* están menos expresados en las HT29 según las tres tecnologías. Respecto de los TRPV, a parte de la mayor expresión de *TRPV6* mostrada por las tres tecnologías, IonTorrent y los microarrays también detectan una menor expresión de *TRPV1*, ambos con un FDR < 0.1. Además, Illumina detecta una menor expresión de *TRPV5* mientras que los microarrays muestran todo lo contrario; es decir, una mayor expresión de *TRPV5* (FDR < 0.1) en HT29. El *TRPA1* aparece menos expresado en HT29 según las dos tecnologías de *RNA-seq*.

Por último, con ambas tecnologías de *RNA-seq* se observa una menor expresión de *TRPP3* las células HT29 y, además, con Illumina también se identifica una menor expresión de *TRPP5*. Sin embargo, en cuanto a *TRPP1* y *TRPP2*, los cuales pueden formar un complejo en la membrana plasmática, muestran un patrón de expresión diferencial totalmente opuesto, pues mientras que con Illumina *TRPP1* está menos expresado en HT29 y *TRPP2* lo está más, con los microarrays se observa una mayor expresión de *TRPP1* y menor de *TRPP2* (FDR < 0.1).

En consecuencia, tras los resultados anteriores, se puede concluir que hay una menor expresión en las células HT29, respecto de las NCM460, de *TRPC7*, *TRPM2*, *TRPM3*, *TRPM5*, *TRPM8*, *TRPML1-3*, *TRPV1* (FDR < 0.1 tanto con IonTorrent como con microarrays), *TRPA1*, *TRPP3*, *TRPP5* (significativo con Illumina y el logFC tiene el mismo signo con los microarrays). También se puede inferir que hay una mayor expresión de *TRPC5*, *TRPV6* y *TRPM7* (significativo con Illumina y el logFC tiene el mismo signo con los microarrays). En cuanto al *TRPP1* y *TRPP2*, puesto que en ambos casos se observa expresión diferencial de ambos (un intercambio) pero en sentido opuesto, y, puesto que ambos forman un complejo funcional, se considerará que el patrón de expresión diferencial es el encontrado en el experimento de microarrays; es decir, menor expresión de *TRPP2* (FDR < 0.1) y mayor de *TRPP1*.

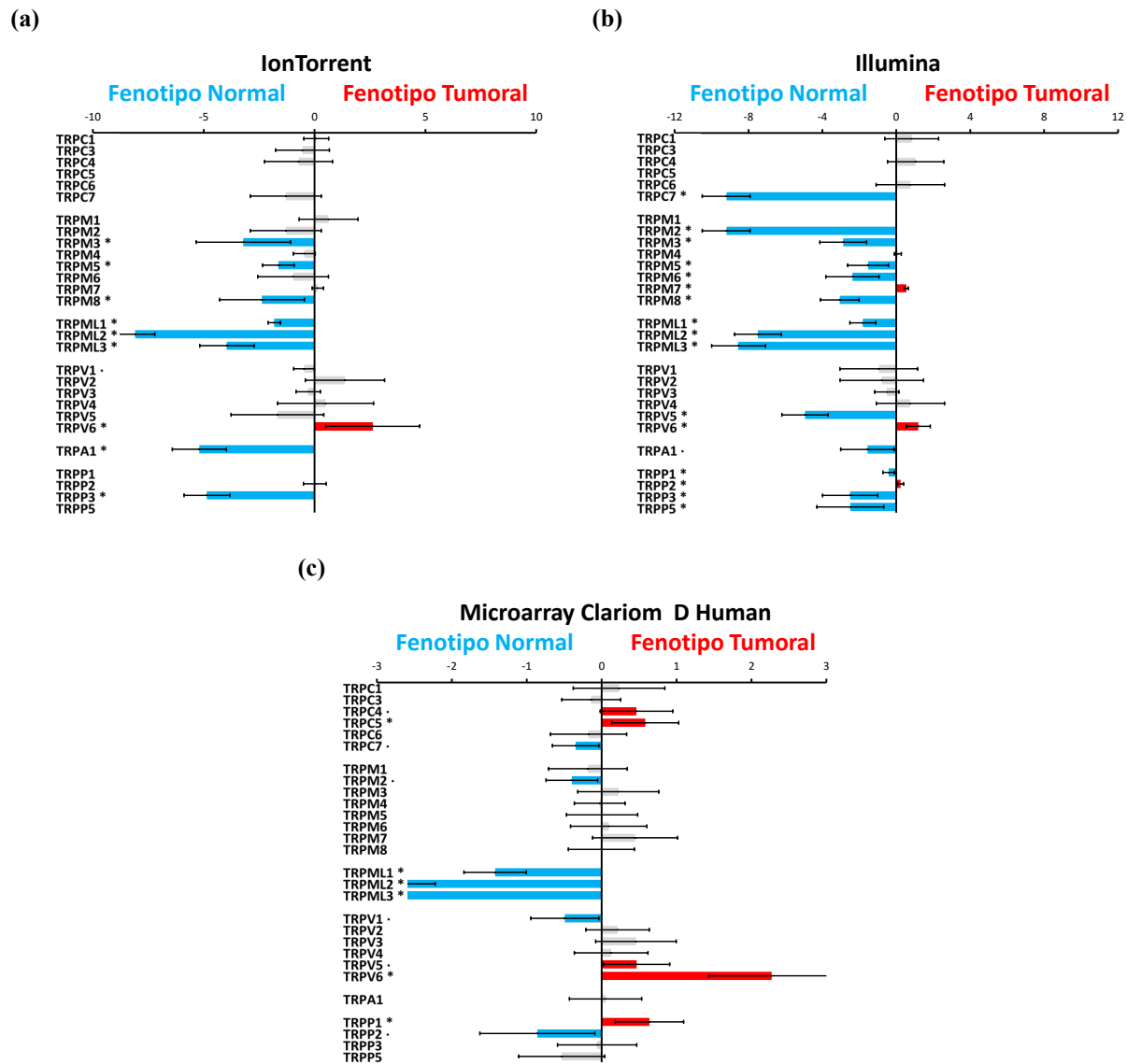


Figura 7.7: Análisis de la expresión diferencial entre las células HT29 (fenotipo tumoral) y NCM460 (fenotipo normal) para los genes de la familia TRP. Los valores de expresión diferencial, en concreto los valores $\log(FC)$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($FDR < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $FDR < 0.05$ y "." $FDR < 0.1$. La expresión diferencial se ha realizado con tres tecnologías transcriptómicas diferentes: **(a) IonTorrent, (b) Illumina y (c) microarray Clariom D Human de Affymetrix.**

En las familias de las bombas y transportadores de Ca^{2+} (CPT) el experimento con IonTorrent revela una mayor expresión en las células HT29 de *PMCA1*, *SERCA2* y *NCX2*, y una menor de *PMCA1*, *SERCA1* y *SPCA1*. Con Illumina, a diferencia de IonTorrent, no se observa la menor expresión de *SERCA1* y, por el contrario, sí se observa una mayor expresión de *SERCA3* y menor de *PMCA3*. En cuanto al experimento con microarrays, también se observa la mayor expresión de *SPCA2*, pero no se observa expresión diferencial de *PMCA3* ni de *SERCA2-3* (figura 7.8). En conclusión, se puede deducir que en las HT29 hay una mayor expresión de *PMCA1*, *SERCA2*, *NCX2* y *SPCA2*, y una menor expresión de *PMCA4* y *SPCA1*.

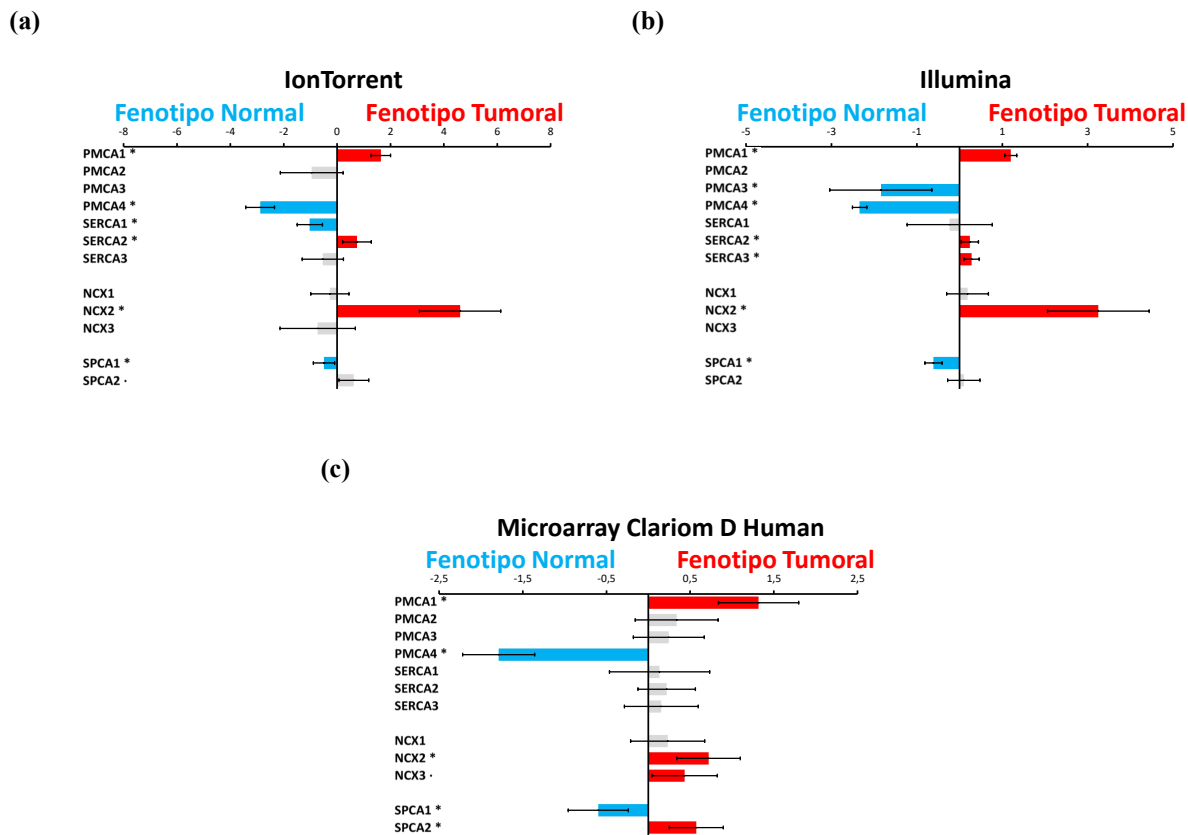


Figura 7.8: Análisis de la expresión diferencial entre las células HT29 y NCM460 para los genes de la familia CPT. Los valores de expresión diferencial, en concreto los valores $\log(\text{FC})$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($\text{FDR} < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $\text{FDR} < 0.05$ y "." $\text{FDR} < 0.1$. La expresión diferencial se ha realizado con tres tecnologías transcriptómicas diferentes: (a) IonTorrent, (b) Illumina y (c) microarray Clariom D Human de Affymetrix.

Respecto a los genes de la familia de canales de liberación de Ca^{2+} (ERC), se observa un perfil de expresión muy parecido mediante las tres tecnologías. En concreto, en las tres se observa una mayor expresión en las células HT29 de *IP3R3* y una menor de *IP3R2* y *RyR2*. Además, con ambas tecnologías de *RNA-seq* se identifica una mayor expresión de *IP3R1*. No obstante, con Illumina también se observa una mayor expresión de *RyR3* (figura 7.9). En conclusión, en HT29, respecto de NCM460, hay una mayor expresión de *IP3R1,3* y una menor de *IP3R2* y *RyR2*.

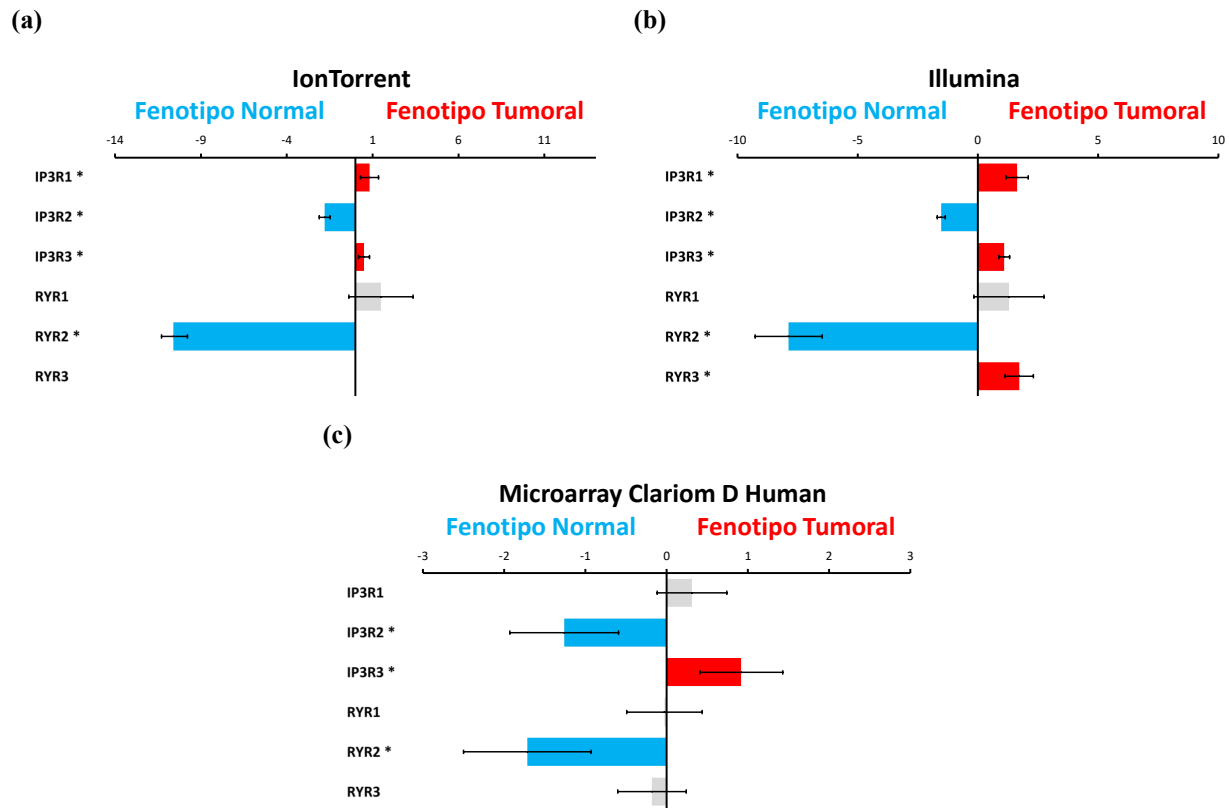


Figura 7.9: Análisis de la expresión diferencial entre las células HT29 y NCM460 para los genes de la familia ERC. Los valores de expresión diferencial, en concreto los valores $\log(\text{FC})$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($\text{FDR} < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $\text{FDR} < 0.05$ y "*" $\text{FDR} < 0.1$. La expresión diferencial se ha realizado con tres tecnologías transcriptómicas diferentes: **(a) IonTorrent, (b) Illumina y (c) microarray Clariom D Human de Affymetrix.**

Por último, al igual que sucede con los genes de ERC, el perfil de expresión de los genes de la familia proteínas transportadoras del Ca^{2+} mitocondrial (MCT) muestra un comportamiento muy parecido en los tres experimentos, puesto que se observa una mayor expresión en las células HT29 de *MCU* y una menor de *MICU2*, *MCUb*, *VDAC1* y *VDAC3*. No obstante, mientras que con Illumina y los microarrays se observa una menor expresión de *NCLX*, con IonTorrent dicho gen no se ha detectado. Además, con las tecnologías de *RNA-seq* también se identifica mayor expresión de *MICU1*, *MCUR1* y de *VDAC2* (figura 7.10). Por lo tanto, de los resultados anteriores puede concluirse que en las células HT29, respecto de las NCM460, hay una mayor expresión de *MCU*, *MICU1*, *MCUR1* y *VDAC2*, y una menor de *MICU2*, *MCUb*, *VDAC1*, *VDAC3* y *NCLX*.

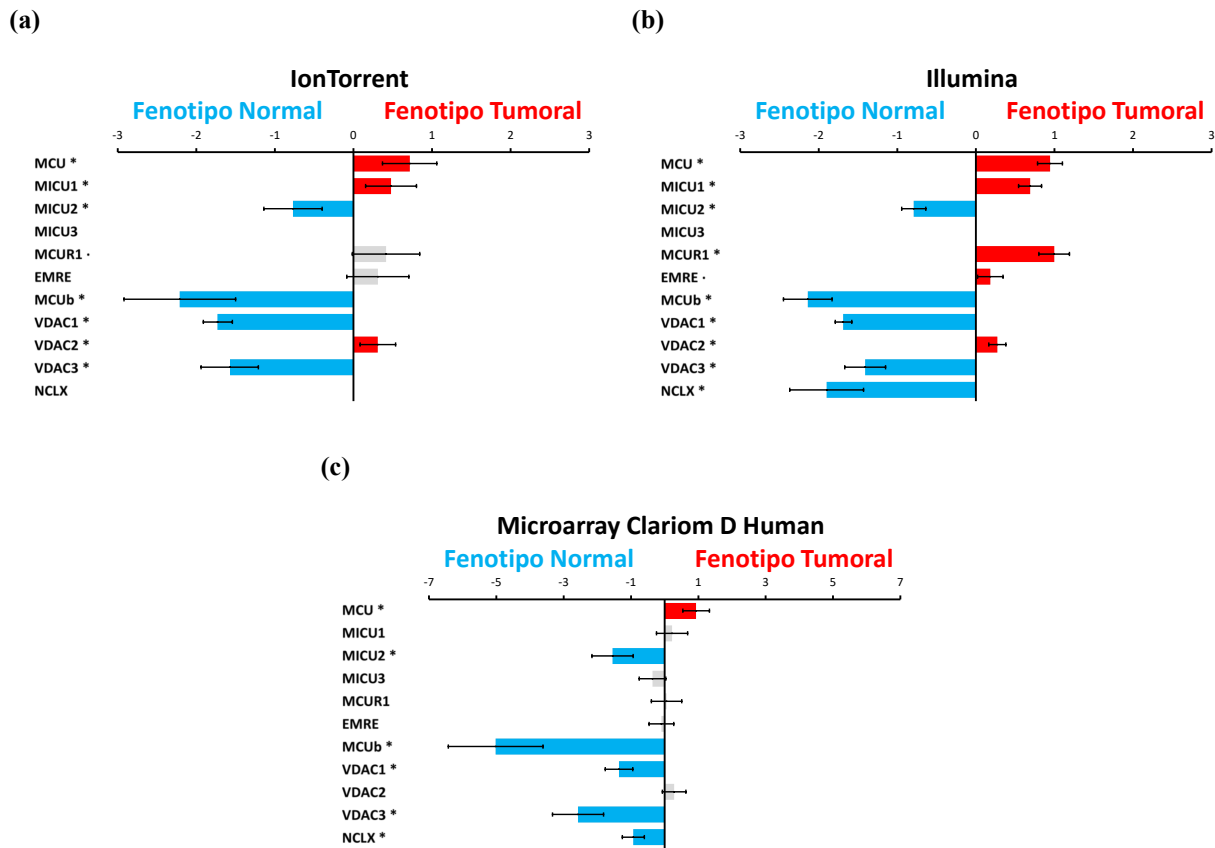


Figura 7.10: Análisis de la expresión diferencial entre las células HT29 y NCM460 para los genes de la familia MCT. Los valores de expresión diferencial, en concreto los valores $\log_2(\text{FC})$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($\text{FDR} < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $\text{FDR} < 0.05$ y "*" $\text{FDR} < 0.1$. La expresión diferencial se ha realizado con tres tecnologías transcriptómicas diferentes: **(a) IonTorrent, (b) Illumina y (c) microarray Clariom D Human de Affymetrix.**

A continuación, se muestra un esquema en el que se resume el análisis de expresión diferencial anterior. Los genes que presentan una menor expresión en las células HT29 (fenotipo tumoral) respecto de las NCM460 se representan sobre las células NCM460 (fenotipo normal), mientras que los que están más expresados se representan sobre las HT29 (figura 7.11).

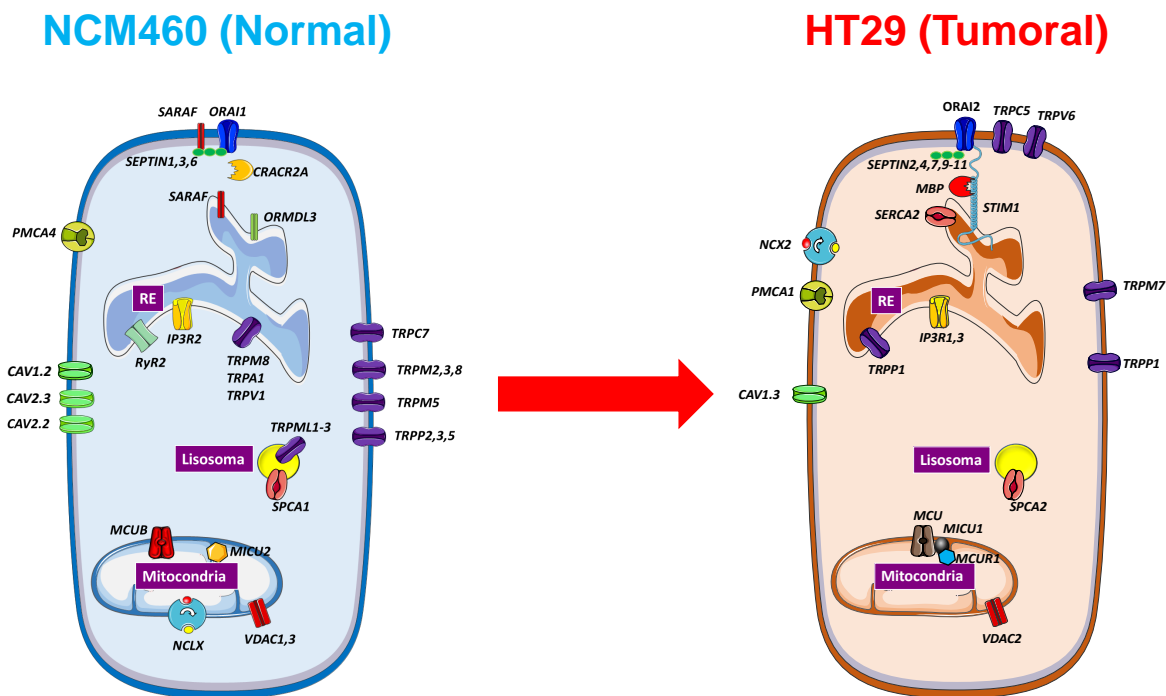


Figura 7.11: Resumen del análisis de expresión diferencial entre las células HT29 vs NCM460 de los genes implicados en el transporte de Ca^{2+} intracelular. A la izquierda se representan la célula NCM460 (fenotipo normal) y a la derecha la célula HT29. Por un lado, sobre las NCM460 se representan únicamente los genes que presentan una mayor expresión en las NCM460 respecto de las HT29. Por otro lado, sobre la HT29 se representan únicamente los genes menos expresados en las HT29 respecto de las NCM460

7.1.2.2. Análisis de significación biológica

Los anteriores análisis se han centrado en los genes relacionados con el transporte de Ca^{2+} , pero hay que tener en cuenta que, salvo para los datos mostrados de los experimentos basados en IonTorrent, el análisis de expresión diferencial se ha realizado sobre un total de 16375 genes para Illumina y 12973 sondas para los microarrays, de los cuales 14511 y 6623, respectivamente, mostraron expresión diferencial. Por lo tanto, el siguiente paso fue extraer unas conclusiones más globales de estos datos y, en la medida de lo posible, relacionarlas con la homeostasis del Ca^{2+} intracelular y el CCR. Así, estas listas de DEG se analizaron con el fin de obtener una información biológicamente relevante que no implicase estudiar el comportamiento de cada uno de los genes de forma individual, lo que se conoce en bioinformática como análisis de significación biológica. En concreto, tanto para los experimentos basados en Illumina como para los basados en los microarrays de Affymetrix, puesto que de los tres, son los dos que contienen información sobre el transcriptoma completo, se llevó a cabo el *análisis de perfiles de expresión* y de *enriquecimiento de genes*.

7.1.2.2.1. Perfiles de expresión de las líneas celulares de cáncer de colon y normales

Para evaluar si las muestras estudiadas se agrupan según lo esperado (fenotipo normal o tumoral) y cómo es el patrón de expresión de los DEG de cada fenotipo (más o menos genes sobre o infraexpresados en HT29), se han combinado los mapas de calor (expresión de genes) y el análisis de clúster jerárquico de Ward (agrupación entre muestras). Por un lado, las muestras se agrupan según lo esperado en función de la línea celular, por lo que no se observa ninguna fuente de variabilidad no deseada. Por otro lado, los genes muestran un claro patrón de infraexpresión en las células de fenotipo tumoral, HT29, respecto de las normales, NCM460, si bien es cierto que este patrón no es tan evidente en las muestras analizadas con Illumina (figura 7.12b). En particular, en las HT29 respecto de las NCM460, en el experimento con IonTorrent se han identificado 12/24 genes sobre/infraexpresados, con Illumina 6829/7681, y con microarrays 2165/4458 sondas.

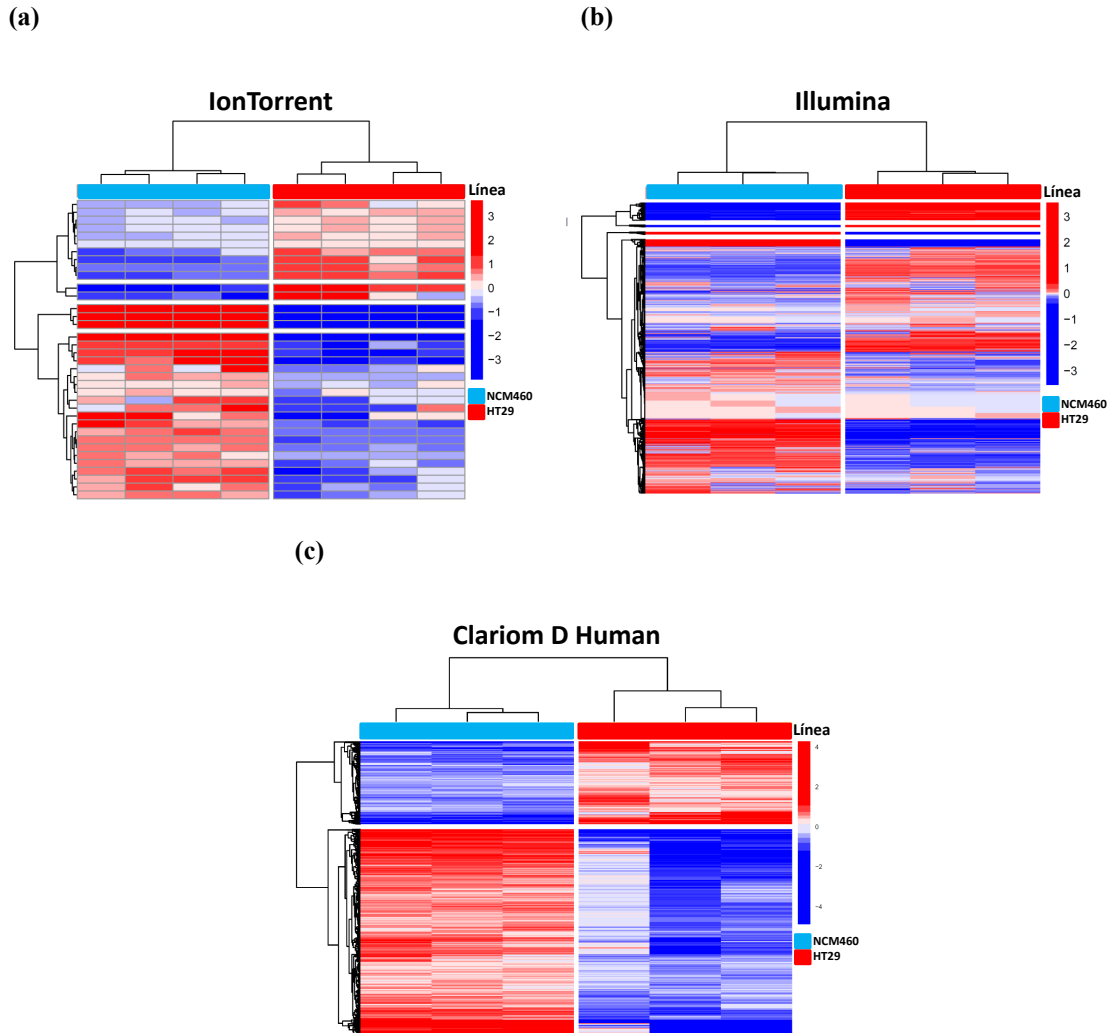


Figura 7.12: Perfiles de expresión. En los mapas de calor se representan los valores de expresión de los genes, previamente transformados (rlog para RNAseq y logaritmo en base 2 para los microarrays) y centrados (restado su media) de los valores normalizados para las muestras de RNA-seq **a) IonTorrent** y **b) Illumina**, y de **microarrays**, **c)**. En el eje X se agrupan las diferentes muestras mediante el análisis clúster jerárquico basado en el método de Ward, y en el eje Y lo hacen los DEG.

7.1.2.2.2. Análisis de enriquecimiento de las líneas celulares de cáncer de colon y normales

Tras evaluar los perfiles de expresión, se realizó un análisis de enriquecimiento a partir de la lista de DEG para responder a la pregunta ¿Qué conjuntos podrían estar relacionados con el fenotipo tumoral? Para ello, se consultaron las bases de datos KEGG PATHWAY, GO y MSigDB en busca de los diferentes conjuntos de genes y se contrastaron las hipótesis nulas competitiva, mediante el método ORA, y autocontenida, mediante CAMERA. Como resultado, en las las tablas 7.3 y 7.2 se muestra la cantidad de conjuntos que rechazan alguna de las dos hipótesis contrastadas para los datos generados por Illumina y los generados por microarrays. Además, puesto que la cantidad de

conjuntos es demasiado elevada como para estudiar el significado de cada uno de ellos, puesto que se han analizado alrededor de 30000 conjuntos diferentes de 9 bases de datos diferentes, se han buscado aquellos conjuntos que en su nombre contienen alguna de las siguientes palabras de interés: *calcium*, *mitochondria*, *reticulum*, *cancer*, *wnt* y *polyamine*. De esta forma, es posible deducir qué tipo de conjuntos de interés podrían estar relacionados con el fenotipo tumoral o contener algún gen que esté diferencialmente expresado (tablas 7.3 y 7.2).

Se ha encontrado que en los experimentos realizados con Illumina la cantidad de conjuntos que rechazan alguna de las hipótesis nulas es demasiado bajo (13 con CAMERA y 87 con ORA). No obstante, de entre ellos, hay 2 conjuntos que contienen la palabra *cancer* (*M929 SABATES COLORECTAL ADENOMA UP* y *M9192 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL DN*) y uno *wnt* (*M5895 HALLMARK WNT BETA CATENIN SIGNALING*), lo que es esperable y, además, válida, en parte, los resultados obtenidos (tabla 7.3).

Tabla 7.2: Análisis de enriquecimiento de genes para la expresión diferencial entre las células HT29 y NCM460 llevada a cabo con el microarray Clariom D Human.

| Base de Datos | Conjuntos | | Conjuntos Significativos (FDR <0.05) | | | | | | | |
|---------------|--------------|---------------------|--------------------------------------|-------------|---------------------|---------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | Métodos | Totales | Calcio ¹ | Mitochondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA | 25 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | | ORA | 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO BP | 12342 | 146 | CAMERA | 35 | 0 | 2 | 0 | 0 | 0 | 0 |
| | | | ORA | 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO MF | 4245 | 66 | CAMERA | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| MH | 50 | 0 | CAMERA | 11 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| MCDB | 6290 | 12 | CAMERA | 567 | 0 | 4 | 0 | 27 | 1 | 0 |
| | | | ORA | 528 | 0 | 6 | 0 | 46 | 3 | 0 |
| MCC | 858 | 0 | CAMERA | 89 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 62 | 0 | 0 | 0 | 0 | 0 | 0 |
| MOG | 189 | 0 | CAMERA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| MIM | 5219 | 0 | CAMERA | 234 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 1204 | 0 | 0 | 0 | 0 | 2 | 0 |
| Total | 29537 | 225 | CAMERA | 967 | 0 | 6 | 1 | 28 | 1 | 0 |
| | | | ORA | 1823 | 0 | 6 | 0 | 46 | 5 | 0 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Onncogenic; MIM: Msigdb Immunology.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

En cuanto al análisis de enriquecimiento obtenido a partir del análisis de expresión diferencial del experimento basado en microarrays, el número de conjuntos que rechazan alguna de las hipótesis autocontenida/competitiva es más elevado que en el caso de Illumina (tabla 7.2: 967/1823, de los cuales, 89/62 pertenecen a la base de datos MCC y 4/6 a MOG, 28/46 contienen en su nombre la

palabra *cancer* (M16740 GRADE COLON AND RECTAL CANCER UP...), 6/6 *mitochondria* (GO: 0032981 mitochondrial respiratory chain complex I assembly, M15112 WONG MITOCHONDRIA GENE MODULE...), 1/0 *reticulum** (hsa04141 Protein processing in endoplasmic reticulum) y 1/5 *wnt* (M7847 REACTOME SIGNALING BY WNT, M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT...).

Tabla 7.3: Análisis de enriquecimiento de genes para la expresión diferencial entre las células HT29 y NCM460 llevada a cabo con Illumina (RNA-seq).

| Base de Datos | Conjuntos | | Métodos | Conjuntos Significativos (FDR <0.05) | | | | | | |
|---------------|--------------|---------------------|---------|--------------------------------------|---------------------|---------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | | Totales | Calcio ¹ | Mitochondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO BP | 12342 | 146 | CAMERA | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO MF | 4245 | 66 | CAMERA | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MH | 50 | 0 | CAMERA | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | | ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MCDB | 6290 | 12 | CAMERA | 4 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | | ORA | 16 | 0 | 0 | 0 | 2 | 0 | 0 |
| MCC | 858 | 0 | CAMERA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MOG | 189 | 0 | CAMERA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MIM | 5219 | 0 | CAMERA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 71 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 29537 | 225 | CAMERA | 13 | 0 | 0 | 0 | 1 | 1 | 0 |
| | | | ORA | 87 | 0 | 0 | 0 | 2 | 0 | 0 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

Si bien es cierto que estos resultados pueden proporcionar una información interesante para futuras investigaciones, tratarlos en más detalle estaría fuera del propósito de la presente tesis. Por un lado, en el apéndice “B.1.2. Análisis de enriquecimiento de genes de la expresión diferencial entre las líneas celulares HT29 y NCM460 de los experimentos basados en RNA-seq” se muestran, para los experimentos basados en Illumina, todos los conjuntos de genes con un FDR < 0.1 -puesto que son un número aceptable como para ser mostrados-, divididos en diferentes tablas en función tanto de la base de datos a la que pertenecen como del método empleado para su análisis. Por otro lado, para los basados en el microarray Clariom D Human, en el apéndice “B.3.3.3. Análisis de enriquecimiento de genes de la expresión diferencial entre las células HT29 y NCM460” se muestran aquellos conjuntos con un FDR < 0.1 que contienen las palabras de interés, puesto que la cantidad total de conjuntos de genes que rechazan alguna de las hipótesis nulas es demasiado elevada como para mostrar todos aquellos con un FDR < 0.1.

7.1.3. Efecto del tratamiento con DFMO sobre la homeostasis del Ca^{2+} intracelular en colonocitos normales y en células de cáncer de colon

El DFMO es un inhibidor suicida de la ruta de síntesis de las poliaminas la cual está sobreactivada en el CCR como consecuencia de alteraciones en la ruta de señalización Wnt -causa más frecuente de aparición de CCR-. Por ello, tras el análisis transcriptómico en los modelos celulares NCM460 y HT29, en el que se observó una relación entre la ruta de señalización Wnt y el fenotipo de CCR, como era de esperar (tabla 7.2), se estudió tanto a nivel funcional como transcriptómico el efecto del DFMO en cada uno de estos modelos, un tratamiento que produce la depleción de poliaminas endógenas. De hecho, autores como Gutiérrez et al. (2019) ya observaron un efecto del DFMO sobre la homeostasis del Ca^{2+} intracelular.

Por un lado, el estudio funcional se llevó a cabo mediante experimentos de imagen de Ca^{2+} en los que se evaluaron los niveles de calcio intracelular basal, las diferencias en el tamaño de los depósitos intracelulares de Ca^{2+} y de SOCE. Por otro lado, el estudio transcriptómico se llevó a cabo mediante un experimento basado en microarrays. Además, cruzando los datos de ambos experimentos también se pretendió identificar que genes podrían explicar los cambios encontrados en los experimentos de imagen de Ca^{2+} . Para asegurar que las condiciones experimentales de ambos fueran lo más parecidas posible, cada una de las muestras del experimento de imagen se cultivó y trató a la par que otra de las muestras del experimento transcriptómico (ver protocolo en el apartado “6.2. Tratamiento de células NCM460 y HT29 con DFMO”).

En cuanto al diseño experimental, el experimento de imagen de Ca^{2+} consta del factor fijo *Línea Celular* a dos niveles (*NCM460* y *HT29*), del factor fijo *Tratamiento* a dos niveles (*Control* y *DFMO*), del factor aleatorio *Día* (agrupa las muestras analizadas el mismo día) y el factor aleatorio *Cristal* (unidad experimental, el cristal en el que se tratan las células y estas son medidas). Respecto del experimento transcriptómico, su diseño es el mismo que el de imagen salvo que en este no existe el factor aleatorio *Cristal*, puesto que el ARN de todas las células de una misma placa petri se extrae a la vez y se agrupa como una sola unidad experimental.

7.1.3.1. Estudio del efecto del DFMO sobre el Ca^{2+} citosólico, el tamaño de los depósitos de Ca^{2+} intracelular y SOCE

Para evaluar el efecto del DFMO a nivel funcional, se ha empleado la técnica de imagen de Ca^{2+} y la sonda de Ca^{2+} citosólico Fura2/AM. Como se explicó en el apartado “6.2. Tratamiento de células NCM460 y HT29 con DFMO”, en cada uno de los registros se ha evaluado tanto el $[\text{Ca}^{2+}]_{\text{cit}}$ basal, el tamaño de los depósitos de Ca^{2+} como la SOCE. El tamaño de los depósitos se evaluó a partir de la liberación del Ca^{2+} desde el RE al citosol causada por el inhibidor de la SERCA CPA. Tras el vaciado de los depósitos, se reveló SOCE mediante la perfusión con medio con Ca^{2+} 1 mM. En cuanto al $[\text{Ca}^{2+}]_{\text{cit}}$ basal, se ha tomado la mediana de la señal en los primeros 30 segundos del registro en medio con Ca^{2+} 1 mM. Se han evaluado cuatro condiciones experimentales generadas por los factores fijos *Línea* = {HT29, NCM460} y *Tratamiento* = {Control, DFMO}, y que

de cada una de ellas hay 8 réplicas realizadas en 4 días diferentes -dos cada día-, hay un total de 32 réplicas en las que se han registrado más de 1000 células. El diseño experimental detallado se explica en el apartado “6.3.2. Procedimiento y diseño experimental para el análisis funcional del efecto del DFMO en los modelos celulares HT29 y NCM460”. En la figura 7.13 se muestra una unidad experimental representativa de cada condición experimental.

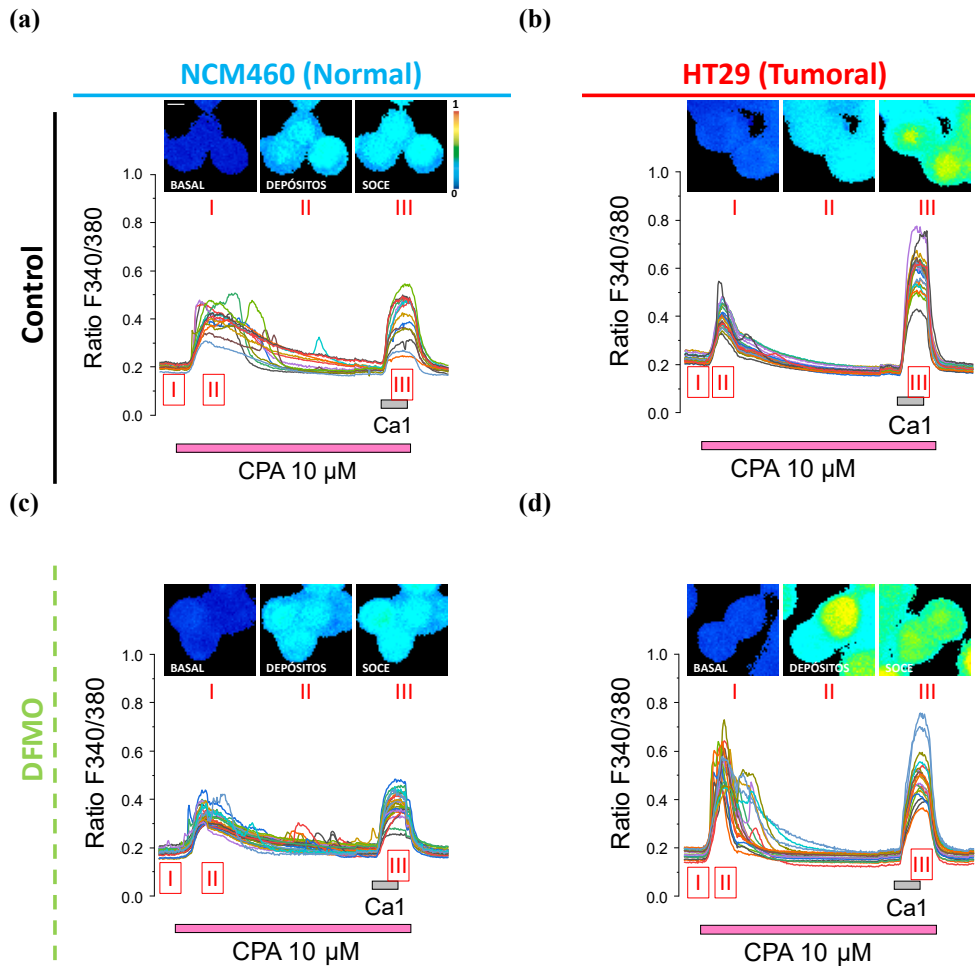


Figura 7.13: Efecto del DFMO en el tamaño de los depósitos y SOCE. Se muestran registros representativos de imagen de Ca^{2+} intracelular en célula única para cada una de las condiciones experimentales (a) células NCM460 Control -sin tratar-, (b) HT29 Control, (c) NCM460 tratadas con DFMO y (d) HT29 tratadas con DFMO. En la parte superior de cada figura se muestra la imagen, en escala de pseudocolor para la condición basal (I), liberación del Ca^{2+} de los depósitos (II) y SOCE (III). En la parte inferior se representa el ratio de fluorescencia 340/380 en función del tiempo y los estímulos empleados. En total, se han obtenido 8 registros para cada condición experimental los cuales han sido realizados en tandas de dos réplicas de cada una de las diferentes condiciones experimentales cada día a lo largo de cuatro días diferentes espaciados en el tiempo una semana mínimo. El número total de células en los 8 registros, para cada condición experimental, ha sido 272 para NCM460 Control, 230 para HT29 Control, 303 para NCM460 DFMO y 222 para HT29 DFMO. En estos experimentos se ha empleado la sonda de Ca^{2+} ratiométrica Fura2/AM para detectar el Ca^{2+} presente en el citosol y se ha registrado la fluorescencia a 340 y 380 nm. En particular, se estudió el tamaño de los depósitos de Ca^{2+} a través de la liberación del Ca^{2+} al citosol con CPA $10 \mu\text{M}$, inhibidor reversible de la SERCA, en medio libre de Ca^{2+} y, posteriormente, en el mismo registro, se evaluó SOCE a través de la perfusión de medio con Ca^{2+} 1 mM. La escala de pseudocolor a la derecha de las imágenes de la primera figura muestra el código de colores para el ratio de las fluorescencias que va desde el 0, azul, al 1, rojo. La barra blanca representa $10 \mu\text{m}$.

A partir de estos registros, se han modelizado, de forma independiente, cada una de las cinco variables respuesta de interés: $[Ca^{2+}]_{cit}$ basal (*CaBasal*), Área Bajo la Curva del Ratio de Fluorescencia 340/380 (AUC) de la señal correspondiente a los depósitos (*AreaDepositos*), Máximo Incremento del Ratio de Fluorescencia 340/380 de la señal (ΔMax) correspondiente a los depósitos (*MaxDepositos*), AUC de la señal correspondiente a SOCE (*AreaSOCE*) y (ΔMax) de la señal correspondiente a SOCE (*MaxSOCE*). Para ello, tal y como se explica en el apartado “6.3.3. Análisis de los datos de experimentos de imagen de Ca^{2+} ”, se han empleado modelos lineales mixtos lo que permite analizar datos con estructura agrupada (en este caso por las células agrupadas dentro de un mismo cubreobjetos o cristal que, a su vez, se agrupan dentro de un mismo día en los que se realizan dos réplicas de cada condición experimental. Es decir, que no sólo se consideran las variables explicativas de interés (factores *Línea* y *Tratamiento*), sino también los efectos aleatorios debidos a los factores aleatorios *Cristal* y *Día*.

Los resultados correspondientes al análisis de datos de este experimento (modelización de la señal, diagnóstico del modelo, contrastes de hipótesis, estimaciones, métodos *bootstrapping*...), para cada una de las señales estudiadas, se presentan en los diferentes subapartados del apéndice “C. Análisis funcional del efecto del DFMO sobre la homeostasis del Ca^{2+} intracelular”. Además, en el apéndice “C.1. Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos” se presenta un ejemplo simplificado de aplicación con lineales mixtos a este tipo de experimentos, seleccionar el mejor modelo con la variable transformada según la familia de transformaciones potencia de Box-Cox (Hawkins and Weisberg (2017)), evaluar si es o no adecuado el modelo y, en caso de que no lo sean -no cumplan la normalidad, la homocedasticidad...-, usar técnicas más complejas de remuestreo paramétrico como *bootstrapping*.

Tras ajustar los diferentes modelos lineales mixtos clásicos y *bootstrap* a cada una de las 4 variables respuesta, en aquellos casos en los que el término de interacción resultó ser significativo (o deba incluirse según AIC) se compararon dos a dos todas las condiciones experimentales y se realizó el contraste de hipótesis oportuno para identificar cuales de ellas son diferentes entre sí (p-valor ajustado con Tukey inferior a 0.05). Además, a partir de aplicar la teoría de grafos (algoritmo de Bron-Kerbosch⁶) sobre estas diferencias dos a dos se agruparon las diferentes condiciones experimentales según puedan ser o no consideradas como iguales.

⁶Este algoritmo busca los cliques maximales. En un grafo no dirigido, un clique es un subgrafo en el cual todo vértice está conectado por una arista a todos los demás vértices pertenecientes a dicho subgrafo.

En las figuras 7.14, 7.15 y 7.16 se muestran los resultados mediante diagramas de barras para $[Ca^{2+}]_{cit}$ basal, para los depósitos y para SOCE, respectivamente. Además, los resultados numéricos correspondientes a los modelos lineales mixtos estimados mediante métodos clásicos⁷ se muestran en las tablas C.12, C.27, C.42 y C.58, y los correspondientes a los modelos *bootstrap* en las tablas C.14, C.30, C.45, C.61.

En primer lugar, las medianas de $[Ca^{2+}]_{cit}$ basal durante el primer minuto del experimento en medio con Ca^{2+} 1 mM han mostrado, por un lado, que las células HT29 sin tratar tienen una $[Ca^{2+}]_{cit}$ basal superior al de las NCM460 sin tratar. Por otro lado, el tratamiento con DFMO reduce $[Ca^{2+}]_{cit}$ basal en ambos casos, de tal manera que la disminución tiene la misma magnitud en ambas líneas celulares. Curiosamente, las células HT29 tratadas con DFMO reducen su $[Ca^{2+}]_{cit}$ basal hasta unos niveles idénticos a los de las NCM460 sin tratar, por lo que se consigue revertir parcialmente el fenotipo tumoral en este contexto.

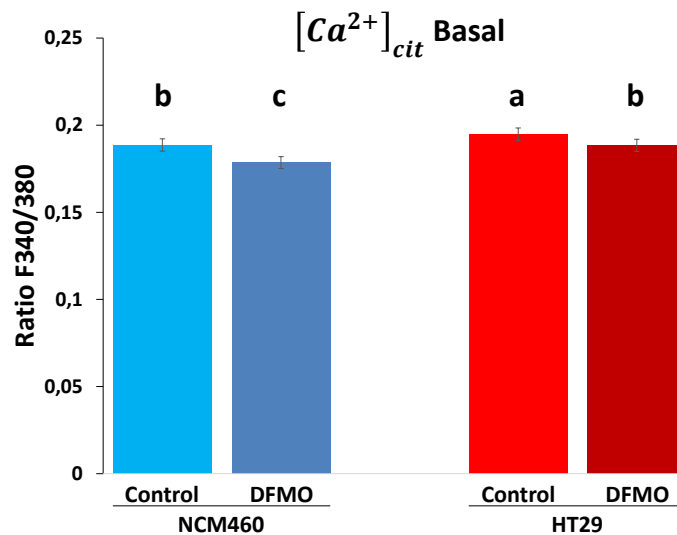


Figura 7.14: Estimaciones del efecto del DFMO sobre la $[Ca^{2+}]_{cit}$ basal. Los diagramas de barras representan los valores esperados (estimados mediante modelos lineales mixtos) de $[Ca^{2+}]_{cit}$ basal (mediana de la señal durante el primer minuto en medio Ca^{2+} 1 mM) para cada una de las condiciones experimentales evaluadas y las barras su error \pm SEM (estimados mediante modelos lineales mixtos). En este caso el termino de interacción entre los factores *Célula* y *Tratamiento* no ha resultado significativo, pero sí los dos efectos principales. No obstante, se ha realizado el contraste de hipótesis para evaluar si es o no diferente la $[Ca^{2+}]_{cit}$ entre las HT29 tratadas con DFMO y las NCM460 sin tratar. En la figura 7.13 se observa una unidad experimental representativa para cada condición experimental, de un total de 32 unidades experimentales, con 272 células para NCM460 Control, 230 para HT29 Control, 303 para NCM460 DFMO y 222 para HT29 DFMO.

⁷Los resultados de los contrastes de hipótesis para las comparaciones dos a dos son prácticamente idénticos en los modelos clásicos y *bootstrap*, lo que evidencia que la ligera violación de los supuestos del modelo, y más aun con el tamaño muestral tan grande (más de 1000 células), no es preocupante.

En relación al tamaño de los depósitos, se puede observar cómo este es inferior en las células HT29 sin tratar respecto de las NCM460 sin tratar; es decir, que el fenotipo tumoral muestra un tamaño de los depósitos inferior, tanto a nivel de incremento de la señal como de la AUC. No obstante, es evidente que el tratamiento con DFMO aumenta considerablemente el tamaño de los depósitos hasta tal punto que, por un lado, elimina las diferencias existentes entre las células HT29 tratadas y NCM460 sin tratar a nivel de AUC; sin embargo, el incremento observado en las HT29 tratadas es muy superior al mostrado por las otras tres condiciones experimentales.

En cuanto al efecto observado en las células NCM460, se aprecia un patrón de comportamiento contrario al de las HT29; es decir, que el AUC disminuye con el tratamiento, si bien es cierto que tal efecto no se observa en el incremento. En concreto, tal y como puede observarse en la figura 7.15b, la condición NCM460 tratada está a caballo entre las HT29 y NCM460 sin tratar ya que, a pesar de que el incremento en las HT29 sin tratar es inferior al de las NCM460 también sin tratar, las NCM460 tratadas no es diferente a ninguna de ellas. Por lo tanto, el DFMO parece revertir parcialmente las diferencias en el tamaño de los depósitos entre las células NCM460 y HT29.

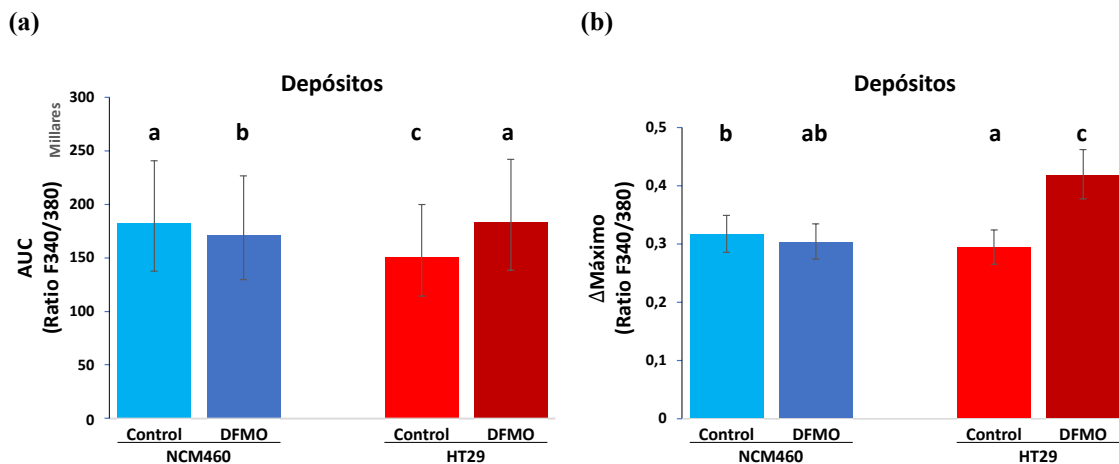


Figura 7.15: Estimaciones del efecto del DFMO en el tamaño de los depósitos. Los diagramas de barras representan los valores esperados (estimados mediante modelos lineales mixtos) del tamaño de los depósitos para cada una de las condiciones experimentales evaluadas y las barras su error \pm SEM (estimados mediante modelos lineales mixtos). El Δ Max de la señal de los depósitos se muestra en la figura b) y el AUC en a). Las agrupaciones entre las condiciones experimentales se han realizado a partir de las diferencias dos a dos y el algoritmo de Bron-Kerbosch, de tal manera que dos condiciones experimentales con la misma letra son iguales, dos con diferente letra son diferentes y en caso de que alguna condición tenga más de una letra, por ejemplo *ab*, se considera similar tanto a la condición *a* como a la *b* aunque las condiciones *a* y *b* sean diferentes entre sí. En la figura 7.13 se observa una unidad experimental representativa para cada condición experimental, de un total de 32 unidades experimentales, con 272 células para NCM460 Control, 230 para HT29 Control, 303 para NCM460 DFMO y 222 para HT29 DFMO.

En cuanto a SOCE, se observa que esta es superior en las células HT29 respecto de las NCM460. Además, el tratamiento parece no tener efecto alguno sobre este mecanismo en las NCM460, pero sí en las HT29 (ver apéndice “C.35, C.42” y figura 7.16).

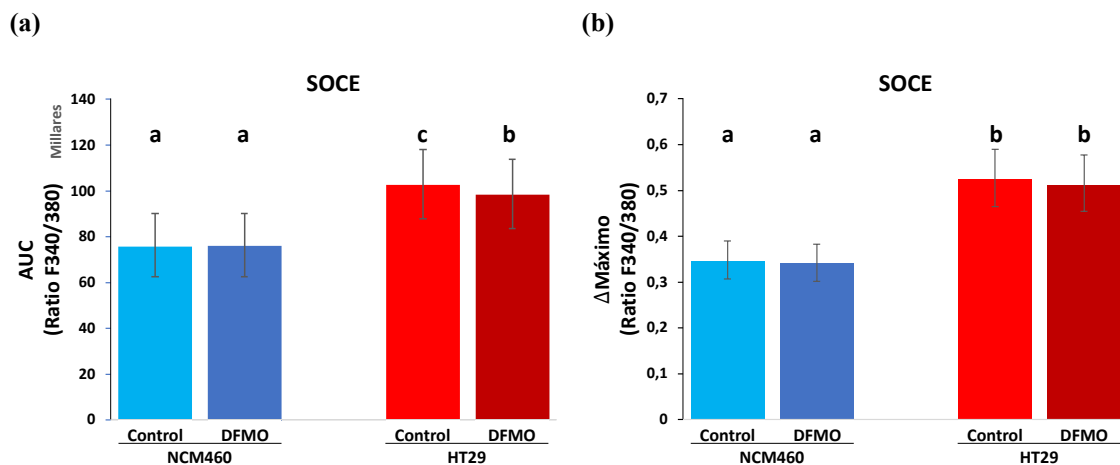


Figura 7.16: Estimaciones del efecto del DFMO en SOCE. Los diagramas de barras representan los valores esperados (estimados mediante modelos lineales mixtos) de SOCE para cada una de las condiciones experimentales evaluadas y las barras de error su valor \pm SEM (estimados mediante modelos lineales mixtos). El Δ Max de SOCE se muestra en la figura (b) y el AUC en a). Las agrupaciones entre las condiciones experimentales se han realizado a partir de las diferencias dos a dos y el algoritmo de Bron-Kerbsch, de tal manera que dos condiciones experimentales con la misma letra son iguales, dos con diferente letra son diferentes y en caso de que alguna condición tenga más de una letra, por ejemplo *ab*, se considera similar tanto a la condición *a* como a la *b* aunque las condiciones *a* y *b* sean diferentes entre sí. En la figura 7.13 se observa una unidad experimental representativa para cada condición experimental, de un total de 32 unidades experimentales, con 272 células para NCM460 Control, 230 para HT29 Control, 303 para NCM460 DFMO y 222 para HT29 DFMO.

7.1.3.2. Análisis transcriptómico de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO

Para el estudio transcriptómico del efecto del DFMO en las líneas celulares HT29 y NCM460 se han empleado un total de 3 muestras para cada condición experimental y se ha usado el microarray Clariom D Human de Affymetrix. Como se comentó en el apartado “6.2. Tratamiento de células NCM460 y HT29 con DFMO”, el diseño experimental consta de dos factores fijos (*Línea Celular* -HT29 y NCM460- y *Tratamiento* -Control y DFMO-) y un factor aleatorio (*Día* del comienzo del tratamiento de la unidad experimental o muestra en cuestión). Tras obtener los datos crudos, *.CEL*, de los microarrays, estos se sometieron a un control de calidad -no se detectó ningún microarray problemático que debiera ser eliminado- y fueron preprocesados con el método RMA (ver apéndice ”B.3. Análisis transcriptómico de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las líneas HT29 y NCM460). Posteriormente, se sometieron a los análisis de expresión diferencial y de significación biológica.

7.1.3.2.1. Análisis de expresión diferencial del efecto del DFMO sobre las líneas celulares HT29 y NCM460

Tras el control de calidad y el preprocesamiento con el método RMA, previo al análisis de expresión diferencial se realizó un filtrado no específico de las sondas, de tal manera que se pasó de 138745 a 69478 sondas de las cuales 12835 tiene correspondencia con UNIGENE y 12973 con ALIAS. Así, a partir del nuevo conjunto de datos que sólo contiene las sondas con correspondencia ALIAS, se procedió al análisis de expresión diferencial mediante el método *limma*, para lo que, previamente, se seleccionó el mejor modelo lineal a partir de los criterios AIC y BIC que, en este caso, resultó ser el modelo lineal mixto completo; es decir, con los dos factores, la interacción entre ellos y el factor aleatorio (apéndice “B.3.1. Control de calidad y preprocesado de los datos transcryptómicos de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO”). En consecuencia, se ajustó el modelo completo a los datos y se realizaron los contrastes de expresión diferencial de interés para cada una de las sondas, donde se rechaza la H_0 de no expresión diferencial para aquellas sondas con un $FDR < 0.05(0.1)$ (figura 7.17, tabla 7.4):

Tabla 7.4: Resumen de los DEG para diferentes contrastes de interés tras el tratamiento con DFMO.

| Contraste | Sondas | |
|-----------------------------|------------|----------------|
| | Número | % ¹ |
| HT29.Control-NCM460.Control | 6623(7516) | 51.05(57.93) |
| HT29.DFMO-NCM460.Control | 5707 | 43.99 |
| HT29.DFMO-NCM460.DFMO | 5286 | 40.75 |
| HT29.DFMO-HT29.Control | 3394(4402) | 26.16(33.93) |
| NCM460.DFMO-NCM460.Control | 62(165) | 0.48(1.27) |

Notas:

¹ El porcentaje está realizado respecto de un total de 12973 sondas con correspondencia con ALIAS que han quedado tras el filtrado.

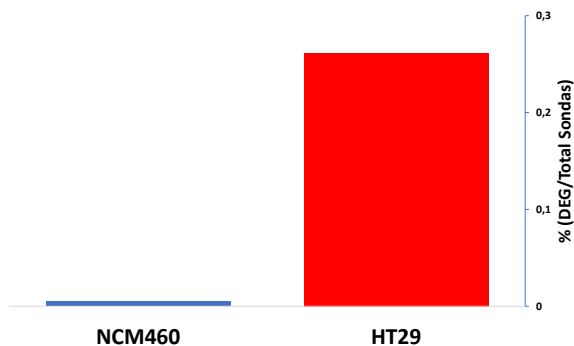


Figura 7.17: Porcentaje de DEG tras el tratamiento con DFMO. En la figura se muestra el porcentaje de DEG, respecto de un total de 12973 sondas, tras el tratamiento con DFMO el cual muestra un evidente comportamiento selectivo sobre las HT29 (fenotipo tumoral) respecto de las NCM460 (fenotipo normal).

A grandes rasgos, de los contrastes anteriores se puede extraer una información sumamente importante. En primer lugar, lo más destacable es la evidente selectividad del tratamiento con DFMO sobre las células de fenotipo tumoral HT29, pues mientras que este tratamiento sólo afecta a 62 sondas para las células NCM460, para las HT29 se ven afectadas un total de 3394. De las 6623 sondas (51.05 %) diferencialmente expresadas entre NCM460 y HT29 en condiciones control, el tratamiento con DFMO reduce esta diferencia en unos 7 puntos, pues entre las células HT29 tratadas con DFMO y las NCM460 control hay 5707 sondas diferencialmente expresadas (44 %).

Por lo tanto, ante los resultados anteriores es lógico querer dar respuesta a preguntas tales como ¿Qué genes están afectados tanto en NCM460 y HT29? o ¿Cuáles lo están sólo en las NCM460?. Para ello, para estos dos contrastes se muestran los diagramas de Venn de DEG sobreexpresados (7.18a), infraexpresados (7.18b), sobreexpresados en HT29 e infraexpresados en NCM460 (7.18c) y viceversa (7.18d). Como se puede comprobar, el patrón de expresión diferencial que domina tras el efecto del DFMO es, en ambas líneas celulares, de sobreexpresión.

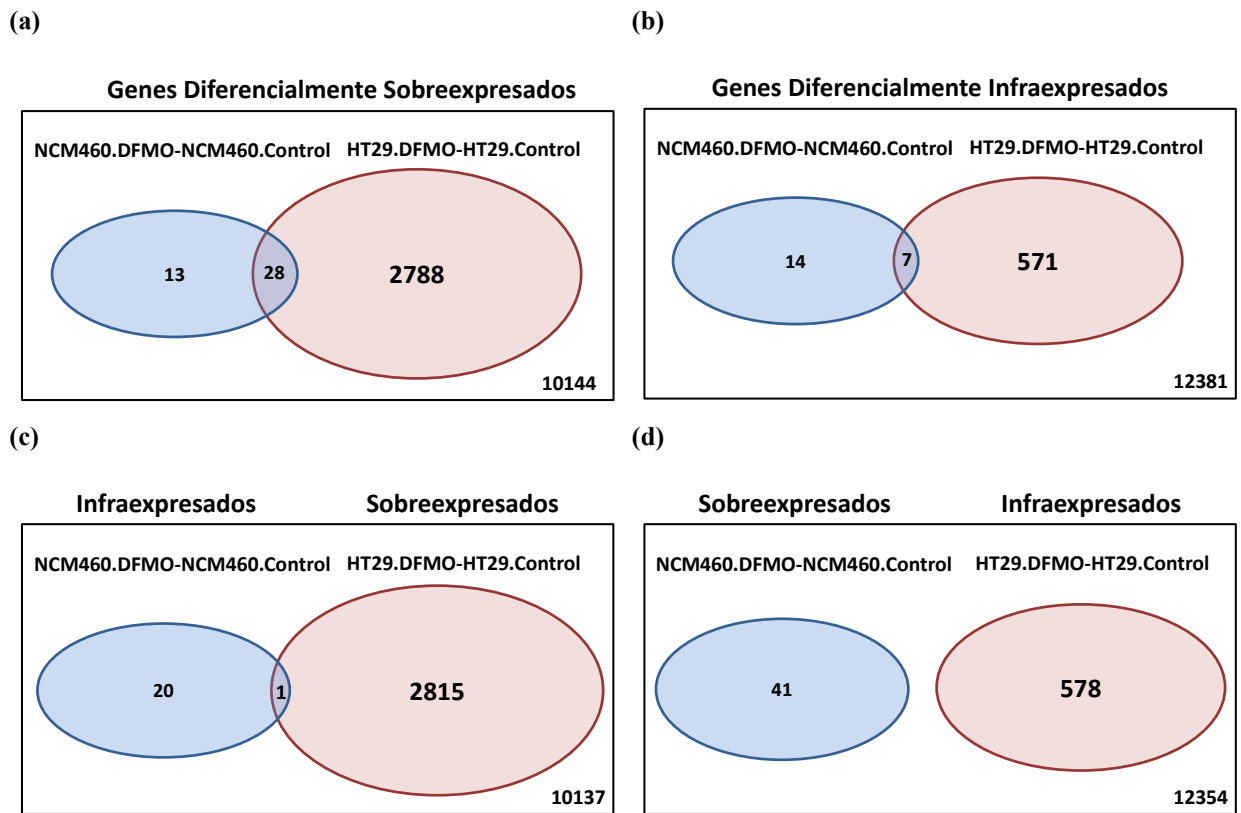


Figura 7.18: Diagramas de Venn para el análisis de expresión diferencial. a) DEG sobreexpresados con el tratamiento en ambas líneas celulares; b) DEG infraexpresados con el tratamiento en ambas líneas celulares; c) DEG sobreexpresados con el tratamiento en HT29 e infraexpresados en NCM460, y d) DEG infraexpresados con el tratamiento en HT29 y sobreexpresados en NCM460.

Aunque los diagramas de Venn muestran una información bastante interesante a nivel global, también sería interesante poder poner cara a estos números. Por ello, en las figuras 7.19 y 7.20 se muestran los resultados de expresión diferencial de todos los DEG tras el tratamiento con DFMO sobre las NCM460 y se comparan con los resultados obtenidos en HT29 (el resto de resultados de los DEG de HT29 no se muestran debido a que es una cantidad demasiado elevada). Curiosamente, mientras que los genes que muestran un comportamiento de expresión diferencial idéntico en ambas líneas celulares tienen un claro patrón de sobreexpresión, los genes que sólo muestran expresión diferencial tras el tratamiento en NCM460 tienen una cantidad similar de genes tanto infra como sobreexpresados.

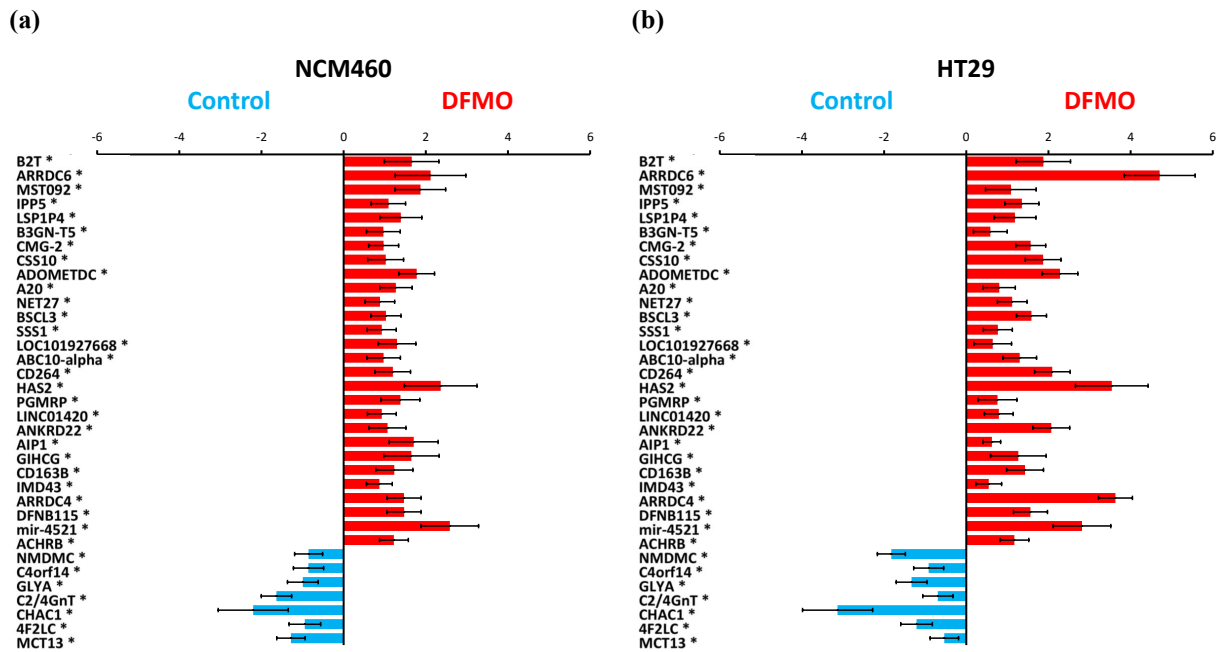


Figura 7.19: DEG tras el tratamiento con DFMO en el mismo sentido en ambas líneas celulares a) NCM460 y b) HT29. Los valores de expresión diferencial, en concreto los valores $\log(FC)$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($FDR < 0.05$) en las células tratadas con DFMO y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $FDR < 0.05$ y "**" $FDR < 0.1$.

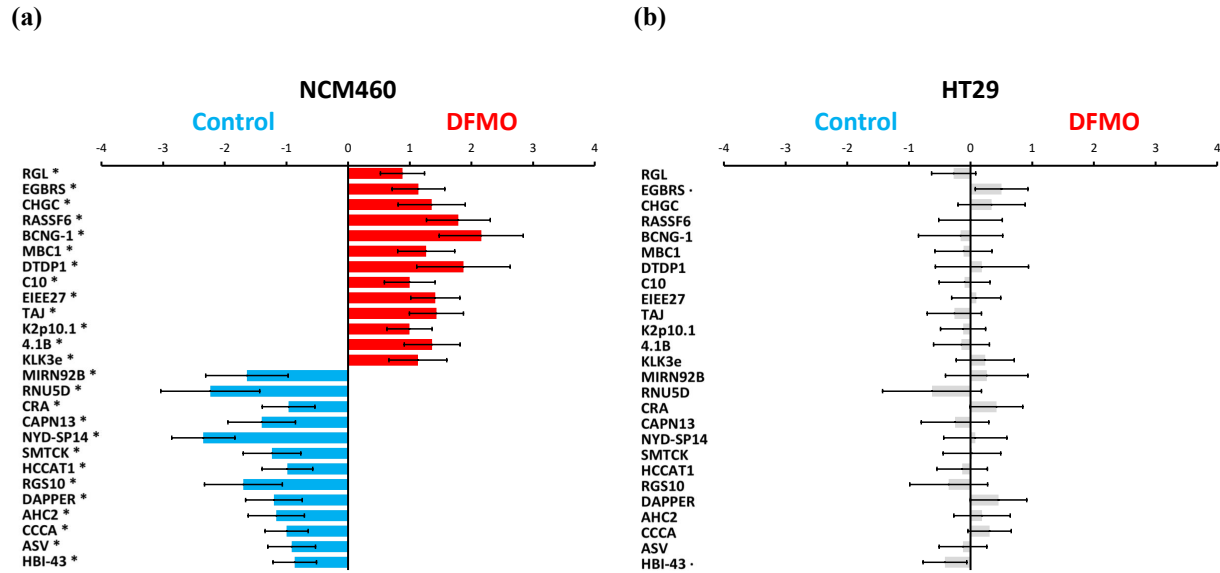


Figura 7.20: DEG tras el tratamiento en NCM460 con un patrón de comportamiento diferente en HT29. a) NCM460 b) HT29. Los valores de expresión diferencial, en concreto los valores $\log(FC)$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($FDR < 0.05$) en las células tratadas con DFMO y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: “*” $FDR < 0.05$ y “.” $FDR < 0.1$.

Puesto que el interés de la presente tesis gira en torno a los genes implicados en la homeostasis del Ca^{2+} intracelular, se van a mostrar los resultados del análisis de expresión diferencial para dichos genes siguiendo el mismo esquema de gráficos de barras que en ocasiones anteriores y divididos en función de las familias en las que se han agrupado cada uno de ellos (figuras 7.21-7.26 y apéndice “B.3.2. Análisis de expresión diferencial de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO”). No obstante, hay que destacar que, con el fin de detectar cambios más sutiles, en estos resultados las barras en rojo o en azul también indican un $FDR < 0.1$ en lugar de $FDR < 0.05$ ⁸.

Al igual que se observa a nivel global (considerando todos los genes), el patrón de expresión tras el tratamiento con DFMO de los genes de la homeostasis del Ca^{2+} intracelular induce, preferentemente, sobreexpresión de algunos de ellos. Además, también sobre este subconjunto de genes el efecto del DFMO parece manifestarse únicamente en las células HT29. En particular, el único gen relacionado con el transporte de Ca^{2+} que muestra una expresión diferencial en las células NCM460 tras el tratamiento con DFMO es el IP3R1, pero lo hace con un FDR inferior a 0.1 y no 0.05. En cuanto a las HT29, en general, se observa, para un $FDR \approx 0.05$, sobreexpresión en genes de SOCE, ERC, CPT y MCT, e infraexpresión en TRP. No obstante, a continuación, se van a analizar estos resultados en más detalle.

⁸Este cambio en las representaciones gráficas no implica ningún cambio en las conclusiones obtenidas, puesto que no hay que estancarse en buscar aquello que sea inferior a un nivel de significación dado, sino ver los datos y resultados en su conjunto.

Un análisis más en detalle muestra que ningún gen de la familia VOCC presenta expresión diferencial tras el tratamiento con DFMO en ninguna de las líneas celulares estudiadas.

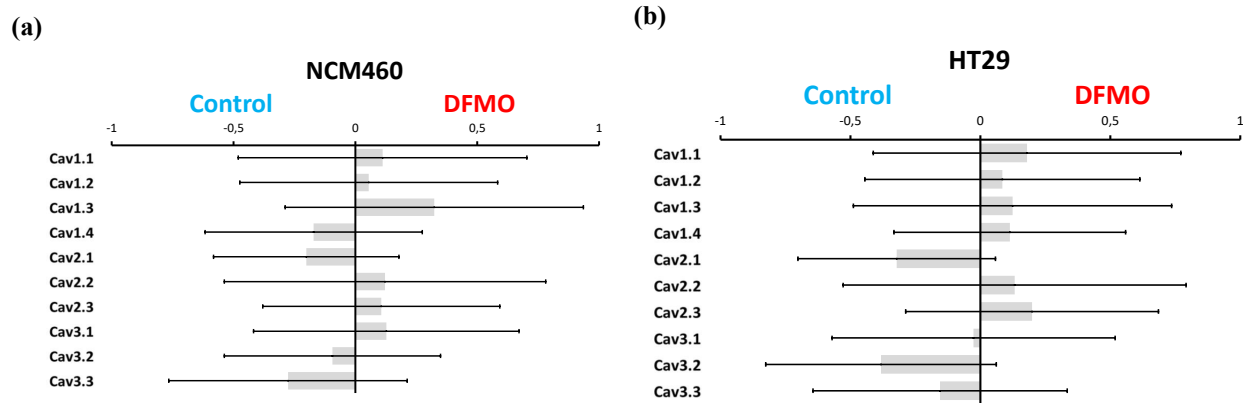


Figura 7.21: Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia VOCC. Los valores de expresión diferencial, en concreto los valores log(FC) estimados mediante el método limma, se representan en un diagrama de barras.

El tratamiento con DFMO no muestra ningún efecto sobre los genes de SOCE en las NCM460 pero sí en las HT29: sobreexpresión de *STIM2*, *ORMDL3*, *SEPTIN6*, *8* y, para un FDR < 0.1, *STIM1*, *CRACR2A* y *SEPTIN9*.

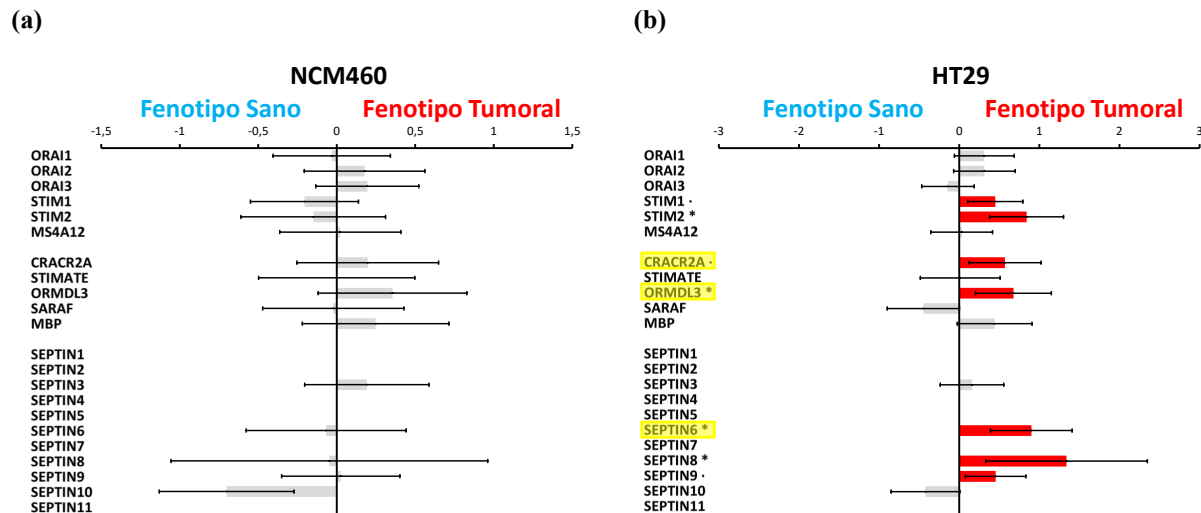


Figura 7.22: Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia SOCE. Los valores de expresión diferencial, en concreto los valores log(FC) estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa (FDR < 0.1) en las células tratadas con DFMO y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*"FDR < 0.05 y "."FDR < 0.1. A la izquierda se representan las células NCM460 (a) y a la derecha las (b) HT29. Los genes subrayados en amarillo son aquellos cuya expresión diferencial entre las HT29 y las NCM460 (apartado 7.1.2.1) se revierte tras el tratamiento con DFMO.

La depleción de poliaminas no induce cambios en los TRP en las células NCM460 pero sí en las HT29, pues disminuyen *TRPC1*, *TRPC5*, *TRPV6* y *TRPP1* y aumenta *TRPP2* (FDR < 0.1).

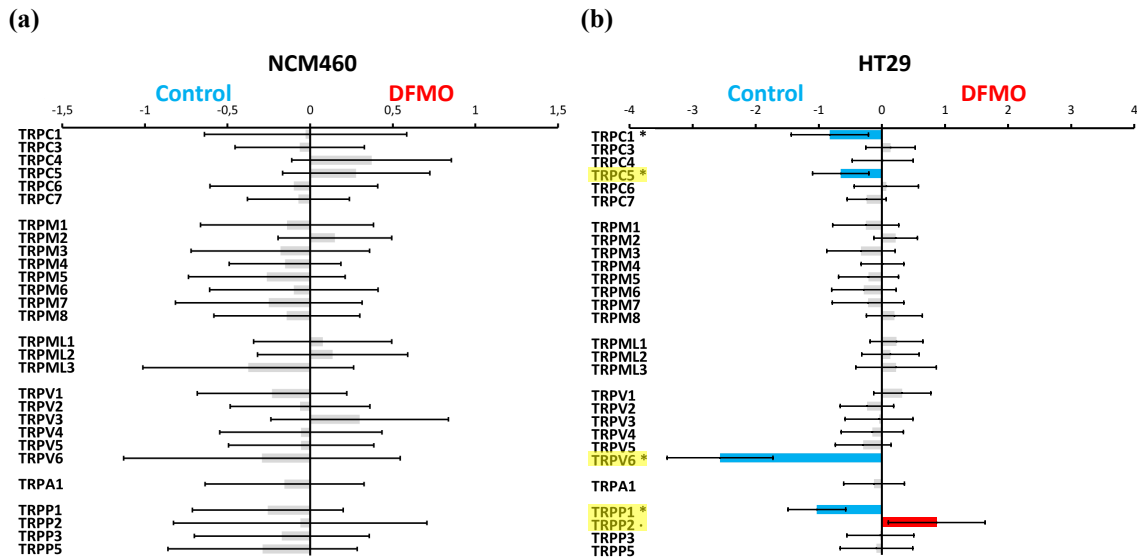


Figura 7.23: Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia TRP. Los valores de expresión diferencial, en concreto los valores log(FC) estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa (FDR < 0.1) en las células tratadas con DFMO y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "**" FDR < 0.05 y "*" FDR < 0.1. A la izquierda se representan las células NCM460 (a) y a la derecha las (b) HT29. Los genes subrayados en amarillo son aquellos cuya expresión diferencial entre las HT29 y las NCM460 (apartado 7.1.2.1) se revierte tras el tratamiento con DFMO.

Para CPT, el DFMO sólo afecta a las HT29: aumenta *PMCA4* y disminuye *SPCA2*.

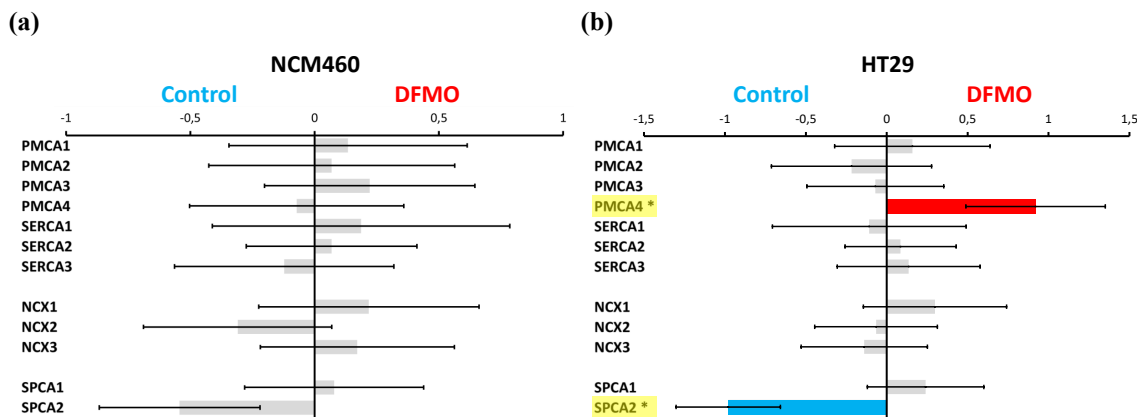


Figura 7.24: Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia CPT. Los valores de expresión diferencial, en concreto los valores log(FC) estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa (FDR < 0.1) en las células tratadas con DFMO y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "**" FDR < 0.05 y "*" FDR < 0.1. A la izquierda se representan las células NCM460 (a) y a la derecha las (b) HT29. Los genes subrayados en amarillo son aquellos cuya expresión diferencial entre las HT29 y las NCM460 (apartado 7.1.2.1) se revierte tras el tratamiento con DFMO.

A diferencia de lo que se ha observado en los genes de SOCE, en los genes de la familia ERC se observa un patrón contrario, pues en estos el DFMO tiene un ligero efecto sobre las células *NCM460* al aumentar la expresión de *IP3R1* ($0.05 < \text{FDR} < 0.1$), pero no sobre las HT29.

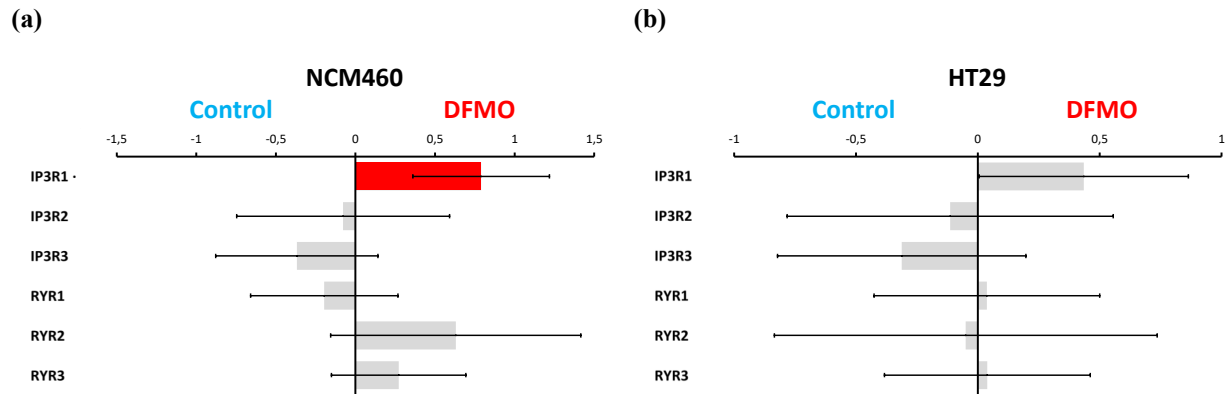


Figura 7.25: Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia ERC. Los valores de expresión diferencial, en concreto los valores $\log(\text{FC})$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($\text{FDR} < 0.1$) en las células tratadas con DFMO y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $\text{FDR} < 0.05$ y "." $\text{FDR} < 0.1$. A la izquierda se representan las células **NCM460** (a) y a la derecha las (b) **HT29**.

Siguiendo el mismo patrón de comportamiento que para el resto de familias estudiadas, a excepción de ERC y VOCC, la expresión de los genes de la familia MCT sólo se ve afectada por el DFMO en las células HT29, pero no en las NCM460, pues hay un aumento en la expresión de *MCU* y *VDAC3*.

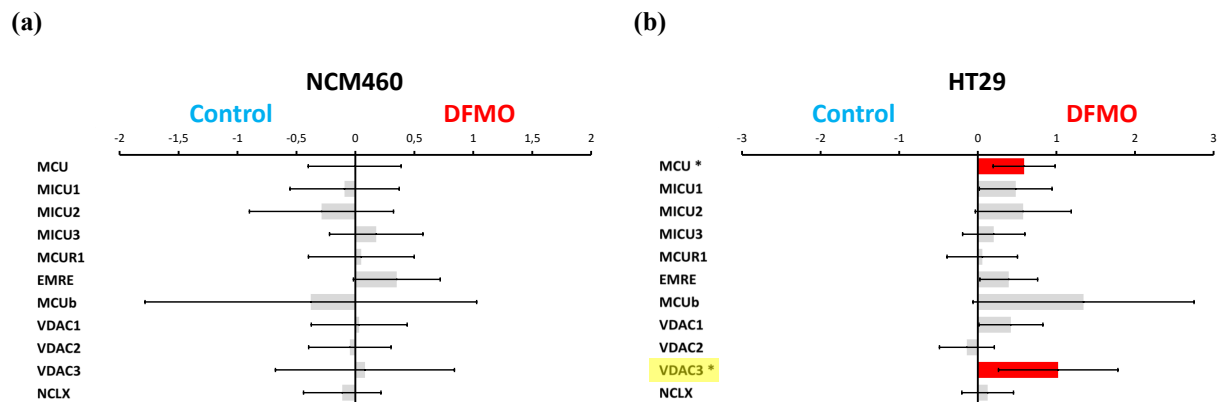


Figura 7.26: Análisis de la expresión diferencial entre las células HT29 o NCM460 tratadas con DFMO vs. Control para los genes de la familia MCT. Los valores de expresión diferencial, en concreto los valores $\log(\text{FC})$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($\text{FDR} < 0.1$) en las células tratadas con DFMO y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $\text{FDR} < 0.05$ y "." $\text{FDR} < 0.1$. A la izquierda se representan las células **NCM460** (a) y a la derecha las (b) **HT29**. Los genes subrayados en amarillo son aquellos cuya expresión diferencial entre las HT29 y las NCM460 (apartado 7.1.2.1) se revierte tras el tratamiento con DFMO.

En conclusión, se observa que el efecto de la depleción de las poliaminas es prácticamente selectivo sobre las células de fenotipo tumoral HT29, ya que en las células NCM460 únicamente afecta a la expresión de 62 genes (165 para un FDR < 0.1) frente a los 3394 (4402 para un FDR < 0.1) que se expresan diferencialmente en las HT29 tras el tratamiento con DFMO. En concreto, mientras que para los genes relacionados con la homeostasis del Ca^{2+} intracelular sólo se observa la sobreexpresión de *IP3R1* en las NCM460 para un FDR < 0.1, en las HT29 se identifican 12 para un FDR < 0.05 y 16 para un FDR < 0.1. Por lo tanto, esta depleción de poliaminas induce, entre otras, la alteración de la homeostasis del Ca^{2+} intracelular a nivel transcriptómico de forma selectiva sobre las células HT29 (figura 7.27).

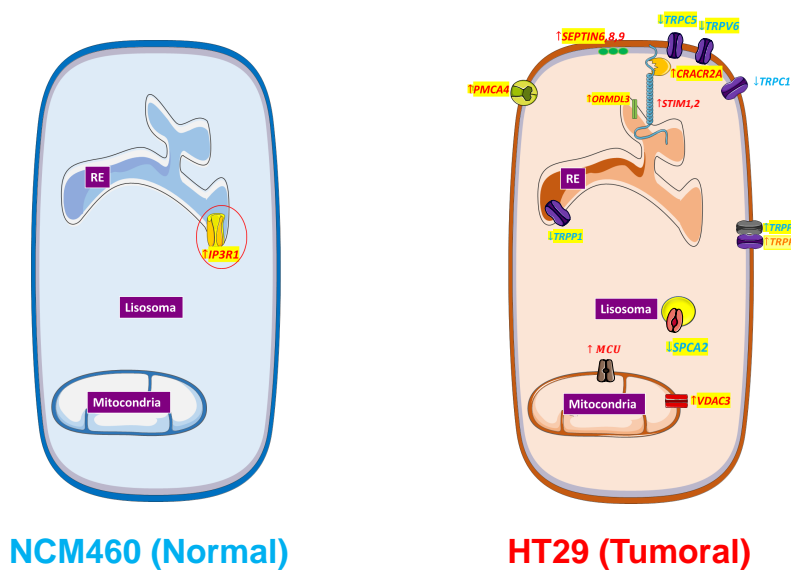


Figura 7.27: Resumen del análisis de expresión diferencial de los genes implicados en el transporte de Ca^{2+} tras la depleción de poliaminas en cada una de las líneas celulares HT29 y NCM460. Sobre cada una de las células se representan los cambios inducidos por el tratamiento con DFMO en cada una de las líneas celulares. Además, los genes subrayados en amarillo son aquellos cuya expresión diferencial observada entre las células HT29 y NCM460 en condiciones control se revierte tras el tratamiento con DFMO.

Finalmente, con el fin de resumir los análisis transcriptómicos anteriores realizados sobre los modelos celulares, a continuación, se muestra un esquema sobre el que se muestra tanto la expresión diferencial entre las líneas celulares como el efecto del DFMO en cada una de ellas (figura 7.28).

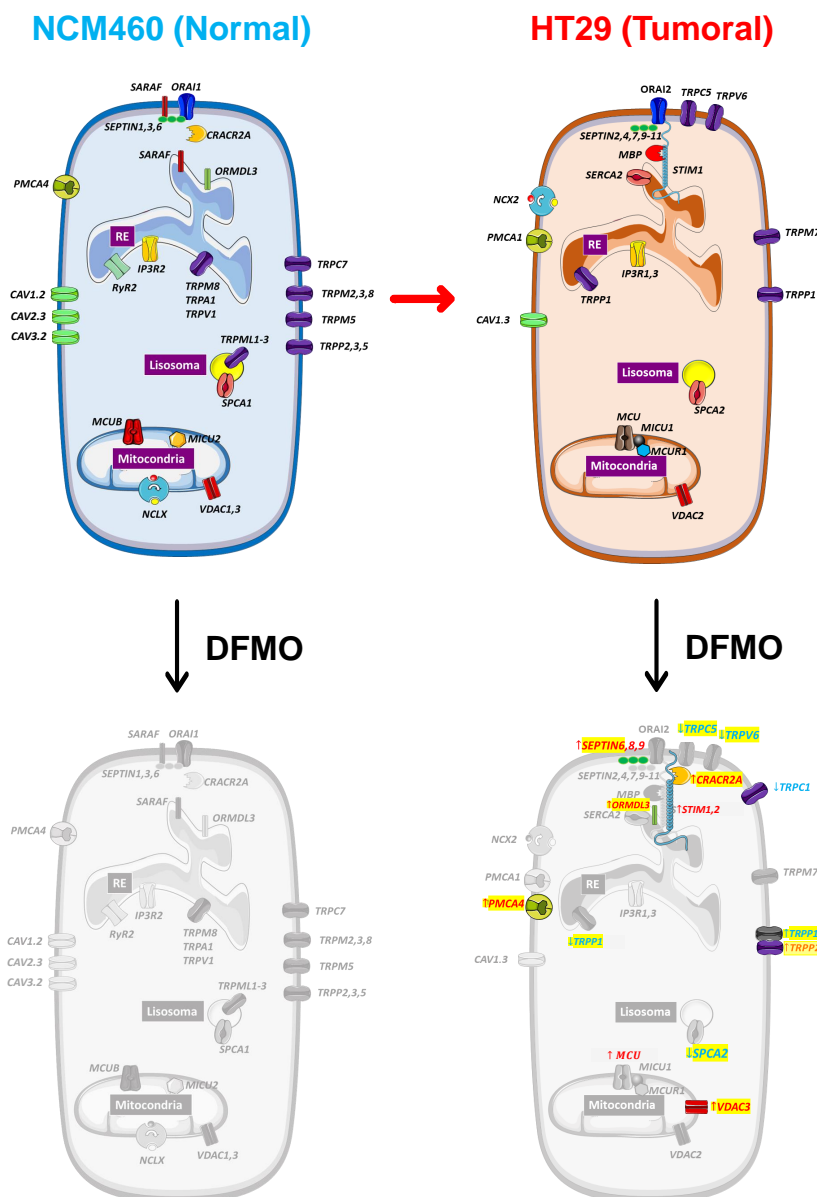


Figura 7.28: Resumen del análisis de expresión diferencial de los genes implicados en el transporte de Ca^{2+} intracelular entre las células HT29 vs NCM460 y entre las células tratadas con DFMO y sin tratar. En la parte superior se muestra la expresión diferencial en la comparación entre las células HT29 (derecha) y NCM460 (izquierda) como modelos de células de cáncer de colon y normales, respectivamente. Sobre la figura correspondiente a HT29 se muestra los genes cuya expresión es superior respecto de las NCM460, mientras que sobre la de NCM460 se representan los genes cuya expresión es menor en HT29 y, por ende, superior en las NCM460. En la parte inferior, sobre el resumen de expresión diferencial entre HT29 y NCM460 (en blanco y negro) se superponen los cambios inducidos por el tratamiento con DFMO en cada una de las líneas celulares. Además, los genes subrayados en amarillo son aquellos cuya expresión diferencial observada entre las células HT29 y NCM460 en condiciones control se revierte tras el tratamiento con DFMO.

7.1.3.2.2. Análisis de significación biológica del efecto del DFMO sobre las líneas celulares HT29 y NCM460

Tras realizar el análisis de expresión diferencial, se procedió a buscar un significado biológico a la lista de DEG, lo que se conoce como análisis de significación biológica. En primer lugar, se realizó una búsqueda de patrones a partir del estudio de los perfiles de expresión mediante mapas de calor y un algoritmo de aprendizaje no supervisado -agrupa muestras y genes sin necesidad de información previa sobre ellas-. En segundo lugar, se realizó el análisis de enriquecimiento con el fin de encontrar conjuntos de genes asociados con el efecto del tratamiento con DFMO dentro de cada una de las líneas celulares.

7.1.3.2.2.1. Perfiles de expresión de las líneas celulares de cáncer de colon y normales tras el tratamiento con DFMO

Mediante la ayuda de un mapa de calor se han representado las expresiones de los genes (considerando sólo aquellos DEG en alguno de los contrastes realizados) de las diferentes muestras analizadas y, además, mediante el algoritmo de clasificación no supervisado *análisis clúster jerárquico con el método de Ward*, tanto las muestras como los genes se han agrupado según su mayor o menor similitud tener en cuenta a qué condición experimental se corresponde cada una de ellas (figura 7.29). Como puede observarse, se forman dos grupos iniciales en las que las muestras se agrupan en función de la línea celular y, dentro de cada uno de estos, otro subgrupo en función del tratamiento con DMFO. Además, se puede ver como el factor aleatorio *Día* no influye en estas agrupaciones. Esto es un buen resultado, pues indica que la calidad de las muestras es buena y que la mayor fuente de variabilidad es el tipo de línea celular. Por otro lado, los genes se agrupan en dos conjuntos: un conjunto más reducido que muestra una sobreexpresión en las muestras correspondientes a las células HT29 y otro más grande que, a su vez, puede subdividirse en otros dos conjuntos de genes de un tamaño similar, de tal manera que uno muestra una sobreexpresión en las HT29 y otro infraexpresión. No obstante, las diferencias en los perfiles de expresión entre las células tratadas y no tratadas con DFMO dentro de la misma línea celular son menos evidentes aunque, lógicamente, han de estar ahí puesto que el algoritmo de aprendizaje no supervisado las ha logrado separar acorde a dicho factor (figura 7.29). Para consultar el número de DEG tras el tratamiento con DFMO en cada una de las líneas celulares y el sentido del cambio, consultar los diagramas de Venn de la figura 7.18.

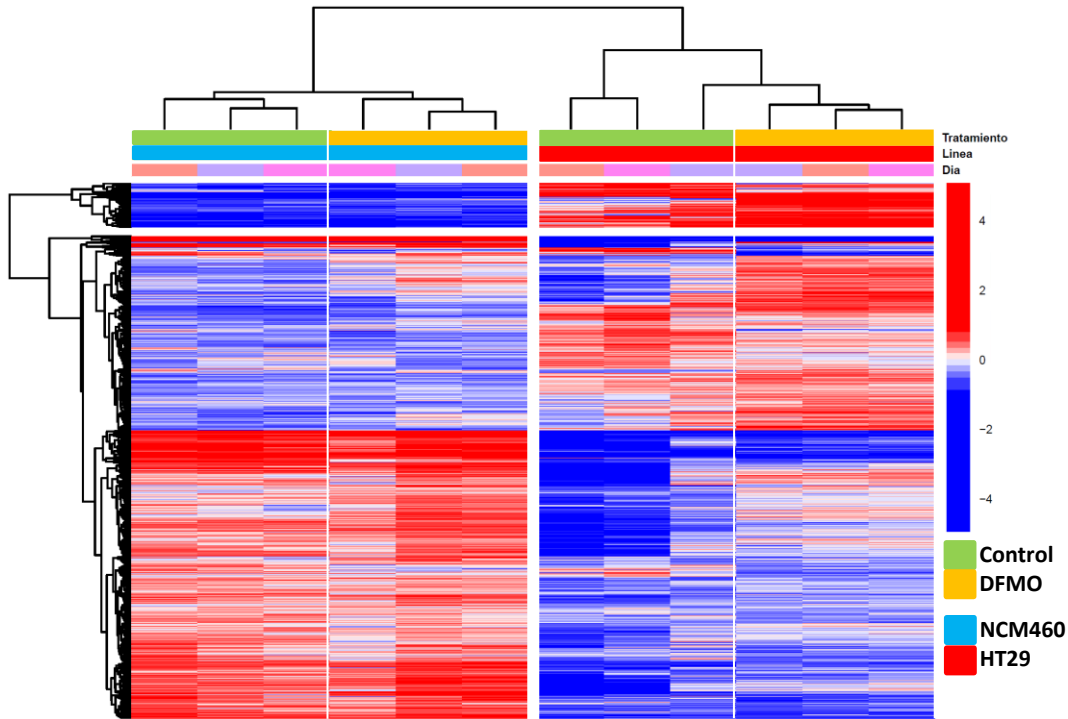


Figura 7.29: Perfiles de expresión. En el mapa de calor se representan los valores centrados de la expresión en escala logarítmica. En el eje X se agrupan las diferentes muestras mediante el análisis clúster jerárquico basado en el método de Ward, y en el eje Y los DEG. El rojo indica sobreexpresión de un gen (fila) en una muestra dada (columna) respecto de la media de los valores de expresión (log) de ese mismo gen a lo largo de todas las muestras.

7.1.3.2.2.2. Análisis de enriquecimiento de las líneas celulares de cáncer de colon y normales tras el tratamiento con DFMO

Tras el análisis de expresión diferencial y a partir de sus resultados, se realizó el análisis de enriquecimiento para buscar, en cada una de las líneas celulares, qué conjuntos podrían estar asociados con el efecto del tratamiento con DFMO. Para ello, se cruzaron los datos de expresión diferencial en cada una de las líneas celulares con las bases de datos KEGG PATHWAY, GO y MSigDB, las cuales contienen alrededor de 30000 conjuntos, un número demasiado elevado como para profundizar en el significado de cada uno de ellos, pues eso estaría fuera del alcance de la presente tesis. Por ello, el análisis de estos datos se ha centrado en obtener un resumen numérico de cuántos conjuntos en total rechazan alguna de las dos hipótesis nulas ($FDR < 0.05$) y, de estos, cuántos pertenecen a las bases de datos relacionadas con el cáncer (MCC y MOG) y cuántos contienen en su nombre alguna de las siguientes palabras clave de interés para la temática de la tesis: *calcium*, *mitochondria*, *reticulum*, *cancer*, *wnt* y *polyamines*.

En primer lugar, se muestra un resumen del análisis de conjuntos de genes para el efecto del DFMO en las células NCM460 (tabla 7.5). Así, de los casi 30000 conjuntos de genes analizados, sólo 380/8 rechazan las hipótesis nula autocontenida/competitiva, de los cuales 61/2 pertenecen a la base de datos MCC, 4/1 pertenecen a MOG y 19/0 contienen la palabra de interés *cancer* (ver apéndice “B.3.3.1. Análisis de enriquecimiento de genes de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las células NCM460”). Algunos de ellos están estrechamente relacionados con la respuesta inmune (M8594 GSE27859 MACROPHAGE VS CD11C INT F480 INT DC UP...), con el funcionamiento de los ribosomas (hsa03008 Ribosome biogenesis in eukaryotes, GO:0003735 structural constituent of ribosome...) y con el propio CCR (M16586 LAIHO COLORECTAL CANCER SERRATED UP, M39740 WP METABOLIC REPROGRAMMING IN COLON CANCER...).

Tabla 7.5: Análisis de enriquecimiento de genes para el tratamiento con DFMO en NCM460.

| Base de Datos | Conjuntos | | Métodos | Conjuntos Significativos (FDR <0.05) | | | | | | |
|---------------|--------------|---------------------|---------|--------------------------------------|----------------------|---------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | | Totales | Calcium ¹ | Micochondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO BP | 12342 | 146 | CAMERA | 19 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO MF | 4245 | 66 | CAMERA | 7 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| MH | 50 | 0 | CAMERA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MCDB | 6290 | 12 | CAMERA | 222 | 0 | 0 | 0 | 19 | 0 | 0 |
| | | | ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MCC | 858 | 0 | CAMERA | 61 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| MOG | 189 | 0 | CAMERA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| MIM | 5219 | 0 | CAMERA | 48 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 29537 | 225 | CAMERA | 380 | 0 | 0 | 0 | 19 | 0 | 0 |
| | | | ORA | 8 | 0 | 0 | 0 | 0 | 0 | 0 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Onncogenic; MIM: Msigdb Immunology.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

En segundo lugar, se analizan los conjuntos de genes para el efecto del DFMO en las HT29 (tabla 7.6). De los 740/2160 conjuntos que rechazan las hipótesis autocontenida/competitiva, 120/131 pertenecen a la base de datos MCC, 4/15 a MOG, ninguno contiene la palabra *calcium*, 6/9 contienen la palabra *mitochondria*, 4/1 *reticulum*, 16/55 *cancer*, 3/5 *WNT* y 1/1 *polyamine* (ver apéndice B.3.3.2). Algunos de estos conjuntos están relacionados con el procesamiento de proteínas en el RE (hsa04141 Protein processing in endoplasmic reticulum, GO:0030968 endoplasmic reticulum unfolded protein response), con la apoptosis y el estrés del RE (GO:0070059 intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress, GO:0034976 response to endoplasmic reticulum stress...), el transporte de electrones mitocondrial (GO:0006120 mitochondrial electron transport, NADH to ubiquinone...), el transporte de proteínas a la mitocondria (M39781 WP MITOCHONDRIAL COMPLEX I ASSEMBLY MODEL OXPHOS SYSTEM...), la expresión de genes mitocondriales (M39544 WP MITOCHONDRIAL GENE EXPRESSION), el CCR y rutas relacionadas con este (GO:0060071 Wnt signaling pathway, GO:0090090 negative regulation of canonical Wnt signaling pathway, M16586 LAIHO COLORECTAL CANCER SERRATED UP, M16740 GRADE COLON AND RECTAL CANCER UP, M16586 LAIHO COLORECTAL CANCER SERRATED UP...) y la síntesis de poliaminas (M747 REACTOME METABOLISM OF POLYAMINES).

Tabla 7.6: Análisis de conjuntos de genes para el tratamiento con DFMO en HT29.

| Base de Datos | Conjuntos | | Conjuntos Significativos (FDR <0.05) | | | | | | | |
|---------------|--------------|---------------------|--------------------------------------|-------------|---------------------|---------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | Métodos | Totales | Calcio ¹ | Micochondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA | 13 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | | ORA | 21 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO BP | 12342 | 146 | CAMERA | 38 | 0 | 2 | 3 | 0 | 0 | 0 |
| | | | ORA | 35 | 0 | 2 | 0 | 0 | 2 | 0 |
| GO MF | 4245 | 66 | CAMERA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| MH | 50 | 0 | CAMERA | 7 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 8 | 0 | 0 | 0 | 0 | 0 | 0 |
| MCDB | 6290 | 12 | CAMERA | 447 | 0 | 4 | 0 | 16 | 3 | 1 |
| | | | ORA | 791 | 0 | 7 | 1 | 55 | 2 | 1 |
| MCC | 858 | 0 | CAMERA | 120 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 131 | 0 | 0 | 0 | 0 | 0 | 0 |
| MOG | 189 | 0 | CAMERA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| MIM | 5219 | 0 | CAMERA | 107 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 1167 | 0 | 0 | 0 | 0 | 1 | 0 |
| Total | 29537 | 225 | CAMERA | 740 | 0 | 6 | 4 | 16 | 3 | 1 |
| | | | ORA | 2160 | 0 | 9 | 1 | 55 | 5 | 1 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Onncogenic; MIM: Msigdb Immunology.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

Como puede comprobarse a partir de estos resultados, mientras que en las células NCM460, modelo de colon normal, el tratamiento con DFMO no tiene efecto alguno sobre potenciales rutas de señalización diana para atacar el CCR, como son las rutas de señalización relacionadas con el CCR, con el estrés del RE, la apoptosis o la propia síntesis de poliaminas, en las HT29, modelo de CCR, el efecto del tratamiento con DFMO es claro sobre estas. Por lo tanto, todo apunta a que este tratamiento podría ser selectivo sobre las células cuyo fenotipo sea de CCR y no afectar a las normales. Con el fin de aportar al lector interesado una información más detallada de estos análisis, por un lado, en el apéndice “B.3.3.1. Análisis de enriquecimiento de genes de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las células NCM460” se muestran, para el efecto del DFMO en NCM460, todos los conjuntos de genes con un FDR < 0.1 que contienen alguna de las palabras de interés, divididos en diferentes tablas en función tanto de la base de datos a la que pertenecen como del método empleado para su análisis. Por otro lado, para el efecto del DFMO en HT29, en el apéndice “B.3.3.2. Análisis de enriquecimiento de genes de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las células HT29” se muestran aquellos conjuntos con un FDR < 0.1 que contienen alguna de las palabras de interés.

7.2. Análisis del remodelado de la homeostasis del Ca^{2+} intracelular en tejido humano de cáncer de colon

El uso de modelos celulares es un buen punto de partida para estudiar diferentes patologías, como es el caso del CCR. No obstante, los modelos celulares tienen ciertas limitaciones en el sentido de que no siempre se acercan lo suficiente a la realidad. Por ello, tras haber observado el evidente remodelado de la homeostasis del Ca^{2+} intracelular en el CCR en los modelos celulares HT29 -tumoral- y NCM460 -normal-, el siguiente paso lógico fue usar unos modelos más cercanos a la realidad del CCR humano. Por un lado, se emplearon 20 muestras pareadas de tejido humano de CCR y normales, donadas por 10 pacientes diferentes a través del Banco de Tumores de Castilla y León, para realizar un experimento transcriptómico de alto rendimiento mediante los microarrays Clariom D Human de Affymetrix.

Por otro lado, se extrajeron del repositorio público NCBI GEO los datos crudos *.CEL*, sin procesar, de diferentes experimentos transcriptómicos de tejido CCR humano basados en microarrays y se reanalizaron desde cero. Posteriormente, los datos resultantes de estos reanálisis junto con los datos del experimento transcriptómico realizado por nuestro grupo en muestras de tejido humano de CCR pareadas se emplearon para realizar un metaanálisis del remodelado de la homeostasis del Ca^{2+} intracelular en el CCR.

7.2.1. Análisis transcriptómico de muestras pareadas de tejido humano de cáncer de colon

Partiendo de 20 muestras pareadas, cedidas por el Banco de Tumores de Castilla y León, de 10 pacientes diferentes, se extrajo su ARN siguiendo el protocolo sugerido por la casa comercial según el microarray escogido (Clariom D Human de Affymetrix en este caso) y se envió a la empresa encargada de realizar dichos microarrays, CABIMER. Tras recibir los datos crudos en formato *.CEL*, estos fueron sometidos a un control de calidad mediante métodos tanto generales (ACP, gráficos de cajas y bigotes...) como específicos de microarrays (NUSE...) para eliminar los posibles microarrays problemáticos o *outliers*, lo que resultó en la eliminación de 4 de ellos (ver apéndice “B.2.0.1. Análisis exploratorio y control de calidad”).

Posteriormente, fueron preprocesados mediante el método RMA para los posteriores análisis transcriptómicos (ver apéndice “B.2.0.2. Preprocesado de los datos”). Tras el preprocesado, se llevaron a cabo los análisis de expresión diferencial y de significación biológica mostrados en los siguientes subapartados. No obstante, en el apéndice “B.2. Caso práctico: análisis de microarrays de las muestras pareadas de tejido humano de cáncer de colon” se muestra el análisis de este experimento en mayor detalle con el fin de proporcionar un ejemplo práctico de análisis de datos de microarrays aplicable al resto de análisis de este tipo de datos realizado a lo largo de toda la tesis.

7.2.1.1. Análisis de expresión diferencial

Debido a la gran dimensión de este tipo de datos (decenas de miles de sondas), fue necesario realizar un filtrado no específico para eliminar aquellas sondas que no se expresaban, que no variaban de una forma evidente su expresión a lo largo de los diferentes microarrays -sin tener en cuenta ninguna covariable o factor- y/o que no disponían de una correspondencia con la anotaciones usadas en los análisis de expresión diferencial -ALIAS- o de enriquecimiento -UNIGENE-. Además, puesto que los diferentes análisis se han centrado en los genes relacionados con el transporte de Ca^{2+} intracelular, las sondas correspondientes a estos se han mantenido tras el filtrado independientemente de su expresión. Como resultado, de un conjunto inicial de datos de 138745 sondas se ha obtenido uno de 69439 sondas diferentes de las cuales 12835 tienen correspondencia con la anotación UNIGENE y 12944 con ALIAS.

Para llevar a cabo el análisis de expresión diferencial entre las muestras de tejido humano de CCR y normal se ha empleado el método *limma*. A diferencia de otros estudios transcriptómicos de expresión diferencial realizados anteriormente, en este caso, aparte de los factores fijos de interés (*Tejido* a dos niveles: Normal y Tumoral), también hay otras posibles variables explicativas que, en caso de ser necesario, han de ser consideradas en el modelo lineal realizado con *limma*. En concreto, estos potenciales predictores son el factor fijo *Sexo* y el factor aleatorio *Paciente*. Por lo tanto, al igual que en los modelos lineales mixtos de los experimentos de imagen de Ca^{2+} para el estudio del efecto del DFMO, en este caso también se han planteado diferentes modelos mixtos y se ha seleccionado el más adecuado. En concreto, se ha obtenido que el modelo más adecuado, que se muestra en la ecuación 7.1, es aquel que solo contiene el factor *Tumor* en la parte fija y el factor aleatorio *Paciente*, mostrado en la ecuación 7.1 (ver apéndice B.2.0.3).

$$\log(\text{expresin})_{ik} = \text{Tejido}_i \alpha_i + A_k + \epsilon_{ik} \quad (7.1)$$

Donde los parámetros del modelo son: α_i el efecto del nivel $i \in \{\text{NORMAL}, \text{TUMORAL}\}$ del factor **Tejido**; A_k el efecto aleatorio del paciente k , y ϵ_{ik} es el error aleatorio correspondiente a la muestra del tejido i del paciente k . Así, el modelo anterior se usó en el método *limma* y se llevó a cabo el análisis de expresión diferencial entre el tejido tumoral de CCR y normal y, posteriormente, para identificar a un gen como diferencialmente expresado se corrige el p-valor resultante de dicho análisis de expresión diferencial controlando **tasa de hipótesis falsamente rechazadas (FDR)** propuesto por Benjamini and Hochberg (1995). No obstante, debido a la gran variabilidad que presentan los datos, el nivel de significación escogido en este caso ha sido $\alpha_{\text{FDR}} = 0.1$ ⁹, pues usar un

⁹En otros análisis de expresión diferencial realizados en la tesis el nivel de significación se ha fijado en $\alpha_{\text{FDR}} = 0.05$. Sin embargo, estudiando la gran variabilidad de los datos de este experimento, se ha considerado que una mejor opción es fijar dicho nivel en $\alpha_{\text{FDR}} = 0.1$ para que, a pesar de cometer algún falso positivo más, detectar más genes diferencialmente expresados. Además, puesto que controlar la *tasa de hipótesis falsamente rechazadas (FDR)* significa que del total de DEG se espera un $\alpha_{\text{FDR}} \cdot 100$ % de falsos positivos, dado el bajo número de DEG se ha considerado aceptable que el 10 % de los DEG sean falsos positivos frente a un 5 % que se obtendría de un modo más convencional. Es importante siempre llegar a una situación de consenso entre la ganancia de potencia del test y el error de tipo I (falsos positivos) cometidos.

nivel de significación inferior, 0.05, podría enmascarar genes que realmente estén diferencialmente expresados aunque, como precio a pagar, la probabilidad de cometer falsos positivos aumentará.

Como resultado del análisis de expresión diferencial, de las 12944 sondas analizadas (las que han quedado tras el filtrado) se han identificado 4435 sondas diferencialmente expresadas para un $\alpha_{FDR} = 0.05$ y 5649 para un $\alpha_{FDR} = 0.1$. Por lo tanto, es evidente que la potencia que se ha ganado (rechazar más hipótesis alternativas verdaderas) al identificar 1214 sondas a mayores al fijar $\alpha_{FDR} = 0.1$ en lugar de $\alpha_{FDR} = 0.05$ compensa los falsos positivos esperados (222 para $\alpha_{FDR} = 0.05$ frente a 565 fijando $\alpha_{FDR} = 0.1$). Como curiosidad, si no se hubiese tenido en cuenta el factor aleatorio, para $\alpha_{FDR} = 0.1$ se hubieran identificado sólo 4420 sondas como diferencialmente expresadas. A continuación, puesto que el interés de la presente tesis gira en torno a los genes implicados en la homeostasis del Ca^{2+} intracelular, sólo se mostrarán, mediante diagramas de barras, los resultados del análisis de expresión diferencial para estos, divididos en función de la familia a la que pertenezcan.

En la figura 7.30 se puede observar cómo ninguno de los genes pertenecientes a la familia VOCC muestra expresión diferencial entre el tejido tumoral de CCR y el normal. Curiosamente, en los modelos celulares sí que se observó una mayor expresión de *CAVI.3* en el fenotipo tumoral (HT29) y una disminución de *CAVI.2*, *CAV2.3* y *CAV3.2*.

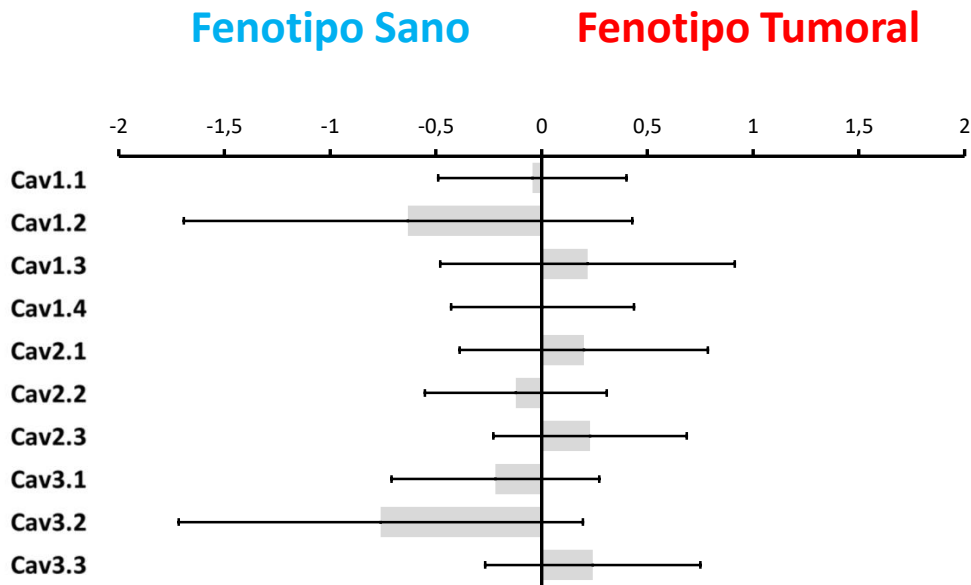


Figura 7.30: Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes VOCC. En el diagrama de barras se representan los valores de expresión diferencial $\log(FC)$ estimados mediante el método limma para 16 muestras pareadas (8 de cada condición experimental). Las barras azules hacia la izquierda indican una menor expresión diferencial significativa ($FDR < 0.1$) en el tejido tumoral y mayor para las barras rojas hacia la derecha. Las barras azules hacia la izquierda indican una menor expresión significativa ($FDR < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $FDR < 0.05$ y "." $FDR < 0.1$.

Para los genes de SOCE, se puede observar cómo sólo el gen *MS4A12* está significativamente infraexpresado en el tejido tumoral. Respecto de los reguladores de SOCE, se identifica la sobreexpresión de *ORMDL3* y *SEPTIN8,11* a la par de una disminución de *SEPTIN1,6* (figura 7.31). Estos resultados muestran un patrón de expresión diferencial algo diferente al observado en las líneas celulares, pues a pesar de que la expresión diferencial para las *SEPTIN1,6,11* es el mismo en ambos modelos, en los modelos no se observa expresión diferencial de *ORAI1,2*, *STIM1*, *CRACR2A*, *SARAF* ni *SEPTIN2,4,7,9,10*. Además, mientras que en las líneas celulares se observó una menor expresión del regulador negativo de SOCE *ORMDL3*, en los tejidos de cáncer de colon se observa una mayor expresión de este. Curiosamente, en cuanto a la sobreexpresión de *SEPTIN8* encontrada en estos tejidos, mientras que con Illumina se identificó una mayor expresión de este gen en las HT29, con los microarrays fue en el sentido contrario.

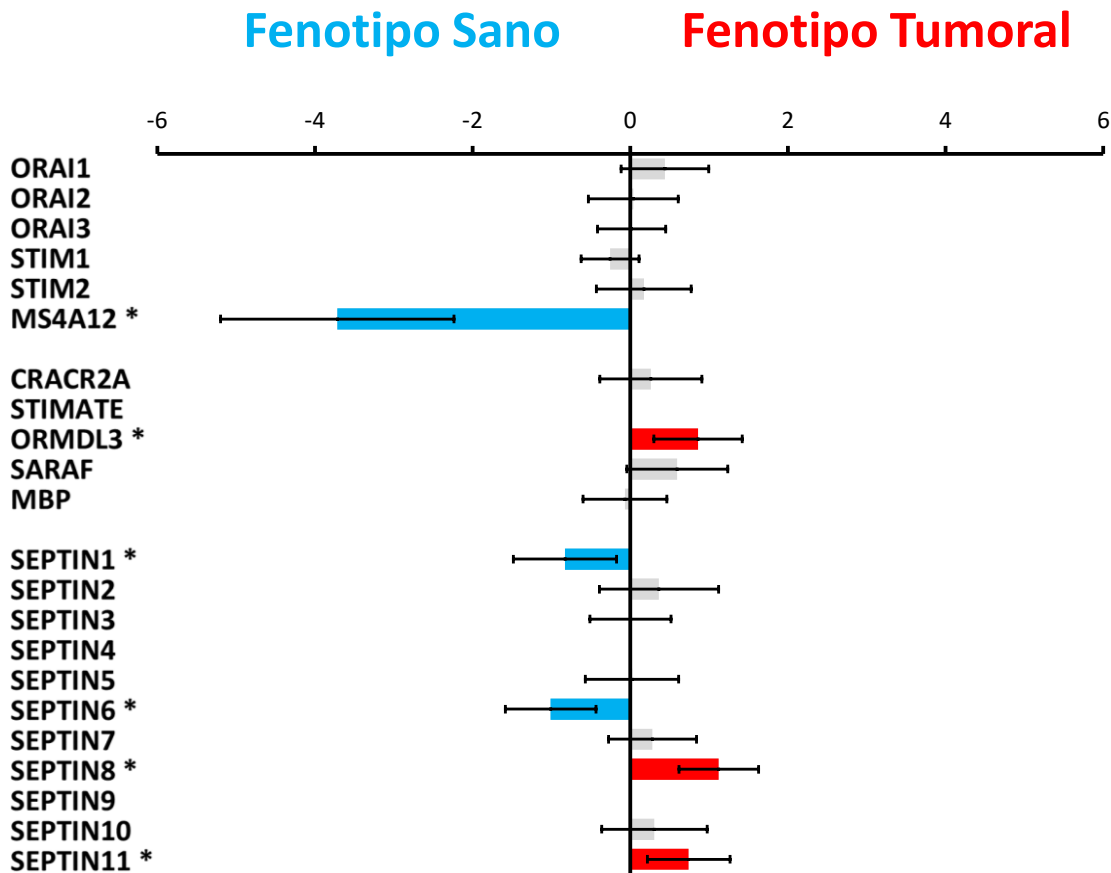


Figura 7.31: Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes SOCE. En el diagrama de barras se representan los valores de expresión diferencial $\log_2(FC)$ estimados mediante el método limma para 16 muestras pareadas (8 de cada condición experimental). Las barras azules hacia la izquierda indican una menor expresión diferencial significativa ($FDR < 0.1$) en el tejido tumoral y mayor para las barras rojas hacia la derecha. Las barras azules hacia la izquierda indican una menor expresión significativa ($FDR < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $FDR < 0.05$ y " " $FDR < 0.1$.

Respecto de los genes que codifican para los TRP, se ha detectado tanto una infraexpresión diferencial de los genes *TRPM2* y *TRPP5* como una sobreexpresión del gen *TRPV1* (figura 7.32). Es muy llamativo el hecho de que, mientras que en los modelos celulares había una evidente expresión diferencial entre el fenotipo tumoral (HT29) y normal (NCM460), pues se han identificado 16 miembros de TRP diferencialmente expresados entre ambos fenotipos cuya expresión en el fenotipo tumoral era prácticamente menor que en el normal, en los tejidos únicamente se han detectado 3 miembros diferencialmente expresados entre el tejido de cáncer de colon y el normal de los cuales coincide la menor expresión de *TRPM2* y de *TRPP5*, pero no la sobreexpresión de *TRPV1* la cual era menor en las células de fenotipo tumoral para un FDR < 0.1.

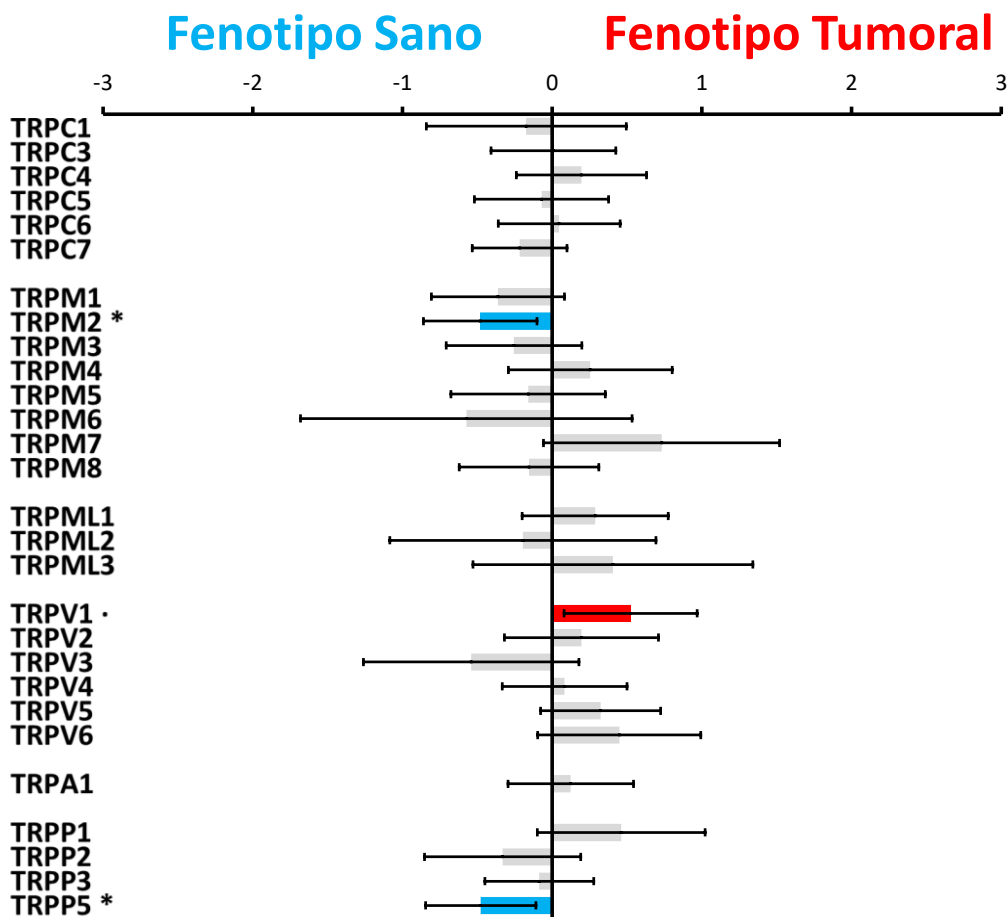


Figura 7.32: Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes TRP. En el diagrama de barras se representan los valores de expresión diferencial $\log(\text{FC})$ estimados mediante el método limma para 16 muestras pareadas (8 de cada condición experimental). Las barras azules hacia la izquierda indican una menor expresión diferencial significativa ($\text{FDR} < 0.1$) en el tejido tumoral y mayor para las barras rojas hacia la derecha. Las barras azules hacia la izquierda indican una menor expresión significativa ($\text{FDR} < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*"FDR < 0.05 y "."FDR < 0.1.

En cuanto a los genes de la familia CPT, se observa cómo tanto la *SERCA2* como las *SPCA1* y *SPCA2* muestran sobreexpresión en el tejido CCR (figura 7.33). Curiosamente, en los modelos celulares tanto *SERCA2* como *SPCA1* están menos expresadas en el fenotipo tumoral. Además, en los tejidos no se observa expresión diferencial de *PMCA1,4* ni *NCX2*.

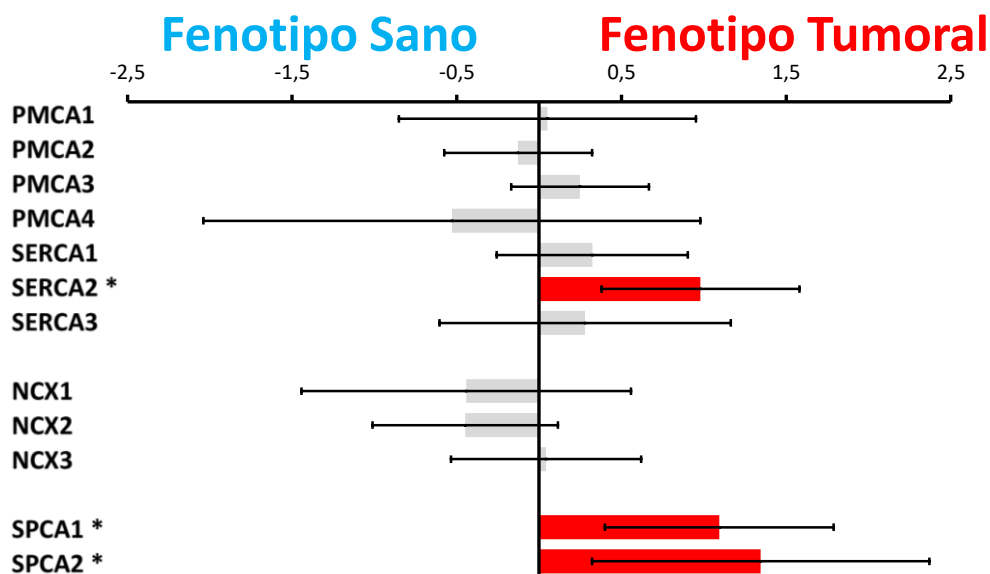


Figura 7.33: Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes CPT. En el diagrama de barras se representan los valores de expresión diferencial $\log(FC)$ estimados mediante el método limma para 16 muestras pareadas (8 de cada condición experimental). Las barras azules hacia la izquierda indican una menor expresión diferencial significativa ($FDR < 0.1$) en el tejido tumoral y mayor para las barras rojas hacia la derecha. Las barras azules hacia la izquierda indican una menor expresión significativa ($FDR < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $FDR < 0.05$ y "**" $FDR < 0.1$.

Respecto a los genes de la familia ERC, en la figura 7.34 se puede observar cómo, de entre todos ellos, sólo el *IP3R3* muestra una sobreexpresión diferencial en el tejido tumoral, si bien es cierto que el *IP3R1* podría estar infraexpresado, pues su $FDR = 0.079$. De estos dos genes, la sobreexpresión de *IP3R3* está de acuerdo con lo observado en los modelos celulares, pero no la expresión de *IP3R1*, pues en dichos modelos se observó un aumento de este canal en las células de fenotipo tumoral. Además, en los tejidos no se ha observado expresión diferencial de *IP3R2* ni *RyR2*.

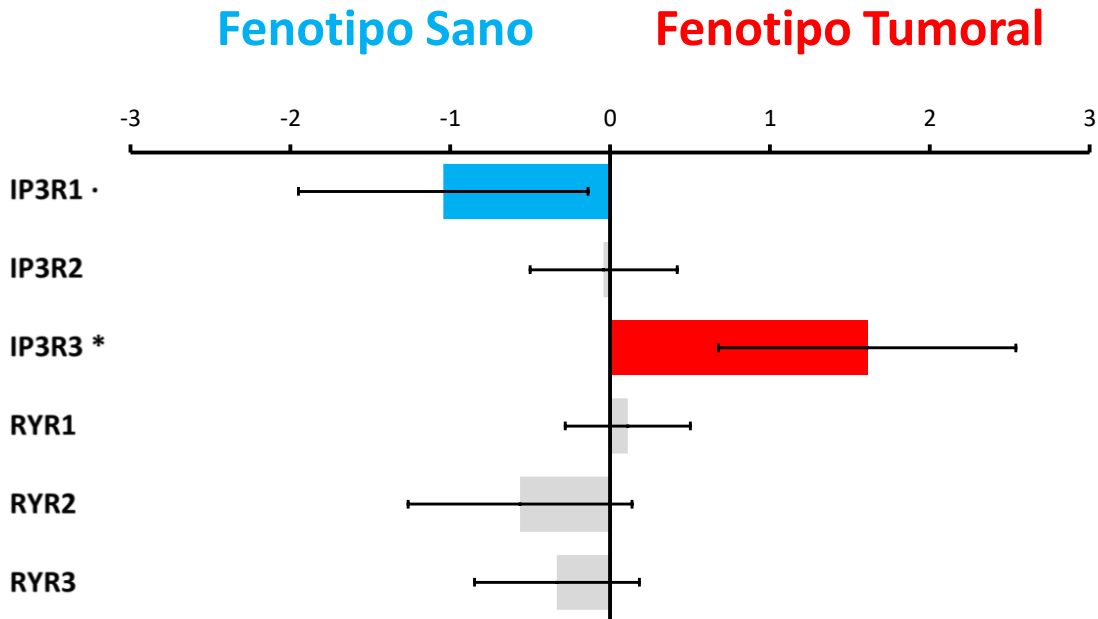


Figura 7.34: Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes ERC. En el diagrama de barras se representan los valores de expresión diferencial $\log(FC)$ estimados mediante el método limma para 16 muestras pareadas (8 de cada condición experimental). Las barras azules hacia la izquierda indican una menor expresión diferencial significativa ($FDR < 0.1$) en el tejido tumoral y mayor para las barras rojas hacia la derecha. Las barras azules hacia la izquierda indican una menor expresión significativa ($FDR < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $FDR < 0.05$ y "." $FDR < 0.1$.

Por último, en la figura 7.35 se puede observar la sobreexpresión de la mayoría de los genes de que transportan Ca^{2+} a través de la mitocondria. En concreto, los genes *MCU*, *MICU1*, *MICU2*, *MCUR1*, *VDAC1*, *VDAC3* y *NCLX*. Este patrón de expresión diferencial es diferente, en parte, del encontrado en los modelos celulares. Así, a pesar de que la expresión diferencial de *MCU*, *MICU1* y *MCUR1* coincide, para *MICU2*, *VDAC1,3* y *NCLX* es totalmente opuesto. Además, en los tejidos no se observa expresión diferencial de *MCUb* ni *VDAC2*.

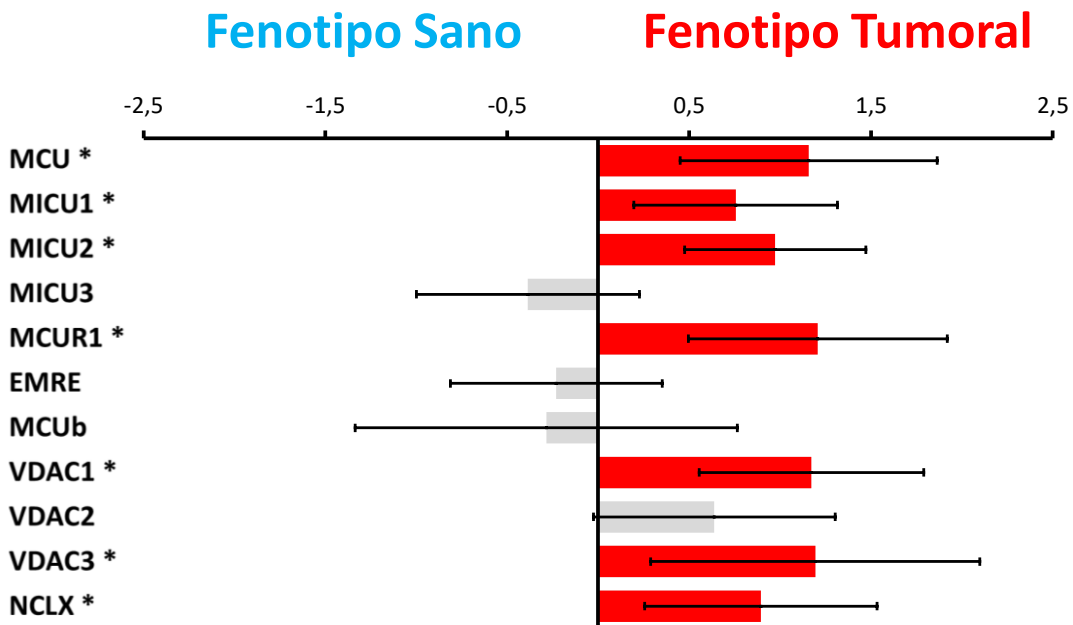


Figura 7.35: Expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes MCT. En el diagrama de barras se representan los valores de expresión diferencial $\log_2(\text{FC})$ estimados mediante el método limma para 16 muestras pareadas (8 de cada condición experimental). Las barras azules hacia la izquierda indican una menor expresión diferencial significativa ($\text{FDR} < 0.1$) en el tejido tumoral y mayor para las barras rojas hacia la derecha. Las barras azules hacia la izquierda indican una menor expresión significativa ($\text{FDR} < 0.1$) en las células HT29 y una mayor expresión para las barras rojas hacia la derecha; los símbolos a la derecha del nombre del gen indican: "*" $\text{FDR} < 0.05$ y "**" $\text{FDR} < 0.1$.

7.2.1.2. Análisis de significación biológica

A pesar de haber mostrado únicamente los resultados de expresión diferencial de los genes de interés, la realidad es que, de las 12944 sondas analizadas, el 43.64 % han mostrado expresión diferencial. Por ello, tras realizar el análisis de expresión diferencial, se procedió a buscar un significado biológico a la lista de DEG (análisis de significación biológica). En primer lugar, se realizó una búsqueda de patrones empleando dos metodologías diferentes. Por un lado, se realizó una *búsqueda de genes relacionados con el fenotipo tumoral respecto del normal* del cáncer de colon, para lo que se realizó el análisis de componentes principales y, usando las componentes principales obtenidas como variables explicativas. A continuación, se aplicó un algoritmo de

aprendizaje supervisado (árboles de clasificación) para encontrar qué componente principal explica mejor las diferencias entre el tejido de CCR y el normal (es capaz de separar las muestras en estos dos grupos). Por último, a partir de la componente principal seleccionada, se evaluó qué genes tienen más peso o influyen más sobre esta, para lo que se estimó la correlación de cada uno de ellos con dicha componente. Por otro lado, se realizó el *estudio de los perfiles de expresión*, en el que a partir del estudio de los perfiles de expresión mediante mapas de calor y un algoritmo de aprendizaje no supervisado -agrupa muestras y genes sin necesidad de información previa sobre ellas-. En segundo lugar, se realizó el análisis de enriquecimiento con el fin de encontrar conjuntos de genes asociados con las diferencias entre el tejido humano de CCR y el normal.

7.2.1.2.1. Búsqueda de patrones: componentes principales y algoritmos de clasificación

A partir de las variables originales, miles de genes, se calculan los componentes principales para reducir la dimensión y poder representar las muestras en dos dimensiones. No obstante, para encontrar qué componentes principales (PC) son las más relevantes a la hora de explicar las diferencias entre el tejido tumoral y el normal se ha empleado el algoritmo de aprendizaje automático supervisado *árboles de clasificación C5.0* (figura 7.36). Afortunadamente, la PC1 es la que mejor explica estas diferencias, lo que supone que la mayor fuente de variabilidad en los datos es precisamente el tipo de tejido ya que la PC1 es capaz, por sí sola, de explicar las diferencias entre el tejido humano de CCR y el normal. Es más, las medidas de rendimiento indican una clasificación con una precisión del 100 %, una sensibilidad del 100 % o una tasa de error del 0 %.

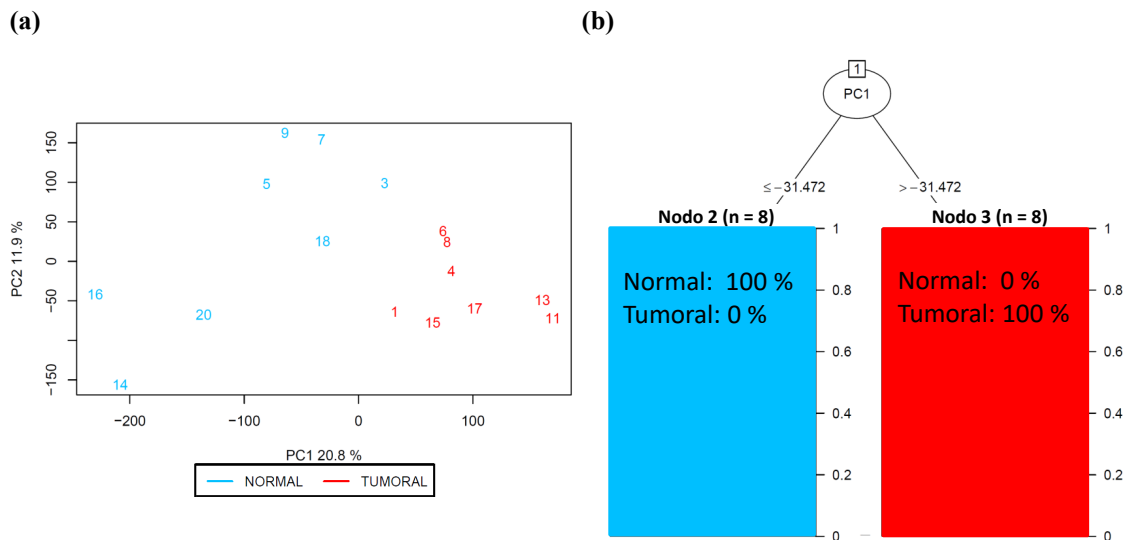


Figura 7.36: Análisis de componentes principales y algoritmo de clasificación supervisada (árboles de clasificación). a) Análisis de componentes principales: sobre las muestras analizadas se ha realizado un análisis de componentes principales para reducir su dimensión. b) Árboles de clasificación: sobre las componentes principales se aplica el algoritmo de clasificación supervisada *árboles de clasificación C5.0* para obtener qué PC es la más relevante en la clasificación de las muestras en función del tipo de tejido.

Tras encontrar la PC más relevante, PC1, y para darle más peso al hecho de que la PC1 es capaz de explicar las diferencias entre el tejido, se realiza un análisis de la varianza para un modelo lineal en el que el factor fijo sea el tipo de tejido y la variable respuesta la PC1, del cual se ha obtenido que, efectivamente, la PC1 se puede explicar por el factor **Tejido** (figura 7.37a). Por otro lado, usando las variables PC1 y PC2 como explicativas se realiza un análisis cuadrático discriminante, otro algoritmo de clasificación no supervisado, donde puede observarse cómo también se obtiene una clasificación perfecta de las muestras en función del tipo de tejido con esta PC y la siguiente que más variabilidad de los datos explica, PC2 (figura 7.37b).

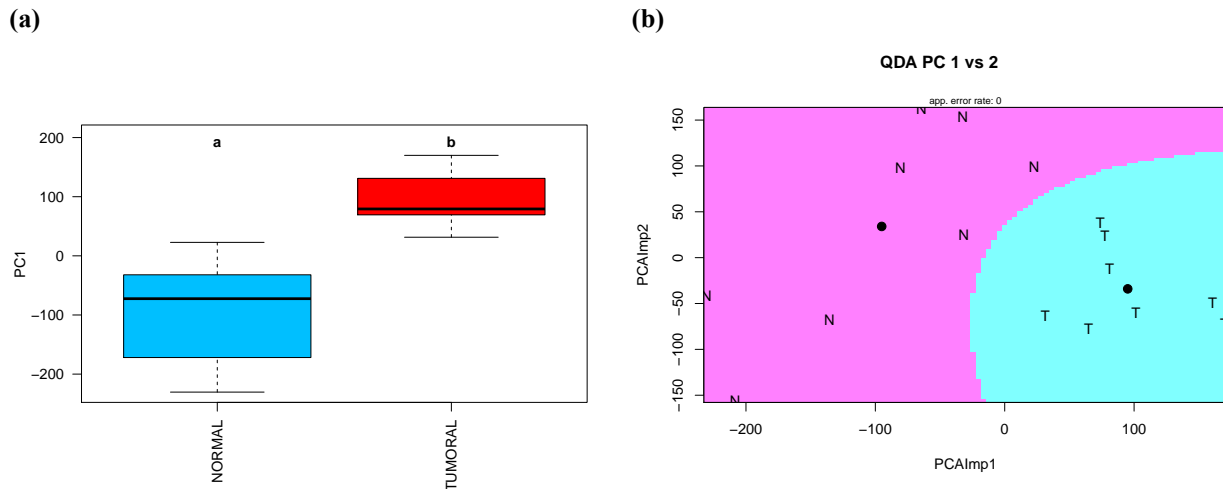


Figura 7.37: Análisis de la varianza de la variable respuesta PC1 en función del tipo de tejido y algoritmo de clasificación supervisada (análisis cuadrático discriminante). a) Análisis de la varianza de la PC1: se hace un ANOVA del modelo lineal cuya variable respuesta es PC1 y el factor fijo es el tipo de tejido. b) Análisis cuadrático discriminante: usando las PC1 y PC2 como variables clasificadoras para explicar las diferencias entre las muestras de diferentes tejidos.

Tras estos resultados, parece lógico estudiar qué variables originales, genes, son las que más peso tienen o más correlacionadas están con esta PC1, pues serán aquellos que mejor expliquen las diferencias entre el tejido tumoral y el sano. Para ello, se estima la correlación para cada uno de los genes con la PC1 y se contrasta la hipótesis nula $H_0 : \rho_{\text{gen},PC1} = 0$ frente a la $H_1 : \rho_{\text{gen},PC1} \neq 0$, tal y como se explica en el apéndice “A.5. Coeficiente de correlación de Pearson”. A continuación, se corrigen los p valores controlando el FDR mediante la aproximación de Benjamini-Hochberg (Benjamini and Hochberg (1995)). Este enfoque podría pensarse como un procedimiento alternativo de análisis de expresión diferencial. Así, de un modo similar a como se han mostrado los resultados de expresión diferencial, se procede a mostrar los valores de correlación entre la PC1 y la expresión de los genes relacionados con la homeostasis del Ca^{2+} intracelular (figuras 7.38 a 7.40).

Como puede apreciarse, en el contexto de los genes de la homeostasis del Ca^{2+} intracelular, que aparte de observarse una correlación significativa con la PC1 en aquellos genes que mostraron tener expresión diferencial, también se han encontrado otros genes que no mostraron expresión diferencial pero que, sin embargo, están correlacionados con PC1 y, por ende, pueden explicar las diferencias entre los tipos de tejido. Cabe destacar que al igual que para el análisis de expresión diferencial con las muestras de tejido humano pareadas, para evaluar la correlación entre la expresión de los genes y la PC1 también se ha fijado un $\alpha_{\text{FDR}} = 0.1$, por lo que en los gráficos se mostrarán coloreados aquellos con un FDR inferior.

En concreto, aquellos genes cuya menor expresión en el tejido de CCR son capaces de explicar las diferencias entre las muestras normales y tumorales, aparte de los identificados como diferencialmente infraexpresados mediante métodos clásicos de análisis de expresión diferencial (limma), son *STIMI*, *TRPC7* y *TRPM1*. De igual modo, aparte de los diferencialmente sobreexpresados, los genes capaces de explicar las diferencias entre las muestras normales y tumorales que están sobreexpresados son *SARAF*, *SEPTIN2,7* (FDR < 0.1), *VDAC2*, *TRPV1* (FDR < 0.1), *TRPV2* y *TRPP1*. Comparando estos resultados con los obtenidos a partir de los modelos celulares, se tiene que la correlación negativa con la PCA1 de *TRPC7* y *VDAC2*, y la positiva de *SEPTIN2,7*, *TRPP1* está de acuerdo con . No obstante, mientras que no se observa expresión diferencial de *TRPM1* y *TRPV2* en los modelos celulares, la correlación con la PC1 de *STIMI*, *SARAF*, *TRPV1* encontrada en los tejidos estaría en contra de la expresión diferencial encontrada para estos genes en dichos modelos celulares.

Hay que tener en cuenta que este análisis es complementario al de expresión diferencial y que, por ello, no ha de extrañar al lector que haya diferencias entre los genes identificados en cada uno de los casos. Es más, esta complementariedad entre diferentes análisis puede, en muchas ocasiones, ayudar a encontrar respuestas que puedan estar en la sombra al solo emplear un punto de vista. Además, no hay que olvidar que cualquier test o cualquier análisis, cada uno basado en unas suposiciones y fundamentos teóricos, puede cometer tanto falsos positivos como falsos negativos, en mayor o menor medida y, por ello, la importancia de evaluar los problemas desde diferentes puntos de vista. Por lo tanto, a continuación, se muestran los genes diferencialmente expresados y/o que contribuyen a explicar las diferencias entre las muestras de tejido de cáncer de colon y normales (figuras 7.38-7.43).

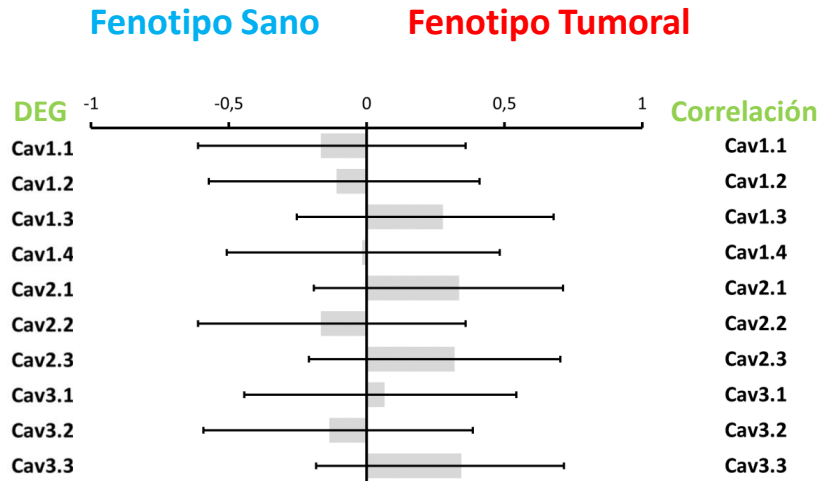


Figura 7.38: Correlación de los genes de la familia VOCC con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. En el diagrama de barras se representan los valores de la correlación de la expresión de cada gen con PC1.

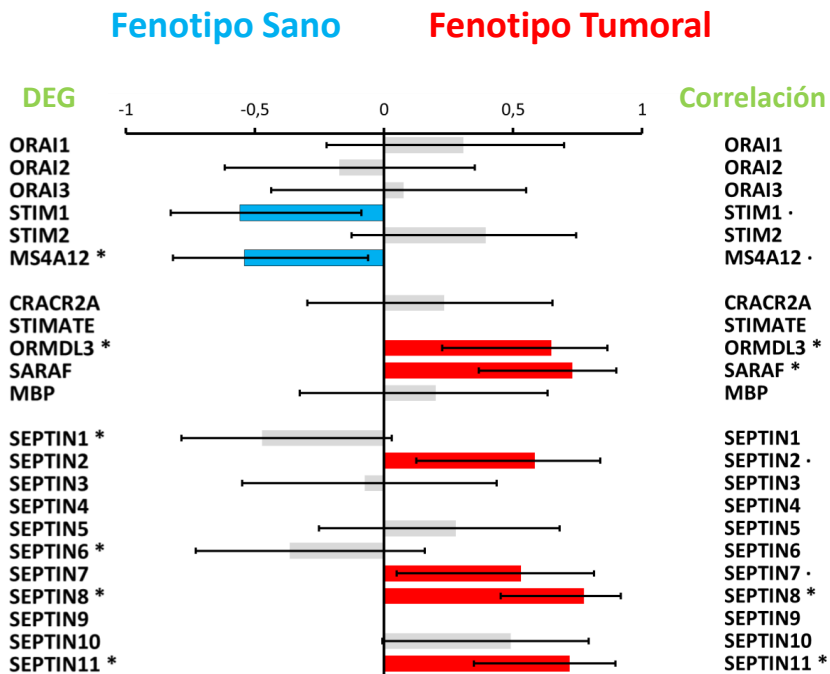


Figura 7.39: Correlación de los genes de la familia SOCE con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. En el diagrama de barras se representan los valores de la correlación de la expresión de cada gen con PC1. Las barras azules hacia la izquierda indican una correlación negativa significativa con PC1, mientras que las rojas hacia la derecha indican una correlación positiva significativa. Los símbolos a la derecha de los nombres de los genes indican un FDR inferior a 0.05 (***) o a 0.1 (**), y, además, la columna de genes de la izquierda indican el resultado del análisis de expresión diferencial del apartado "7.2.1.1" mientras que la de la derecha se refiere al contraste de hipótesis sobre la correlación con PC1.

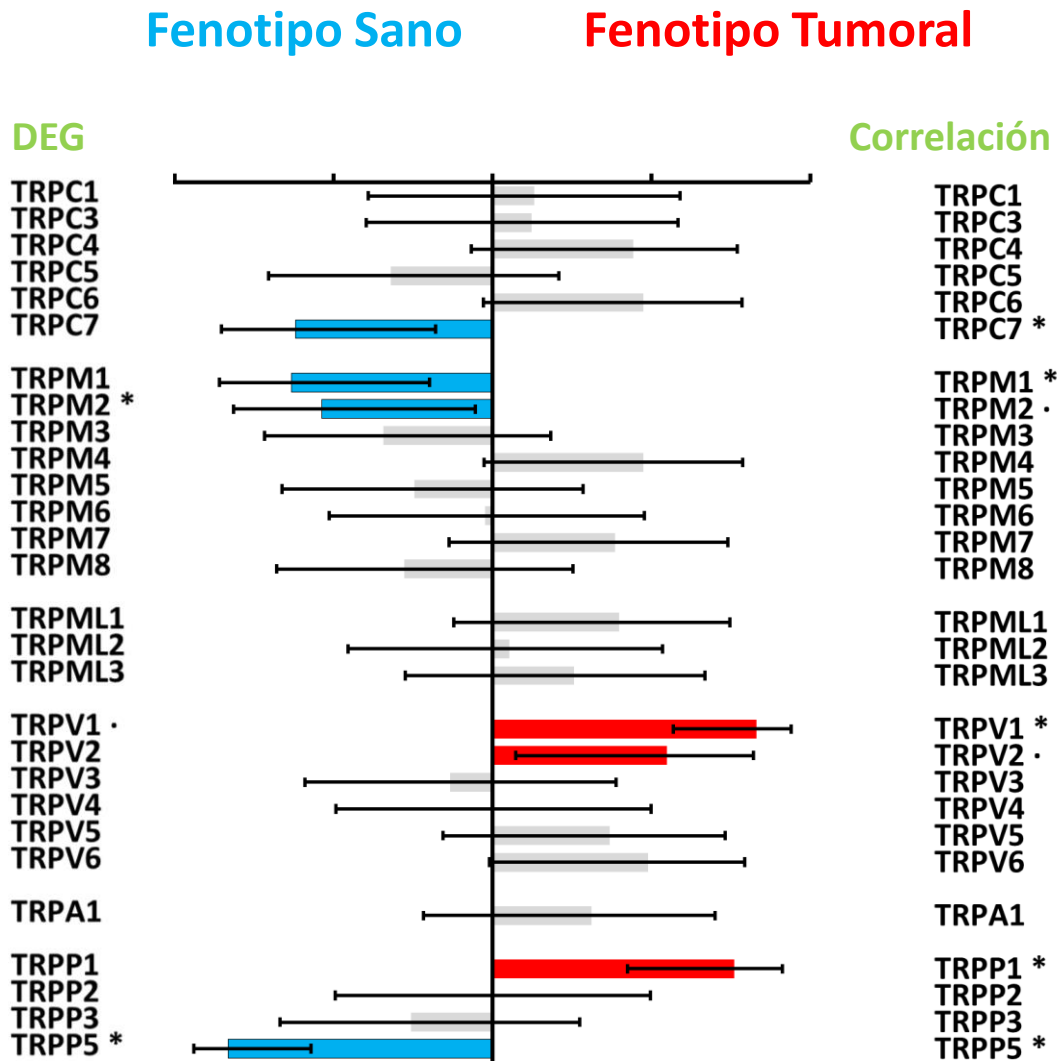


Figura 7.40: Correlación de los genes de la familia TRP con la PC1 capaz de explicar las diferencias entre los tipos de tejido, PC1. En el diagrama de barras se representan los valores de la correlación de la expresión de cada gen con PC1. Las barras azules hacia la izquierda indican una correlación negativa significativa con PC1, mientras que las rojas hacia la derecha indican una correlación positiva significativa. Los símbolos a la derecha de los nombres de los genes indican un FDR inferior a 0.05 (***) o a 0.1 (**), y, además, la columna de genes de la izquierda indican el resultado del análisis de expresión diferencial del apartado "7.2.1.1" mientras que la de la derecha se refiere al contraste de hipótesis sobre la correlación con PC1.

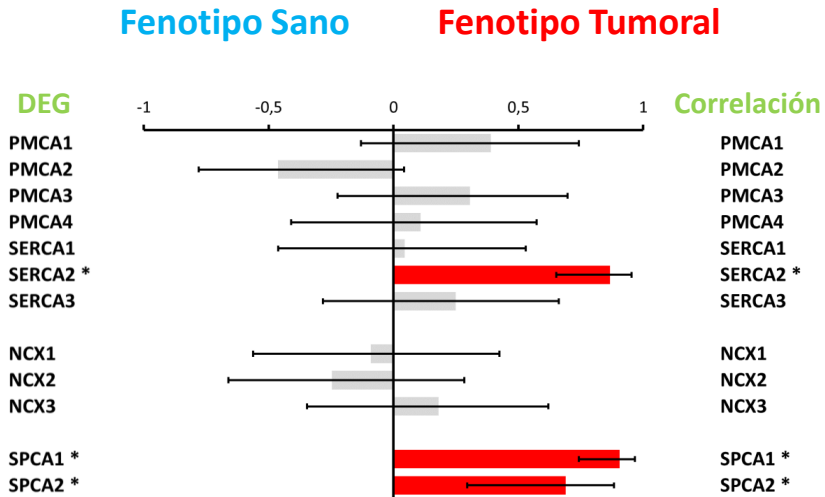


Figura 7.41: Correlación de los genes de la familia CPT (bombas e intercambiadores de calcio) con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. En el diagrama de barras se representan los valores de la correlación de la expresión de cada gen con PC1. Las barras azules hacia la izquierda indican una correlación negativa significativa con PC1, mientras que las rojas hacia la derecha indican una correlación positiva significativa. Los símbolos a la derecha de los nombres de los genes indican un FDR inferior a 0.05 (***) o a 0.1 (**), y, además, la columna de genes de la izquierda indican el resultado del análisis de expresión diferencial del apartado "7.2.1.1" mientras que la de la derecha se refiere al contraste de hipótesis sobre la correlación con PC1.

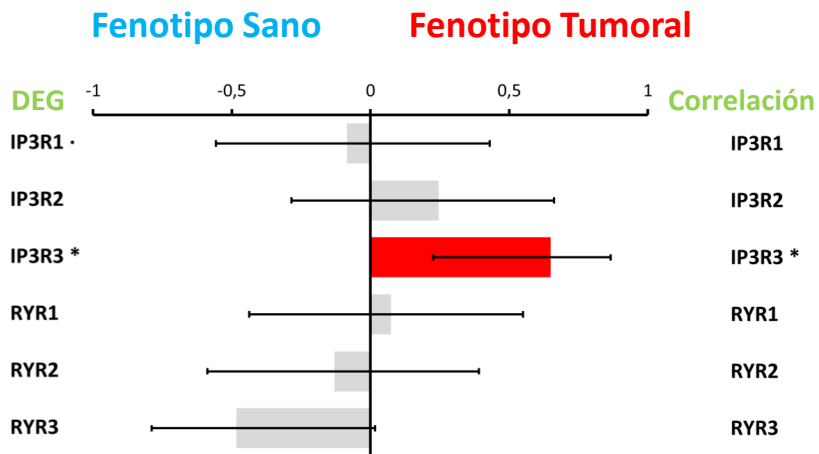


Figura 7.42: Correlación de los genes de la familia ERC con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. En el diagrama de barras se representan los valores de la correlación de la expresión de cada gen con PC1. Las barras azules hacia la izquierda indican una correlación negativa significativa con PC1, mientras que las rojas hacia la derecha indican una correlación positiva significativa. Los símbolos a la derecha de los nombres de los genes indican un FDR inferior a 0.05 (***) o a 0.1 (**), y, además, la columna de genes de la izquierda indican el resultado del análisis de expresión diferencial del apartado "7.2.1.1" mientras que la de la derecha se refiere al contraste de hipótesis sobre la correlación con PC1.

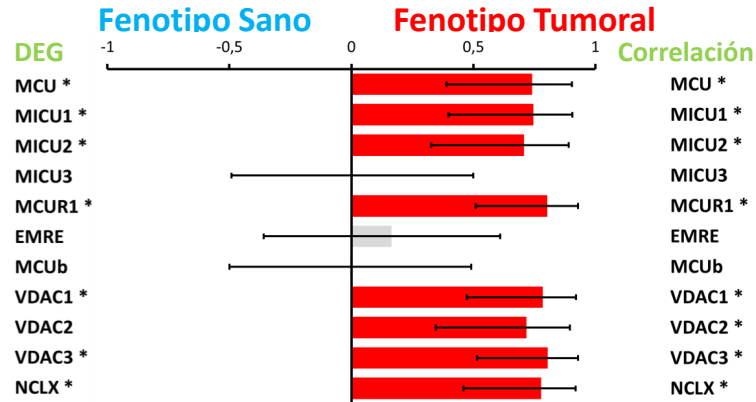


Figura 7.43: Correlación de los genes de la familia MCT con la PC capaz de explicar las diferencias entre los tipos de tejido, PC1. En el diagrama de barras se representan los valores de la correlación de la expresión de cada gen con PC1. Las barras azules hacia la izquierda indican una correlación negativa significativa con PC1, mientras que las rojas hacia la derecha indican una correlación positiva significativa. Los asteriscos a la derecha de los nombres de los genes indican un FDR inferior a 0.05. La columna de genes de la izquierda indican el resultado del análisis de expresión diferencial del apartado 7.2.1.1 a la de la derecha del contraste de hipótesis sobre la correlación con PC1.

En la figura 7.11 se muestra un resumen de los resultados anteriores, tanto del análisis de expresión diferencial como de correlación (7.44).

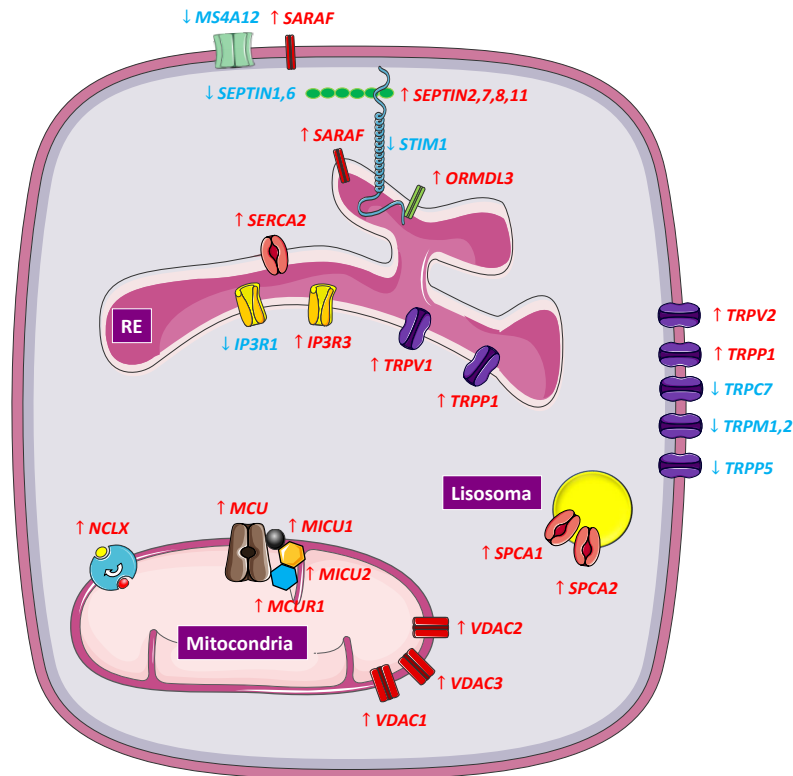


Figura 7.44: Resumen del análisis de expresión diferencial de los genes relacionados con el transporte de Ca^{2+} intracelular y de su influencia en la clasificación entre tejido tumoral y sano. En la célula se muestran aquellos genes que están diferencialmente expresados y/o explican las diferencias entre tejido tumoral y normal de colon. En rojo se representan los que están sobreexpresados y en azul aquellos infraexpresados en el tejido tumoral.

7.2.1.2.2. Búsqueda de patrones: Perfiles de expresión

Con la ayuda de un mapa de calor se han representado las expresiones de las sondas diferencialmente expresadas en el análisis anterior para cada una de las muestras analizadas y, además, mediante el algoritmo de clasificación no supervisado *análisis clúster jerárquico con el método de Ward*, tanto las muestras como los genes se han agrupado según su mayor o menor similitud pero sin conocimiento previo sobre si son o no tejido tumoral (figura 7.45). Por un lado, se observan dos conjuntos de muestras que han sido agrupadas en función del tipo de tejido: a la izquierda se agrupan las muestras tumorales y a la derecha las normales, lo cual es indicativo de la buena calidad de las muestras puesto que la mayor fuente de variabilidad es el tipo de tejido (que no fuera así no significaría mala calidad, sino que hay fuentes de variabilidad que no se han considerado y que habría que buscar). Por otro lado, los genes (sondas en este caso) se agrupan en dos conjuntos: un conjunto más reducido que muestra una menor expresión en las muestras de tejido tumoral (439 sondas) y otro muy superior que muestra una clara sobreexpresión en dicho tejido respecto del normal (3996 sondas); es decir, que se observa una clara sobreexpresión de genes en el tejido humano de CCR (figura 7.45).

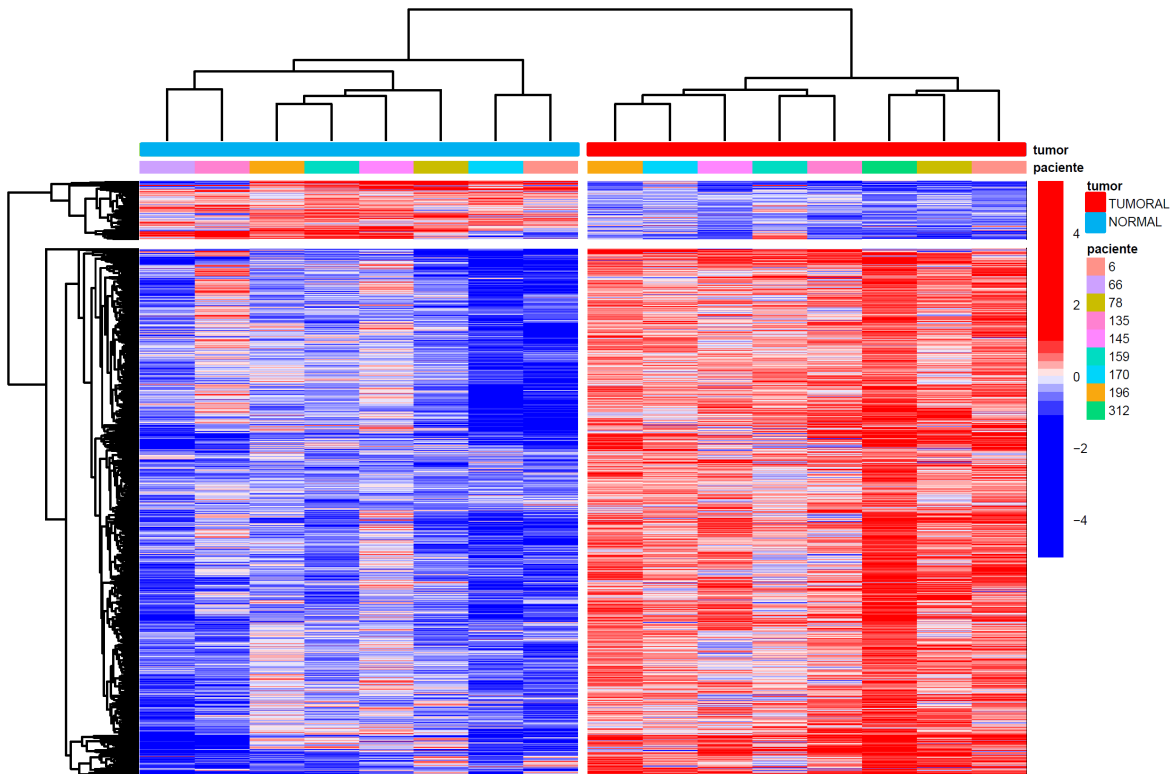


Figura 7.45: Perfiles de expresión. En el mapa de calor se representan los valores centrados de la expresión en escala logarítmica. En el eje X se agrupan las diferentes muestras mediante el análisis clúster jerárquico basado en el método de Ward, y en el eje Y lo hacen los DEG. El rojo indica sobreexpresión de un gen (fila) en una muestra dada (columna) respecto de la media de los valores de expresión (log) de ese mismo gen a lo largo de todas las muestras. La leyenda correspondiente a los pacientes indica qué muestras (tumoral y normal) pertenecen a un mismo paciente.

7.2.1.2.3. Análisis de enriquecimiento

Para estudiar la lista de DEG de forma global, en lugar de sonda a sonda, se ha recurrido a información previamente generada por la comunidad científica sobre grupos de genes con significado biológico. En concreto, se han consultado las bases de datos KEGG PATHWAY, GO y MSigDB que, a su vez, contienen diferentes colecciones de conjuntos de genes con diferente significado biológico. En concreto, se han consultado hasta 9 bases de datos diferentes con cerca de 30000 conjuntos de genes (tabla 7.7). Por lo tanto, se cruzan los datos de la lista de genes generada por el análisis de expresión diferencial y se contrastan la *hipótesis nula autocontenida* -usando el método CAMERA- y la *hipótesis nula competitiva* -mediante el método ORA- para inferir qué conjuntos de genes están asociados con el tejido de CCR respecto del normal según su perfil de expresión diferencial. No obstante, es evidente que esta cantidad de conjuntos es un número demasiado elevado como para profundizar en el significado de cada uno de ellos, puesto que no es el propósito de esta tesis. Para extraer información interesante, el análisis de estos conjuntos se ha centrado en obtener un resumen numérico de cuántos en total rechazan alguna de las dos hipótesis nulas ($FDR < 0.05$) y, de estos, cuántos pertenecen a las bases de datos relacionadas con el cáncer (MCC y MOG) y cuántos contienen en su nombre alguna de las siguientes palabras clave: *calcium*, *mitochon*, *reticulum*, *cancer*, *wnt* y *polyamines* (tabla 7.7).

Tabla 7.7: Análisis de conjuntos de genes relacionados con el tejido humano de CCR.

| Base de Datos | Conjuntos | | Conjuntos Significativos (FDR <0.05) | | | | | | | |
|---------------|-----------|---------------------|--------------------------------------|---------|---------------------|---------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | Métodos | Totales | Calcio ¹ | Micochondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA | 84 | 1 | 0 | 1 | 0 | 0 | 0 |
| | | | ORA | 11 | 0 | 0 | 1 | 0 | 0 | 0 |
| GO BP | 12342 | 146 | CAMERA | 198 | 8 | 3 | 2 | 0 | 0 | 0 |
| | | | ORA | 62 | 0 | 2 | 0 | 0 | 0 | 0 |
| GO MF | 4245 | 66 | CAMERA | 66 | 2 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| MH | 50 | 0 | CAMERA | 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 9 | 0 | 0 | 0 | 0 | 0 | 0 |
| MCDB | 6290 | 12 | CAMERA | 1193 | 3 | 6 | 0 | 130 | 5 | 1 |
| | | | ORA | 767 | 0 | 6 | 0 | 89 | 4 | 1 |
| MCC | 858 | 0 | CAMERA | 424 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 227 | 0 | 0 | 0 | 0 | 0 | 0 |
| MOG | 189 | 0 | CAMERA | 46 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| MIM | 5219 | 0 | CAMERA | 842 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 1124 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 29537 | 225 | CAMERA | 2870 | 14 | 9 | 3 | 130 | 5 | 1 |
| | | | ORA | 2232 | 0 | 8 | 1 | 89 | 4 | 1 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Onncogenic; MIM: Msigdb Immunology.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

Como puede observarse, de los 29537 conjuntos de genes analizados, 2870/2232 rechazan las hipótesis autocontenida/competitiva. De estos, 424/227 pertenecen a la base de datos MCC, 46/15 a la base MOG, 14/0 contienen la palabra *calcium*, 9/8 la palabra *Mitochondria*, 3/1 *reticulum*, 130/89 *cancer*, 5/4 *wnt*, y 1/1 *polyamine* (ver apéndice “B.2.0.4.3. Análisis de enriquecimiento”). Algunos de estos conjuntos están relacionados con:

- El transporte de Ca^{2+} : *hsa04020 Calcium signaling pathway*, *GO:0051924 regulation of calcium ion transport*, *GO:0010880 regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum*, *GO:0019722 calcium-mediated signaling*,...
- El CCR: *M16740 GRADE COLON AND RECTAL CANCER UP*, *M17423 LAIHO COLORECTAL CANCER SERRATED DN*, *M15780 GRADE COLON AND RECTAL CANCER DN*,...
- El procesamiento de proteínas en el RE: *Protein processing in endoplasmic reticulum*.
- La maquinaria de traducción mitocondrial: *GO:0070126 mitochondrial translational termination*, *GO:0070125 mitochondrial translational elongation*...
- La ruta de señalización WNT :*GO:0060071 Wnt signaling pathway*, *planar cell polarity pathway*, *M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT*, *M1758 SANSOM WNT PATHWAY REQUIRE MYC*, *M7847 REACTOME SIGNALING BY WNT*, *M39731 WP LNCRNA INVOLVEMENT IN CANONICAL WNT SIGNALING AND COLORECTAL CANCER*...
- ...

No obstante, como se refleja en la tabla 7.7, el número de conjuntos de interés es mucho mayor. Por ello, en el apéndice “B.2.0.4.3. Análisis de enriquecimiento” y en las tablas B.14-B.21 pueden consultarse, dividido en función de la base de datos a la que pertenecen, aquellos conjuntos con un $\text{FDR} < 0.1$ cuyos nombres contienen alguna de las palabras clave.

7.2.2. Análisis de expresión de proteínas implicadas en SOCE

De las mismas muestras de tejido humano de las que se extrajo ARN para estudiar su transcriptoma (apartado “7.2.1 Análisis transcriptómico de muestras pareadas de tejido humano de cáncer de colon”), en paralelo se extrajo la proteína, tal y como se explica en el paso 12 del apartado “6.6 Extracción de proteínas” . Como se explicó en el apartado “6.7.2 Procedimiento y diseño experimental” , el experimento consta de dos factores fijos que dan lugar a cuatro condiciones experimentales diferentes: *Hombre.Normal*, *Hombre.Tumoral*, *Mujer.Normal* y *Mujer.Tumoral*. Además, para cada una de las condiciones experimentales se ha registrado la expresión de cinco proteínas relacionadas con SOCE: ORAI1, ORAI2, TRPC1, STIM1 y STIM2. Como resultado, se tienen diferentes registros correspondientes al revelado de cada una de las membranas en cada una de las cuales se mide una sola proteína de interés (en las figuras 7.46 y 7.47 se muestra una membrana representativa de cada una de las cinco proteínas analizadas).

A partir de estos registros, con el fin de evaluar el efecto de cada uno de los factores fijos de interés (*Tejido* y *Sexo*) se han modelizado, de forma independiente, los niveles de expresión de cada una de las cinco proteínas de interés. Para ello, se han empleado modelos lineales mixtos, puesto que permiten tanto considerar las variables explicativas de la parte fija como la parte aleatoria debida a los factores aleatorios *Paciente* y *Gel* (para consultar el diseño experimental ver apartado “6.7.2”). En el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos” se presenta un ejemplo simplificado de cómo trabajar con lineales mixtos en el cual se explica brevemente cómo seleccionar el mejor modelo, evaluar si este es o no adecuado y, en caso de que no lo sea -no cumplan la normalidad, la homocedasticidad...-, usar técnicas más complejas de remuestreo paramétrico como *bootstrapping*.

Tras ajustar los diferentes modelos lineales, transformar las variables (en caso de ser necesario) y realizar la selección del modelo más adecuado para cada proteína de interés, se obtuvo un modelo diferente para cada ellas, si bien es cierto que en los cinco casos hubo que transformar la variable respuesta mediante la familia de transformaciones potencia de Box-Cox, indicado con $T(\cdot)$ (Hawkins and Weisberg (2017)). Además, de los cinco, dos resultaron ser modelos lineales mixtos y los otros tres modelos lineales de efectos fijos:

- *STIM1*: se obtuvo que el modelo más adecuado para modelizar la expresión de STIM1 es (ecuación 7.2):

$$\log \left(\frac{DO_{STIM1}}{DO_{\beta\text{-actina}}}_{ijr} \right) = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{sexo}_j + \epsilon_{ijr} \quad (7.2)$$

donde $\log (DO_{STIM1}/DO_{\beta\text{-actina}})_{ijr}$ es el valor de la expresión de STIM1 medido en la réplica r de la muestra cuyo tejido es i (Tumoral o Normal) y sexo j (Hombre o Mujer). En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel Hombre.Normal); β_1 es el cambio de expresión en el tejido tumoral respecto

del de referencia (Hombre.Normal), β_2 el cambio en la expresión de la proteína de interés en las muestras de mujer respecto de las de referencia (Hombre.Normal), y β_3 el efecto de la interacción entre los factores *Tejido* y *Sexo* que indica si el cambio entre el tejido tumoral y normal es diferente en hombres que en mujeres (en caso de ser significativo).

- *STIM2*: el modelo más adecuado para modelizar la expresión de STIM2 es (ecuación 7.3):

$$T \left(\frac{DO_{STIM2}}{DO_{\beta\text{-actina}}_{ijdr}} \right) = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{sexo}_j + b_{0,d} + \epsilon_{ijdr} \quad (7.3)$$

donde $T \left(\frac{DO_{STIM2}}{DO_{\beta\text{-actina}}_{ijdr}} \right)$ es el valor de la variable respuesta transformada según la familia de Box-Cox propuesto por Hawkins and Weisberg (2017), medido en la réplica r de la muestra cuyo tejido es i (Tumoral o Normal) y el sexo j (Hombre o Mujer) medidas en el paciente d . En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel Hombre.Normal); β_1 es el cambio de expresión en el tejido tumoral respecto del de referencia (Hombre.Normal), β_2 el cambio en la expresión de la proteína de interés en las muestras de mujer respecto de las de referencia (Hombre.Normal); β_3 el efecto de la interacción entre los factores *Tejido* y *Sexo* que indica si el cambio entre el tejido tumoral y normal es diferente en hombres que en mujeres (en caso de ser significativo), y $b_{0,d}$ es el intercepto aleatorio debido al día que indica la desviación respecto de β_0 que hay en el paciente d .

- *ORAI1*: el modelo más adecuado para modelizar la expresión de ORAI1 es (ecuación 7.4):

$$\log \left(\frac{DO_{ORAI1}}{DO_{\beta\text{-actina}}_{ijr}} \right) = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{sexo}_j + \epsilon_{ijr} \quad (7.4)$$

donde $\log \left(\frac{DO_{ORAI1}}{DO_{\beta\text{-actina}}_{ijr}} \right)$ es el valor de la expresión de ORAI1 medido en la réplica r de la muestra cuyo tejido es i (Tumoral o Normal) y sexo j (Hombre o Mujer).

- *ORAI2*: el modelo más adecuado para modelizar la expresión de ORAI2 es un modelo lineal mixto (ecuación 7.5):

$$\log \left(\frac{DO_{ORAI2}}{DO_{\beta\text{-actina}}_{ijmpr}} \right) = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{Sexo}_j \gamma_{ij} + b_{0,gelm} + b_{0,pacientep} + \epsilon_{ijmpr} \quad (7.5)$$

donde $\log \left(\frac{DO_{ORAI2}}{DO_{\beta\text{-actina}}_{ijmpr}} \right)$ es el valor de la expresión de ORAI2 medido en la réplica r de la muestra cuyo tejido es i (Tumoral o Normal), sexo j (Hombre o Mujer)

medidas en el gel m . En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel Hombre.Normal); β_1 es el cambio de expresión en el tejido tumoral respecto del de referencia (Hombre.Normal), β_2 el cambio en la expresión de la proteína de interés en las muestras de mujer respecto de las de referencia (Hombre.Normal); β_3 el efecto de la interacción entre los factores *Tejido* y *Sexo* que indica si el cambio entre el tejido tumoral y normal es diferente en hombres que en mujeres (en caso de ser significativo); $b_{0,gelm}$ es el intercepto aleatorio debido al gel que indica la desviación respecto de β_0 que hay en gel m , y $b_{0,paciente}$ es el intercepto aleatorio debido a paciente p que indica la desviación respecto de β_0 que muestra el paciente p . En la tabla D.7 se muestra la selección de la parte aleatoria y en la tabla D.11 la selección de la parte fija.

- *TRPC1*: el modelo más adecuado para modelizar la expresión de TRPC1 es un modelo lineal de efectos fijos, al igual que en el caso de ORAI1 (ecuación 7.6):

$$T \left(\frac{DO_{TRPC1}}{DO_{\beta\text{-actina}}}_{ijr} \right) = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{sexo}_j + \epsilon_{ijr} \quad (7.6)$$

Para cada uno de los modelos anteriores los gráficos, diagnósticos, estimaciones... se muestran en el apéndice “D Análisis de datos de Western Blot” (tablas D.1, D.12, D.29 y D.40, y figuras D.1-[{fig:ModeloSTIM2STIM2SinOut}](#)). Así, una vez se ha ajustado el modelo más adecuado a cada una de las proteínas, en aquellos casos en los que el término de interacción sea significativo (o deba incluirse según AIC) se compararon dos a dos todas las condiciones experimentales y se realizó el contraste de hipótesis oportuno para identificar cuales de ellas son diferentes entre sí (p valor ajustado con Tukey inferior a 0.05 -0.07 para ORAI2-) y se agruparon las diferentes muestras en función de dichas diferencias aplicando el algoritmo Bron-Kerbosch de la teoría de grafos (tablas D.6 a D.46). Cabe destacar que si el término de interacción no es significativo y ambos factores sí que lo son (como en el caso de ORAI2) entonces significa que las cuatro condiciones experimentales son diferentes entre sí.

En las figuras 7.46 y 7.47 se muestran los resultados mediante diagramas de barras de expresión de las cinco proteínas evaluadas. Además, los resultados numéricos correspondientes a los cinco modelos y a sus diferentes contrastes de hipótesis se proporcionan en el apéndice “D Análisis de datos de Western Blot”. Cabe destacar que no ha sido necesario recurrir a métodos de *bootstrapping*, puesto que las suposiciones del modelo se cumplen en todos los casos. Como resultado, en la figura 7.46a y 7.46c se observa que ni la expresión de ORAI1 ni de TRPC1 varía ni entre sexos ni entre el tejido tumoral y normal, respectivamente. En cuanto a ORAI2, la expresión disminuye en las muestras tumorales respecto de las normales y, además, la expresión es superior en mujeres que en hombres, por lo que las cuatro condiciones experimentales son diferentes entre sí (figura 7.46b).

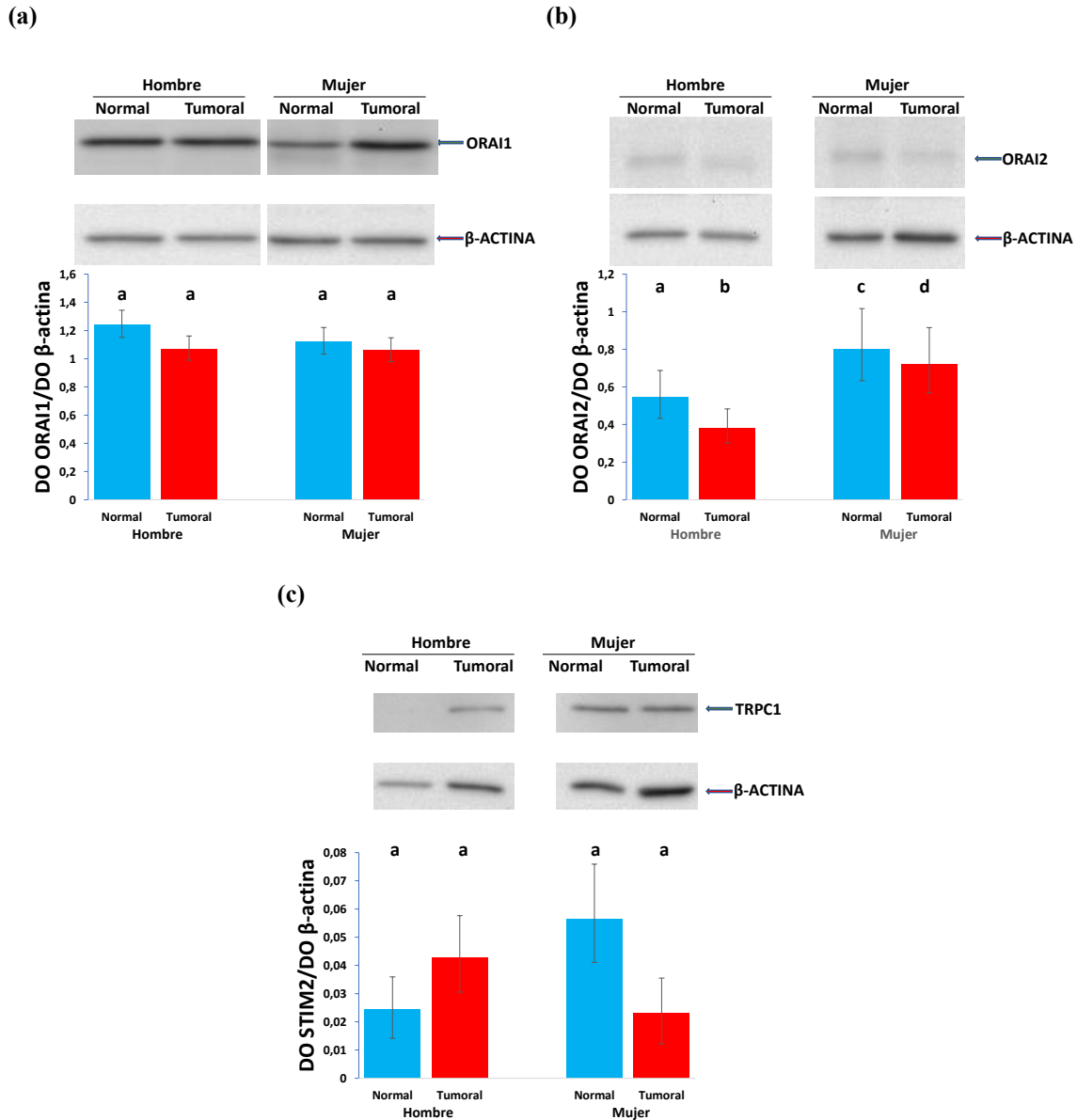


Figura 7.46: Análisis de expresión de las proteínas ORAI1, ORAI2 y TRPC1 en muestras pareadas de tejido humano de cáncer de colon y normal mediante *Western Blot*. a) ORAI1, b) ORAI2 y c) TRPC1. En cada subfigura, en la parte superior se muestra una imagen representativa correspondiente a la proteína de interés y su control endógeno o de carga β -actina. En la parte inferior, se muestran los diagramas de barras en los que se representan las estimaciones de la expresión de la proteína de interés para cada una de las condiciones experimentales evaluadas y las barras de error \pm SEM asociadas a dichas estimaciones, donde la expresión o variable respuesta se expresa como el ratio densidad óptica (DO) de la proteína de interés entre la DO de la β -actina. Las agrupaciones entre las condiciones experimentales, en aquellos casos en los que existe el término de interacción en el modelo, se han realizado a partir de las diferencias dos a dos (significativas para un p valor ajustado por Tukey <0.05-0.07) y el algoritmo de Bron-Kerbosch, de tal manera que dos condiciones experimentales con la misma letra son iguales, dos con diferente letra son diferentes y en caso de que alguna condición tenga más de una letra, por ejemplo *ab*, se considera similar tanto a la condición *a* como a la *b* aunque las condiciones *a* y *b* sean diferentes entre sí. El número total de muestras para cada condición experimental, en función de la proteína, es (se muestran en el orden Hombre.Normal/ Hombre.Tumoral/ Mujer.Normal/ Mujer.Tumoral): **(a) ORAI1, 13/12/11/12, (b) ORAI2, 11/10/10/10 y (c) TRPC1, 13/13/11/11.**

En cuanto a las isoformas de STIM, la expresión de STIM1 muestra una clara disminución en el tejido tumoral respecto del normal sólo en hombres, mientras que en mujeres no se aprecia ninguna variación en la expresión de STIM1 (figura 7.47a). En cuanto a STIM2, se observa el patrón totalmente opuesto a STIM1; es decir, mientras que en hombres no se aprecian diferencias entre el tejido normal y tumoral, en mujeres hay un aumento en la expresión de la proteínas STIM2 en el tejido tumoral respecto del normal (figura 7.47b).

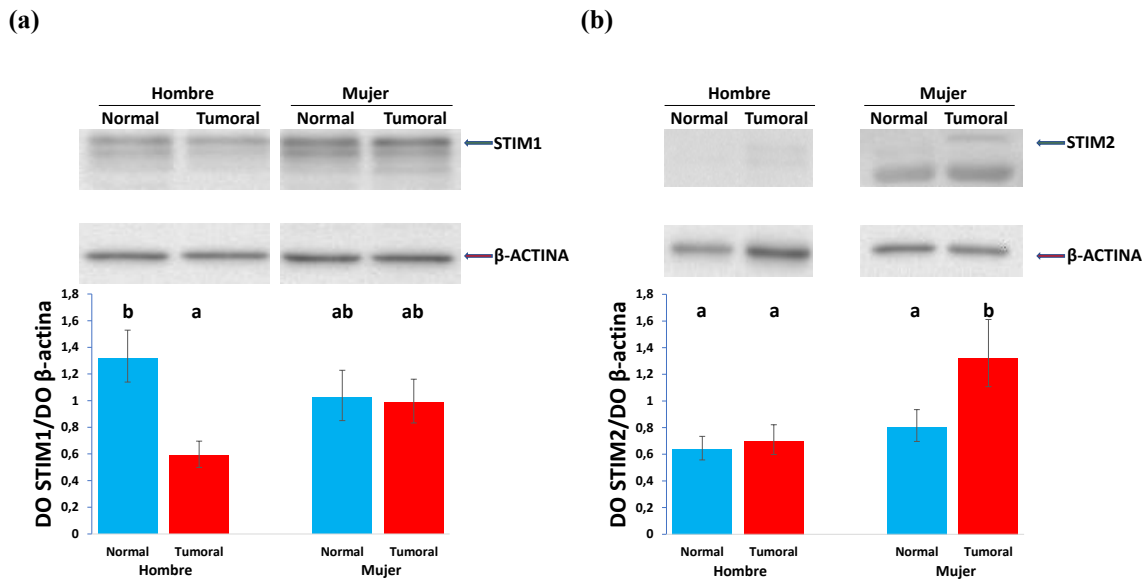


Figura 7.47: Análisis de expresión de las proteínas STIM1 y STIM2 en muestras pareadas de tejido humano de cáncer de colon y normal mediante *Western Blot*. a) STIM1 y b) STIM2. En cada subfigura, en la parte superior se muestra una imagen representativa correspondiente a la proteína de interés y su control endógeno o de carga β -actina. En la parte inferior, se muestran los diagramas de barras en los que se representan las estimaciones de la expresión de la proteína de interés para cada una de las condiciones experimentales evaluadas y las barras de error \pm SEM asociadas a dichas estimaciones, donde la expresión o variable respuesta se expresa como el ratio densidad óptica (DO) de la proteína de interés entre la DO de la β -actina. Las agrupaciones entre las condiciones experimentales, en aquellos casos en los que existe el término de interacción en el modelo, se han realizado a partir de las diferencias dos a dos (significativas para un p valor ajustado por Tukey $<0.05-0.07$) y el algoritmo de Bron-Kerbosch, de tal manera que dos condiciones experimentales con la misma letra son iguales, dos con diferente letra son diferentes y en caso de que alguna condición tenga más de una letra, por ejemplo *ab*, se considera similar tanto a la condición *a* como a la *b* aunque las condiciones *a* y *b* sean diferentes entre sí. El número total de muestras para cada condición experimental, en función de la proteína, es (se muestran en el orden Hombre.Normal/ Hombre.Tumoral/ Mujer.Normal/ Mujer.Tumoral): **(a)** STIM1, 14/11/9/11, y **(b)** STIM2, 10/9/12/8.

7.2.3. Reanálisis transcriptómico de muestras de tejido humano de cáncer de colon extraídas de repositorios públicos y metaanálisis

En el apartado “7.2.1” se ha realizado y analizado un experimento transcriptómico de muestras pareadas de tejido humano de CCR y normal. No obstante, es obvio que estos resultados, aunque interesantes, no son más que la evidencia encontrada en un solo grupo de investigación a partir de un total de 20 muestras procedentes de 10 pacientes de ambos sexos y en diferentes estadios de la enfermedad, por lo que sería interesante comparar estos resultados con la evidencia existente, pues en un campo con tantas variables no controlables como la investigación biomédica es posible que existan diferencias notables entre la evidencia generada por diferentes grupos de investigación.

Es más, para dar respuesta a la hipótesis planteada pueden diseñarse diferentes experimentos realizados con diferentes metodologías, protocolos, en diferentes épocas, diferentes grupos de investigación... Por ello, el experimento realizado por nuestro grupo de investigación no es más que uno de los tantos posibles. En otras palabras, que cada uno de los experimentos individuales que se han llevado a cabo para contestar una pregunta no son más que una muestra aleatoria extraída de la población de todos los posibles experimentos igualmente diseñados y, a su vez, extraído de otra población de experimentos con todos los posibles diferentes diseños experimentales capaces de responder a las mismas preguntas.

Por ello, el siguiente paso fue buscar la evidencia ya existente y, en particular, los datos originales que se usaron para llegar a ella con el fin de reanalizarlos siguiendo el mismo protocolo que en nuestro experimento transcriptómico en tejido humano. Así, el fin último será reunir tanto los resultados de los experimentos reanalizados como los obtenidos por nuestro grupo con el fin de obtener una estimación del efecto global más precisa y exacta sobre el estudio en cuestión. Tal y como se explicó en el apartado “5.1.2. Modelos de tejido humano” , se seleccionaron 5 experimentos transcriptómicos del repositorio público NCBI GEO en función del cumplimiento de los criterios de inclusión ya explicados en dicho apartado: muestras de tejido humano tanto de CCR como normal en cada experimento, al menos cinco réplicas para cada condición experimental, experimentos realizados con microarrays de *Affymetrix* y accesibilidad de los datos transcriptómicos originales *.CEL*.

En particular, las razones de que los microarrays empleados fueran de *Affymetrix* y de que estuvieran disponibles los datos originales sin procesar (*.CEL*) fueron, por un lado, poder reanalizar los datos transcriptómicos siguiendo el mismo protocolo que se ha mostrado en el apartado 7.2.1. *Análisis transcriptómico de muestras pareadas de tejido humano de cáncer de colon* con nuestras muestras y, por otro lado, evitar la heterogeneidad que pueda implicar juntar datos generados por plataformas (microarrays) de diferentes casas comerciales. Así, se han seleccionado las siguientes *Series* de NCBI GEO (tabla 7.8) y, de ellas, se han extraído los datos originales *.CEL*. Cabe destacar que previo a realizar la búsqueda, se planteó realizar un análisis de alrededor de 1000 microarrays y que, tras la selección de los distintos experimentos, este número se ha superado ligeramente hasta alcanzar un total de 1231 más los 20 generados por nuestro grupo.

Tabla 7.8: Series GEO seleccionadas.

| Series | Microarrays | | Pareado | Plataforma | Referencia |
|----------|-------------|---------|-----------------|---|-------------------------------|
| | Normal | Tumoral | | | |
| GSE8671 | 32 | 32 | Si | Affymetrix Human Genome U133 Plus 2.0 Array | Sabates-Bellver et al. (2007) |
| GSE39582 | 19 | 566 | No | Affymetrix Human Genome U133 Plus 2.0 Array | Marisa et al. (2013) |
| GSE33113 | 6 | 90 | No | Affymetrix Human Genome U133 Plus 2.0 Array | Felipe de Sousa et al. (2011) |
| GSE44076 | 148 | 98 | Si ¹ | Affymetrix Human Genome U219 Array | Solé et al. (2014) |
| GSE41258 | 54 | 186 | Si ¹ | Affymetrix Human Genome U133A Array | Sheffer et al. (2009) |

Notas:

¹ No todas las muestras son pareadas.

A continuación, en los siguientes subapartados se muestran los resultados de los reanálisis de los datos originales correspondientes a los experimentos tanto seleccionados y extraídos de NCBI GEO como los de nuestro grupo (apartado 7.2.1) y, además, de los metaanálisis realizados a partir de los reanálisis, tanto a nivel de expresión diferencial como de enriquecimiento de conjuntos de genes.

7.2.3.1. Reanálisis y metaanálisis de expresión diferencial

En primer lugar, cada uno de los datos originales *.CEL* de cada uno de los experimentos extraídos de NCBI GEO han sido reanalizados de forma independiente siguiendo el mismo protocolo de análisis que en el apartado “7.2.1. Análisis transcriptómico de muestras pareadas de tejido humano de cáncer de colon”. En la tabla 7.9 se muestra un resumen del tratamiento de los datos previo a los análisis de expresión diferencial y de enriquecimiento; es decir, las muestras que quedan tras el control de calidad, las sondas que tiene cada microarray previo al filtrado, y las sondas que quedan en la matriz de datos de expresión posterior al filtrado. Como puede observarse, por un lado, el control de calidad ha eliminado, en todos los casos, algún microarray (comparar la tabla 7.9 con la tabla 7.8) y, por otro lado, el filtrado ha mantenido alrededor de la mitad de las sondas en cada uno de los conjuntos de datos (tabla 7.9).

Tabla 7.9: Resumen del pre-tratamiento de los conjuntos de datos transcriptómicos seleccionados.

| Conjunto de Datos | Muestras tras QC | Sondas Pre-Filtrado | Sondas Post-Filtrado | | |
|-------------------|------------------|---------------------|----------------------|--------------------|----------------------|
| | | | Total | ALIAS ¹ | UNIGENE ² |
| Nuestras | 16 | 138745 | 69435 | 12944 | 12835 |
| GSE8671 | 47 | 54675 | 27607 | 22483 | 22152 |
| GSE39582 | 509 | 54675 | 27628 | 22507 | 22152 |
| GSE33113 | 90 | 54675 | 27624 | 22495 | 22152 |
| GSE44076 | 198 | 49386 | 25005 | 24733 | 24372 |
| GSE41258 | 204 | 22283 | 11297 | 10729 | 10532 |

Abreviaturas:

QC: control de calidad y detección de *outliers*.

Notas:

¹ Estas sondas tienen correspondencia con el identificador ALIAS, y no necesariamente 1:1.

² Estas sondas tienen correspondencia con el identificador ENTREZID, y no necesariamente 1:1.

Posteriormente, se seleccionó, de forma independiente para cada uno de los experimentos, el modelo más adecuado para explicar la variable respuesta $\log(\text{Expresin})$. Para ello, se partió del modelo completo, que es aquel que contiene todas las posibles variables explicativas -mostradas en las columnas *Factores Fijos:Niveles* y *Factores aleatorios* tabla 7.10-, y se seleccionó según los criterios AIC y BIC. Curiosamente, en todos los experimentos el modelo seleccionado fue aquel que sólo consideraba al factor fijo de interés -aquel que contiene las condiciones tumoral y normal- y , en caso de existir, al factor aleatorio correspondiente a los pacientes.

Tabla 7.10: Resumen de la selección de los modelos para cada conjunto de datos transcriptómicos.

| Conjunto de Datos | Factores Fijos:Niveles ¹ | Factores Aleatorios ¹ | Modelo Seleccionado ¹ |
|-------------------|---|----------------------------------|----------------------------------|
| Nuestras | Tejido: {Tumoral, Normal} Sexo: {Hombre, Mujer} | Paciente | Tejido+Paciente |
| GSE8671 | Tissue: {adenoma, normal} Location: {rectum...} | Patient | Tissue+Patient |
| GSE39582 | Stage: {S0, S1, S2, S3, S4, NonTumoral} Gender: {Male, Female} Age: [22, 92.4] | | Stage |
| GSE44076 | sampletype: {Tumor, Normal, Mucosa} gender: {Male, Female} age: [25, 88] | patient | sampletype + patient |
| GSE44076 | tissue: {tumor, normal} Sex: {m, f} Age: [34.64, 95.13] | | tissue |
| GSE41258 | tissue: {PrimaryTumor, NormalColon} Stage: {I, II, III, IV, NonTumoral} Gender: {Male, Female} Age: [19, 87] | patientid | tissue + patientid |

Notas:

¹ Se han usado los nombres originales de cada conjunto de datos.

Tras la selección del modelo, se realizó el análisis de expresión diferencial para cada uno de los experimentos, por lo que se obtiene un análisis de expresión diferencial diferente para cada uno de ellos. Posteriormente, a partir de estos resultados, se realizó un metaanálisis de expresión diferencial entre el tejido humano de CCR y normal. Para ello, de los diferentes métodos de metaanálisis existentes, se han escogido dos: el método del *tamaño del efecto moderado* (Choi et al. (2003), Marot et al. (2009)), en el que, mediante un modelo de efectos aleatorios, se combinan los efectos individuales de cada experimento y para cada gen se obtiene el efecto global μ_g y su p-valor asociado, y el método de la *combinación de los p valores*, en el que se calcula el *puntaje-Z* asociado a la combinación de cada uno de los p valores de una cola de cada experimento, que sigue una $N(1, 0)$, y su p-valor asociado.

No obstante, antes de combinar los valores de expresión diferencial obtenidos en cada uno de los experimentos, hay que tener en cuenta que no todos se han realizado en la misma plataforma -tipo de microarray- y, además, que cada plataforma tiene unas sondas específicas que no tienen por qué coincidir con las de otra plataforma. Es más, dentro de la misma plataforma pueden existir conjuntos de sondas redundantes -se corresponden con un mismo gen-. Por todo ello, para hacer

frente a esta situación, en primer lugar, y dentro de cada experimento, se combinan los resultados del análisis de expresión diferencial de aquellas sondas que tienen correspondencia con un mismo ALIAS (de igual modo a como se combinan los efectos en un metaanálisis, pero combinando el $\log(FC)$ en vez del tamaño del efecto). En segundo lugar, una vez se tienen los valores de expresión diferencial combinados y en la anotación ALIAS para cada uno de los experimentos, ya es posible combinar los resultados de expresión diferencial de estos.

En adelante, al experimento realizado por nuestro laboratorio mostrado en el apartado apartado “7.2.1. Análisis transcriptómico de muestras pareadas de tejido humano de cáncer de colon” se le denominará *IBGM*. Como resultado, en la tabla 7.11 se muestra un resumen de los reanálisis de expresión diferencial de cada uno de los experimentos extraídos de NCBI GEO, del análisis de expresión diferencial del experimento transcriptómico realizado por nuestro grupo en muestras pareadas de tejido humano de CCR y normal, y de los dos metaanálisis de expresión diferencial realizados a partir de estos. Como puede apreciarse en la tabla 7.11, tanto en los experimentos como en los metaanálisis se observa un claro patrón de sobreexpresión en las muestras correspondientes al tejido humano de CCR respecto del normal.

Tabla 7.11: Resumen del reanálisis y metaanálisis de expresión diferencial.

| Conjunto de Datos | Total Genes | DEG | Sobreexpresados | Infraexpresados |
|-----------------------------|-------------|-------|-----------------|-----------------|
| IBGM | 15505 | 5807 | 5295 | 512 |
| GSE8671 | 15864 | 12358 | 7150 | 5208 |
| GSE39582 | 15502 | 9138 | 4769 | 4369 |
| GSE33113 | 15420 | 7518 | 4042 | 3476 |
| GSE44076 | 14888 | 13434 | 7355 | 6079 |
| GSE41258 | 9964 | 7776 | 4428 | 3348 |
| Metaanálisis p Comb. | 21382 | 15190 | 8432 | 6600 |
| Metaanálisis Efectos | 21382 | 11166 | 6855 | 4311 |

Abreviaturas:

DEG: genes (sondas) diferencialmente expresados.

No obstante, como ha sido costumbre a lo largo de la tesis, el análisis de expresión diferencial y, en este caso, también el metaanálisis, se ha centrado en los genes implicados en la homeostasis del Ca^{2+} intracelular. Por ello, a continuación se muestran, divididos en las familias de genes VOCC, SOCE, TRP, CPT, ERC y MCT, los resultados del metaanálisis en mediante diagramas de barras y en las tablas B.51-B.56 del apéndice B.4.1 los resultados de los reanálisis de cada uno de los experimentos extraídos de GEO así como del metaanálisis resultante. Cabe destacar que tanto en las gráficas como en las tablas, los valores correspondientes a los metaanálisis no son los valores de $\log(FC)$, sino el tamaño del efecto o del puntaje-Z en función del método de metaanálisis empleado. Además, se ha considerado que un gen está diferencialmente expresado si lo está con al menos uno de los dos métodos para un nivel de significación $\alpha_{FDR} = 0.05$ o si es inferior a $\alpha_{FDR} = 0.05$. Como se verá en las siguientes figuras, hay un claro remodelado de la homeostasis del Ca^{2+} intracelular, hecho que apoya la hipótesis planteada en la presente tesis.

Como puede observarse tanto en la figura 7.48 como en la tabla B.51 (apéndice B.4.1), que muestra los resultados de los genes pertenecientes a la familia VOCC, se ha detectado la infraexpresión en el tejido de CCR humano de los genes *CAV2.1* y *CAV3.2* y una sobreexpresión del *CAV1.3*. Este perfil de expresión recuerda al de las líneas celulares en las que también se identificó una mayor expresión de *CAV1.3* y menor de *CAV3.2* en el fenotipo tumoral. Al igual que en las líneas, también se detecta una infraexpresión de una isoforma de *CAV2*, *CAV2.1*, en el tejido tumoral. Cabe destacar que en el experimento *IBGM* no se detectó expresión diferencial de ningún VOCC (tabla 7.12).

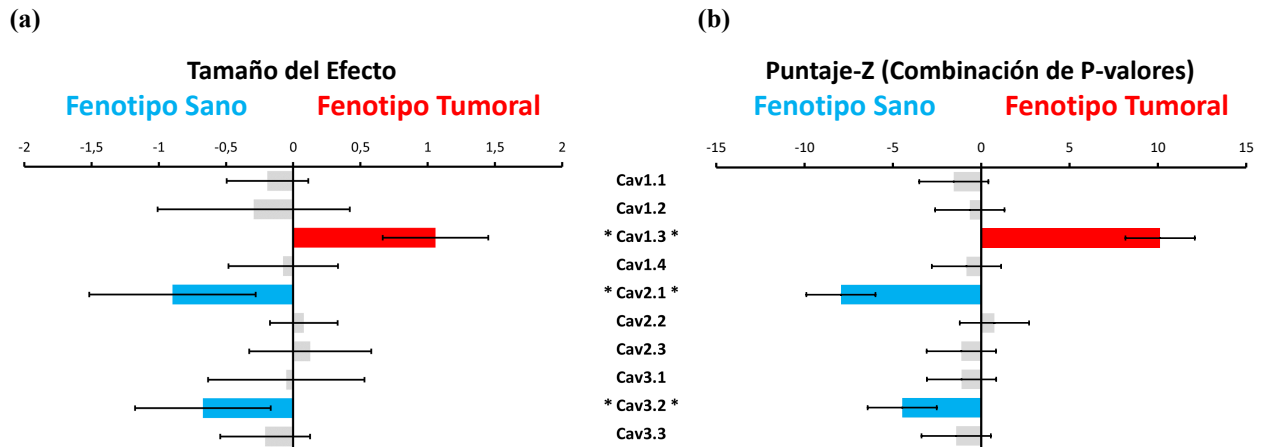


Figura 7.48: Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes VOCC. Los valores representados son (a) el tamaño del efecto moderado, donde los genes que muestran expresión diferencial significativa se indican mediante "*" para $FDR < 0.05$ y "." para $FDR < 0.1$ a la izquierda del nombre del gen, y (b) el puntaje-Z asociado a la combinación de los p valores de cada experimento, cuya expresión diferencial se indica a la derecha del nombre del gen. Además, para un $\alpha_{FDR} = 0.05$, las barras azules hacia la izquierda indican una menor expresión significativa en el tejido tumoral y una sobreexpresión para las barras rojas hacia la derecha.

Tabla 7.12: Comparación entre modelos del análisis de expresión diferencial para VOCC.

| Gen | Modelos | | | DFMO |
|---------------|------------------|------|--------------|------|
| | Líneas Celulares | IBGM | Metaanálisis | |
| <i>CAV1.2</i> | ↓ | | | |
| <i>CAV1.3</i> | ↑ | | | ↑ |
| <i>CAV2.1</i> | | | | ↓ |
| <i>CAV2.3</i> | ↓ | | | |
| <i>CAV3.2</i> | ↓ | | | ↓ |

En cuanto a los genes de la familia SOCE, por un lado, se observa sobreexpresión de *ORAI1*, *ORAI2*, *ORAI3* e infraexpresión de *STIM1* y *MS4A12* (figura 7.49, tabla B.52). Este perfil de expresión no coincide con el encontrado en las líneas celulares, salvo por la mayor expresión de *ORAI2* en el fenotipo tumoral. Además, la sobreexpresión de *ORAI1* e infraexpresión de *STIM1* en el tejido de cáncer de colon muestra un perfil contrario al de los modelos celulares. Respecto del

experimento transcriptómico *IBGM* que, recordemos, es una especie de “muestra” empleada en el metaanálisis, también muestra una disminución de *MS4A12* y *STIM1* en el tejido de cáncer de colon. Cabe destacar que el metaanálisis, a diferencia de los experimentos en líneas celulares e *IBGM*, es capaz de identificar también la sobreexpresión de *ORAI3*.

Por otro lado, en relación a los reguladores de SOCE, se ha identificado la sobreexpresión de *CRACR2A*, *SEPTIN2,4,7* y la disminución de *STIMATE*, *SARAF*, *MBP* y *SEPTIN1,6,8,10,11*. (figura 7.49, tabla 7.13). La expresión diferencial encontrada para *SEPTIN1,2,4,6,7* y *SARAF* coincide con lo observado en las líneas celulares, si bien es cierto que la expresión diferencial de *CRACR2A*, *MBP*, *SEPTIN10,11* muestra un perfil de expresión totalmente opuesto. Cabe destacar que a pesar de que en los modelos celulares se observó expresión diferencial de *SEPTIN3,9*, el metaanálisis no detectó tal comportamiento de ninguno de ellos. En cuanto a su similitud con el experimento *IBGM*, el metaanálisis también ha identificado la sobreexpresión de *SEPTIN2,7* infraexpresión de *SEPTIN1,6*; sin embargo, para *SARAF*, *SEPTIN8,11* el sentido de la expresión diferencial es el contrario. Curiosamente, mientras que tanto en los tejidos como en el experimento *IBGM* se ha identificado la expresión diferencial de *ORAMD3* (en sentido contrario en cada caso), no se observó tal efecto en el metaanálisis. Por el contrario, mediante el metaanálisis, a diferencia de los otros dos experimentos, sí se identifica la infraexpresión de *STIMATE* en el tejido de cáncer de colon (tabla 7.13).

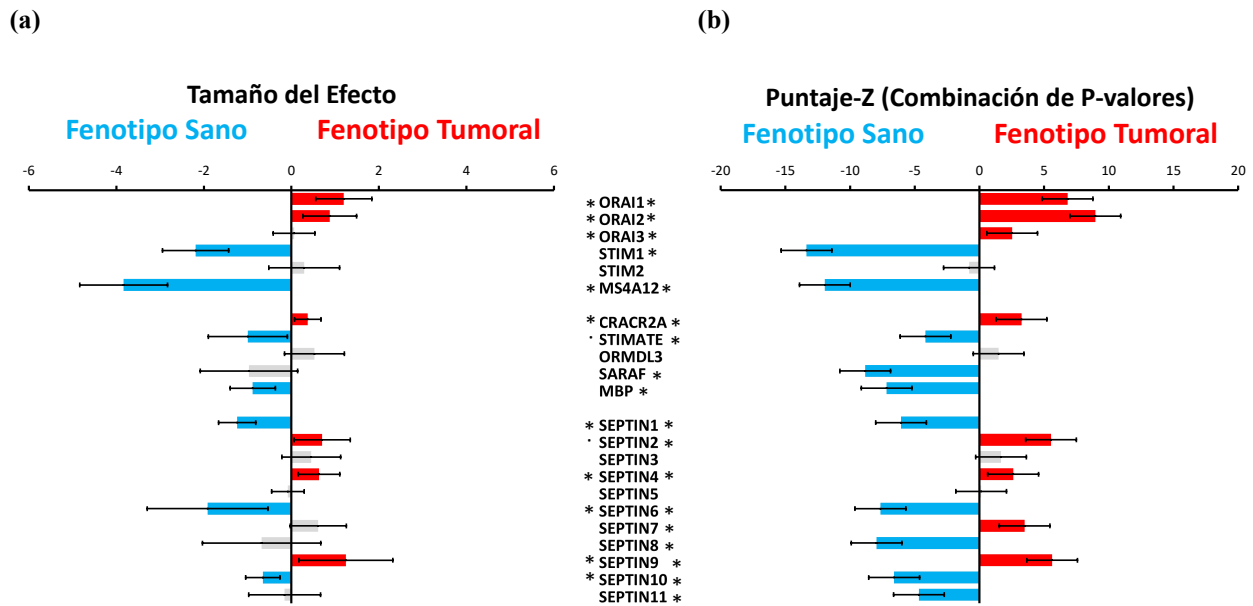


Figura 7.49: Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes SOCE. Los valores representados son (a) el tamaño del efecto moderado, donde los genes que muestran expresión diferencial significativa se indican mediante “*” para $FDR < 0.05$ y “.” para $FDR < 0.1$ a la izquierda del nombre del gen, y (b) el puntaje-Z asociado a la combinación de los p valores de cada experimento, cuya expresión diferencial se indica a la derecha del nombre del gen. Además, para un $\alpha_{FDR} = 0.05$, las barras azules hacia la izquierda indican una menor expresión significativa en el tejido tumoral y una sobreexpresión para las barras rojas hacia la derecha.

Tabla 7.13: Comparación entre modelos del análisis de expresión diferencial para SOCE.

| Gen | Modelos | | | DFMO |
|-----------------|------------------|------|--------------|------|
| | Líneas Celulares | IBGM | Metaanálisis | |
| <i>ORAI1</i> | ↓ | | ↑ | |
| <i>ORAI2</i> | ↑ | | ↑ | |
| <i>ORAI3</i> | | | ↑ | |
| <i>STIM1</i> | ↑ | ↓ | ↓ | ↑ |
| <i>STIM2</i> | | | | ↑ |
| <i>MS4A12</i> | | ↓ | ↓ | |
| <i>CRACR2A</i> | ↓ | | ↑ | ↑ |
| <i>STIMATE</i> | | | ↓ | |
| <i>ORMDL3</i> | ↓ | ↑ | | ↑ |
| <i>SARAF</i> | ↓ | ↑ | ↓ | |
| <i>MBP</i> | ↑ | | ↓ | |
| <i>SEPTIN1</i> | ↓ | ↓ | ↓ | |
| <i>SEPTIN2</i> | ↑ | ↑ | ↑ | |
| <i>SEPTIN3</i> | ↓ | | | |
| <i>SEPTIN4</i> | ↑ | | ↑ | |
| <i>SEPTIN6</i> | ↓ | ↓ | ↓ | ↑ |
| <i>SEPTIN7</i> | ↑ | ↑ | ↑ | |
| <i>SEPTIN8</i> | | ↑ | ↓ | ↑ |
| <i>SEPTIN9</i> | ↑ | | ↑ | ↑ |
| <i>SEPTIN10</i> | ↑ | | ↓ | |
| <i>SEPTIN11</i> | ↑ | ↑ | ↓ | |

En el caso de los TRP, se observa un claro patrón de infraexpresión en el tejido tumoral respecto del normal. En concreto, mediante, al menos, uno de los dos métodos empleados se ha detectado la infraexpresión de los genes *TRPC1*, *TRPC6*, *TRPM1,4-8*, *TRPML1*, *TRPML2*, *TRPA1*, *TRPV1*, *TRPV3* y *TRPP1* (figura 7.50, tabla 7.14). Además, también se ha observado una sobreexpresión de los genes *TRPC7*, *TRPV4* y *TRPP3* en el tejido tumoral (figura 7.50, tabla 7.14). Por un lado, la expresión diferencial de algunos de estos TRP no se ha identificado en los modelos celulares: *TRPC1,6*, *TRPM1,4,6*, *TRPV3,4*. Por el contrario, no detecta la expresión diferencial de *TRPC5*, *TRPPM2,3*, *TRPML3*, *TRPV6* ni *TRPP5*. Además, la expresión diferencial de *TRPC7*, *TRPM7*, *TRPP1,3* observada en los modelos celulares es la contraria a la obtenida por el metaanálisis. Por otro lado, en el experimento transcriptómico *IBGM* tampoco se ha detectado la expresión diferencial de ciertos TRP que sí han sido identificados en el metaanálisis: *TRPC1,6*, *TRPM4-8*, *TRPML1,2*, *TRPA1*, *TRPV3,4*, *TRPP3*. Por el contrario, no se detecta expresión diferencial de *TRPM2*, *TRPV2* ni de *TRPP5*. Además, la dirección de la expresión diferencial de *TRPC7*, *TRPV1* y *TRPP1* en dicho experimento es la opuesta a la del metaanálisis (tabla 7.14).

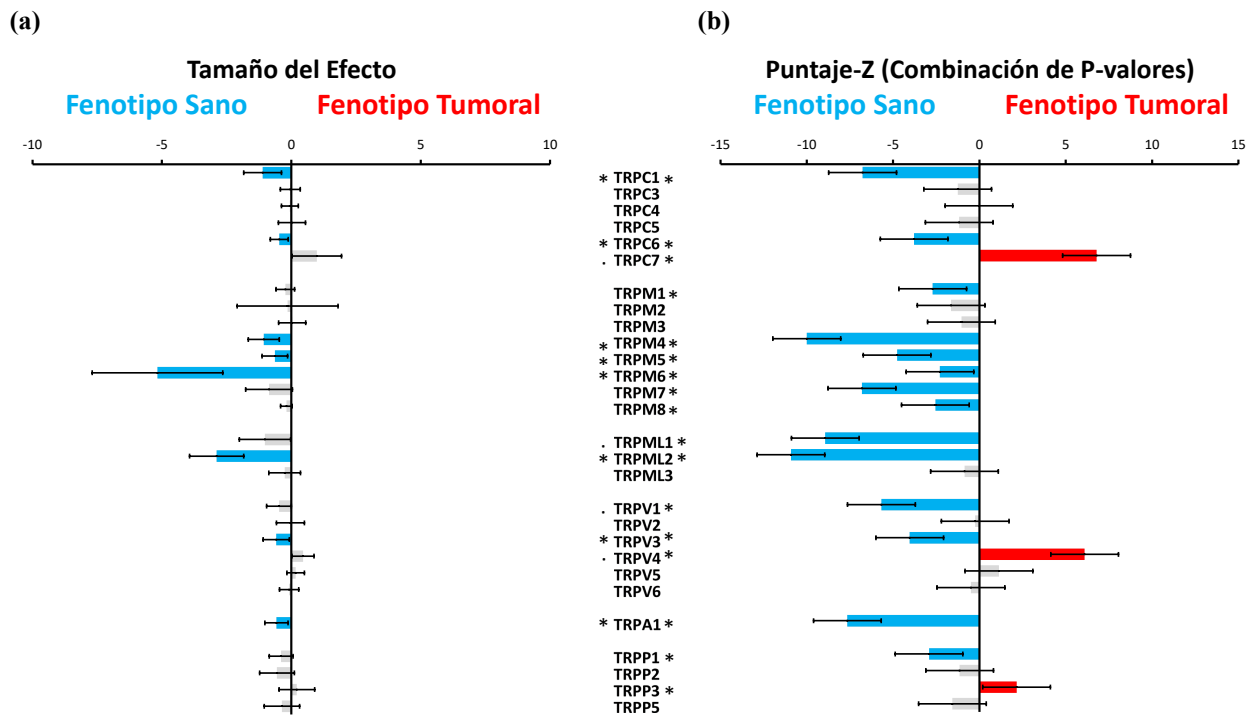


Figura 7.50: Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes TRP. Los valores representados son (a) el tamaño del efecto moderado, donde los genes que muestran expresión diferencial significativa se indican mediante "*" para $FDR < 0.05$ y "." para $FDR < 0.1$ a la izquierda del nombre del gen, y (b) el puntaje-Z asociado a la combinación de los p valores de cada experimento, cuya expresión diferencial se indica a la derecha del nombre del gen. Además, para un $\alpha_{FDR} = 0.05$, las barras azules hacia la izquierda indican una menor expresión significativa en el tejido tumoral y una sobreexpresión para las barras rojas hacia la derecha.

Tabla 7.14: Comparación entre diferentes modelos del análisis de expresión diferencial para TRP.

| Gen | Modelos | | | DFMO |
|---------------|------------------|------|--------------|------|
| | Líneas Celulares | IBGM | Metaanálisis | |
| <i>TRPC1</i> | | | ↓ | ↓ |
| <i>TRPC5</i> | ↑ | | | ↓ |
| <i>TRPC6</i> | | | ↓ | |
| <i>TRPC7</i> | ↓ | ↓ | ↑ | |
| <i>TRPM1</i> | | ↓ | ↓ | |
| <i>TRPM2</i> | ↓ | ↓ | | |
| <i>TRPM3</i> | ↓ | | | |
| <i>TRPM4</i> | | | ↓ | |
| <i>TRPM5</i> | ↓ | | ↓ | |
| <i>TRPM6</i> | | | ↓ | |
| <i>TRPM7</i> | ↑ | | ↓ | |
| <i>TRPM8</i> | ↓ | | ↓ | |
| <i>TRPML1</i> | ↓ | | ↓ | |
| <i>TRPML2</i> | ↓ | | ↓ | |
| <i>TRPML3</i> | ↓ | | | |
| <i>TRPV1</i> | ↓ | ↑ | ↓ | |
| <i>TRPV2</i> | | ↑ | | |
| <i>TRPV3</i> | | | ↓ | |
| <i>TRPV4</i> | | | ↑ | |
| <i>TRPV6</i> | ↑ | | | ↓ |
| <i>TRPA1</i> | ↓ | | ↓ | |
| <i>TRPP1</i> | ↑ | ↑ | ↓ | ↓ |
| <i>TRPP2</i> | ↓ | | | ↑ |
| <i>TRPP3</i> | ↓ | | ↑ | |
| <i>TRPP5</i> | ↓ | ↓ | | |

En cuanto a los genes de la familia CPT, hay un claro patrón de infraexpresión en el tejido tumoral de algunos de ellos: *PMCA1*, *PMCA3*, *PMCA4*, *SERCA2*, *SERCA3*, *NCX2* y *SPCA2*. No obstante, también se observa la sobreexpresión de *SPCA1* (figura 7.51, tabla B.54). La expresión diferencial de *SPCA1* coincide con lo observado tanto en los modelos celulares como en el experimento transcriptómico *IBGM*. Además, en los modelos celulares también coincide la expresión diferencial de *PMCA4*. No obstante, la expresión diferencial de *PMCA1*, *SERCA2*, *NCX2* y *SPCA2* está en sentido contrario a lo observado en el metaanálisis, mientras que en el experimento transcriptómico *IBGM* está en sentido contrario la expresión diferencial de *SERCA2* y *SPCA2*. Además, a diferencia de lo encontrado en los modelos celulares, el metaanálisis también detecta la expresión diferencial de *PMCA1* y *SERCA3*, mientras que, en comparación con el experimento transcriptómico *IBGM*, el metaanálisis es capaz de identificar la expresión diferencial de *PMCA1,3,4*, *NCX2* y *SERCA3* (tabla 7.15).

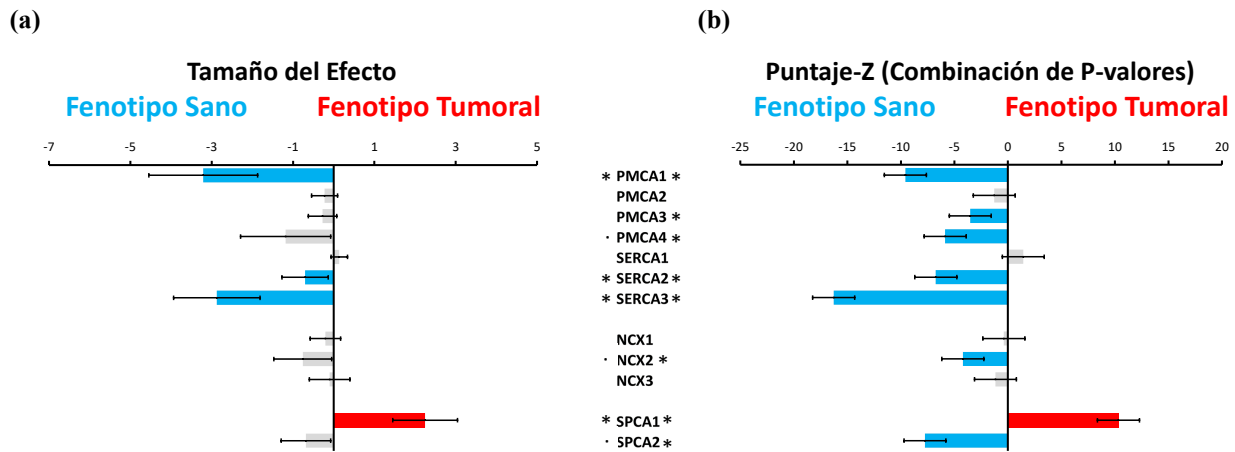


Figura 7.51: Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes CPT. Los valores representados son (a) el tamaño del efecto moderado, donde los genes que muestran expresión diferencial significativa se indican mediante "*" para $FDR < 0.05$ y "." para $FDR < 0.1$ a la izquierda del nombre del gen, y (b) el puntaje-Z asociado a la combinación de los p valores de cada experimento, cuya expresión diferencial se indica a la derecha del nombre del gen. Además, para un $\alpha_{FDR} = 0.05$, las barras azules hacia la izquierda indican una menor expresión significativa en el tejido tumoral y una sobreexpresión para las barras rojas hacia la derecha.

Tabla 7.15: Comparación entre modelos del análisis de expresión diferencial para CPT.

| Gen | Modelos | | | DFMO |
|---------------|------------------|------|--------------|------|
| | Líneas Celulares | IBGM | Metaanálisis | |
| <i>PMCA1</i> | ↑ | | ↓ | |
| <i>PMCA3</i> | | | ↓ | |
| <i>PMCA4</i> | ↓ | | ↓ | ↑ |
| <i>SERCA2</i> | ↑ | ↑ | ↓ | |
| <i>SERCA3</i> | | | ↓ | |
| <i>NCX2</i> | ↑ | | ↓ | |
| <i>SPCA1</i> | ↓ | ↑ | ↑ | |
| <i>SPCA2</i> | ↑ | ↑ | ↓ | ↓ |

De los genes que codifican para canales de liberación de Ca^{2+} desde el RE, se ha identificado la infraexpresión de *IP3R1* y *RyR3*, y la sobreexpresión de *IP3R2* (figura 7.52, tabla B.55). Estos resultados son muy diferentes de los encontrados en los modelos celulares y en el experimento transcriptómico *IBGM*. Por un lado, en los modelos celulares no se identifica expresión diferencial de *RyR3* pero sí, aunque en sentido contrario que el metaanálisis, de *IP3R1,2*, mientras que, a diferencia del metaanálisis, sí se detecta expresión diferencial de *IP3R3* y *RyR2*. Por otro lado, en el experimento *IBGM* la expresión diferencial de *IP3R1* coincide con lo observado en el metaanálisis, si bien es cierto que la expresión diferencial de *IP3R3* observada en dicho experimento no se detecta en el metaanálisis. Además, el metaanálisis es capaz de identificar la expresión diferencial de *IP3R2* y *RyR3* (tabla 7.16).

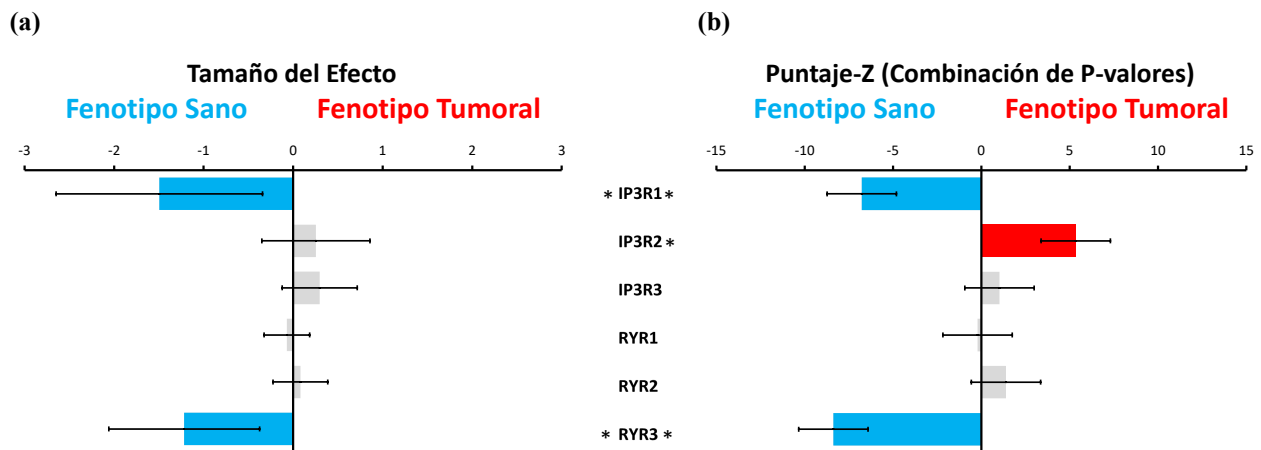


Figura 7.52: Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes ERC. Los valores representados son (a) el tamaño del efecto moderado, donde los genes que muestran expresión diferencial significativa se indican mediante "*" para $\text{FDR} < 0.05$ y "" para $\text{FDR} < 0.1$ a la izquierda del nombre del gen, y (b) el puntaje-Z asociado a la combinación de los p valores de cada experimento, cuya expresión diferencial se indica a la derecha del nombre del gen. Además, para un $\alpha_{\text{FDR}} = 0.05$, las barras azules hacia la izquierda indican una menor expresión significativa en el tejido tumoral y una sobreexpresión para las barras rojas hacia la derecha.

Tabla 7.16: Comparación entre diferentes análisis de expresión diferencial para CLC.

| Gen | Modelos | | | DFMO |
|--------------|------------------|------|--------------|------|
| | Líneas Celulares | IBGM | Metaanálisis | |
| <i>IP3R1</i> | ↑ | ↓ | ↓ | |
| <i>IP3R2</i> | ↓ | | ↑ | |
| <i>IP3R3</i> | ↑ | ↑ | | |
| <i>RYR2</i> | ↓ | | | |
| <i>RYR3</i> | | | ↓ | |

En cuanto al transporte de Ca^{2+} a través de la mitocondria, es evidente la clara expresión diferencial que estos presentan. Se observa infraexpresión significativa de *MCU*, *MICU1-3*, *EMRE*, *VDAC2* y *NCLX*, y sobreexpresión de *MCUb* y *VDAC1* (figura 7.53, tabla 7.17). Estos resultados difieren de los encontrados tanto en los modelos celulares como en el experimento transcriptómico *IBGM*. Por un lado, respecto de los modelos celulares, únicamente coinciden *MICU2* y *NCLX*, mientras que *MCU*, *MICU1*, *MCUb* y *VDAC1,2* muestran un patrón de expresión diferencial contrario. A mayores, el metaanálisis detecta la expresión diferencial de *MICU3* y *EMRE*. Por otro lado, respecto del experimento *IBGM*, sólo coinciden *VDAC1*, mientras que la expresión diferencial de *MCU*, *MICU1,2*, *VDAC2* y *NCLX* es contraria. Además, a diferencia del experimento *IBGM*, el metaanálisis detecta expresión diferencial de *MICU3*, *MCUb* y *EMRE* (tabla 7.17).

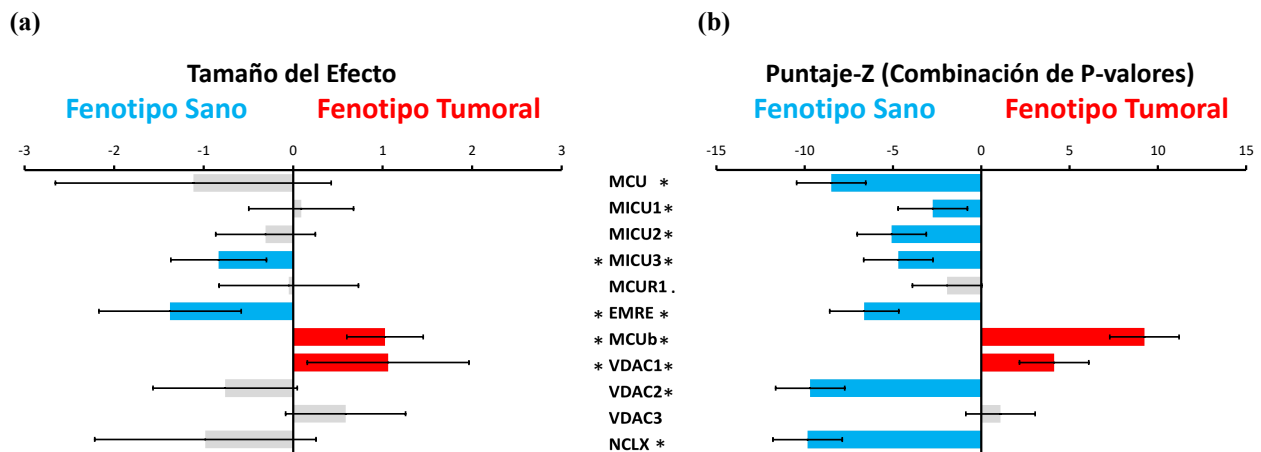


Figura 7.53: Metaanálisis de la expresión diferencial en tejido humano de CCR vs. sano del conjunto de genes MCT. Los valores representados son (a) el tamaño del efecto moderado, donde los genes que muestran expresión diferencial significativa se indican mediante "*" para $FDR < 0.05$ y "." para $FDR < 0.1$ a la izquierda del nombre del gen, y (b) el puntaje-Z asociado a la combinación de los p valores de cada experimento, cuya expresión diferencial se indica a la derecha del nombre del gen. Además, para un $\alpha_{FDR} = 0.05$, las barras azules hacia la izquierda indican una menor expresión significativa en el tejido tumoral y una sobreexpresión para las barras rojas hacia la derecha.

Tabla 7.17: Comparación entre diferentes análisis de expresión diferencial para MCT.

| Gen | Modelos | | | DFMO |
|--------------|------------------|------|--------------|------|
| | Líneas Celulares | IBGM | Metaanálisis | |
| <i>MCU</i> | ↑ | ↑ | ↓ | ↑ |
| <i>MICU1</i> | ↑ | ↑ | ↓ | |
| <i>MICU2</i> | ↓ | ↑ | ↓ | |
| <i>MICU3</i> | | | ↓ | |
| <i>MCUR1</i> | ↑ | ↑ | | |
| <i>EMRE</i> | | | ↓ | |
| <i>MCUb</i> | ↓ | | ↑ | |
| <i>VDAC1</i> | ↓ | ↑ | ↑ | |
| <i>VDAC2</i> | ↑ | ↑ | ↓ | |
| <i>VDAC3</i> | ↓ | ↑ | | ↑ |
| <i>NCLX</i> | ↓ | ↑ | ↓ | |

A continuación se presenta un resumen esquemático del análisis de expresión diferencial obtenido a través del metaanálisis (7.54).

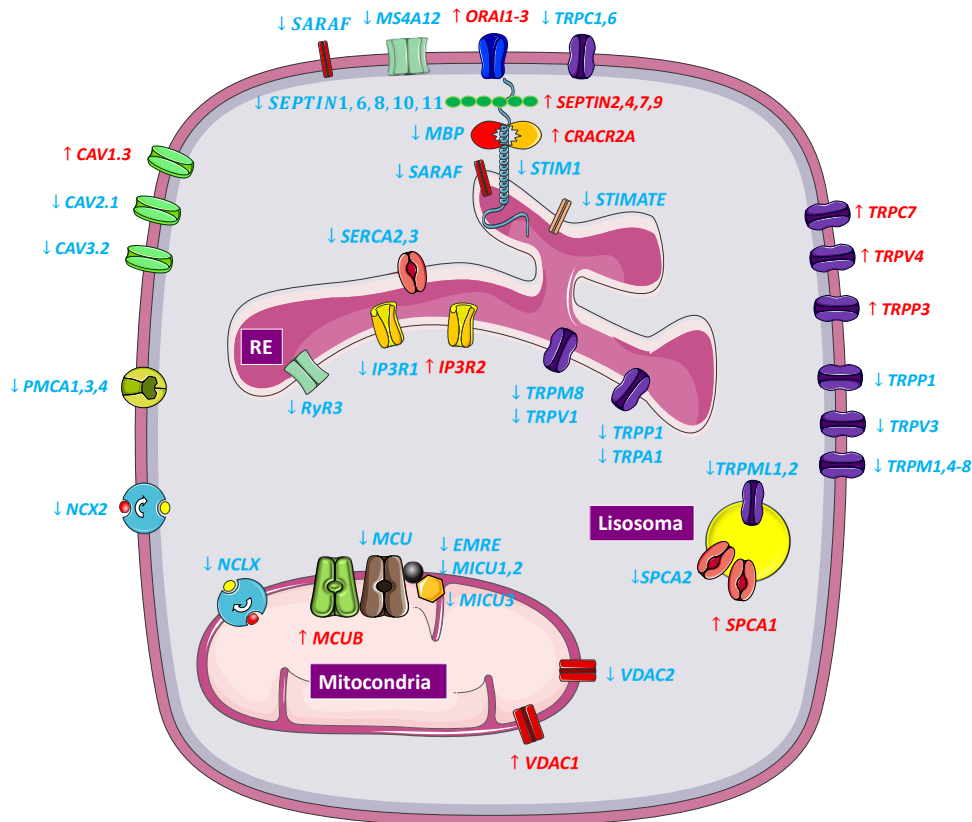


Figura 7.54: Resumen del metaanálisis de expresión diferencial entre el tejido de cáncer de colon y normal de los genes relacionados con el transporte de Ca^{2+} intracelular. En la representación esquemática de la célula tumoral se muestran aquellos genes que están diferencialmente expresados entre el tejido tumoral y normal de colon. En rojo se representan los que están sobreexpresados y en azul aquellos infraexpresados

7.2.3.2. Metaanálisis de enriquecimiento

Tras el metaanálisis de expresión diferencial, se procedió a realizar un análisis de enriquecimiento a partir los resultados obtenidos de cada uno de los experimentos transcriptómicos reanalizados. No obstante, en vez de usar la lista de genes obtenida en el metaanálisis anterior, se optó por realizar el análisis de enriquecimiento de forma independiente en cada uno de los experimentos empleados en el metaanálisis, tanto los extraídos de NCBI GEO como el realizado por nuestro grupo en muestras pareadas de tejido humano de CCR (tablas 7.19-7.21). Cabe destacar que, al igual que en los análisis de enriquecimiento anteriores, para cada uno de los experimentos se han contrastado las hipótesis nulas *autocontenida* y *competitiva* con los métodos *CAMERA* y *ORA*, respectivamente.

Tras obtener los resultados del análisis de enriquecimiento para cada uno de los experimentos, se realizó el metaanálisis de enriquecimiento combinando sus p valores (unilaterales y sin corregir) para así obtener un único p-valor global para cada uno de los conjuntos de genes. Por último, estos p valores fueron corregidos controlando el FDR mediante la aproximación de Benjamini and Hochberg (1995) y se fijó el nivel de significación en $\alpha_{FDR} = 0.05$. En la tabla 7.18 se muestran tanto el reanálisis de enriquecimiento independiente para cada experimento como el metaanálisis de enriquecimiento. Evidentemente, los resultados obtenidos con cada uno de los experimentos es diferente; de ahí la importancia de realizar un metaanálisis a partir de toda esta evidencia. Se han identificado 9115 conjuntos de genes que rechazan la *hipótesis nula autocontenida* y 3588 que rechazan *competitiva*.

No obstante, siguiendo el mismo planteamiento que en los análisis de enriquecimiento realizados a lo largo de la tesis se evaluó cuántos de los conjuntos de genes que muestran asociación con el fenotipo de CCR humano pertenecen a las bases de datos relacionadas directamente con el cáncer -MCC y MOG- (tabla 7.18) y/o contienen en su nombre alguna de las 6 palabras clave *calcium, mitochondria, reticulum, cancer, wnt* y *polyamine* (tablas de 7.19 a 7.24, respectivamente). Cabe destacar que el análisis de enriquecimiento se ha realizado de forma independiente tanto con cada uno de los experimentos extraídos de GEO como con el experimento *IBGM* y, posteriormente, combinando, para cada gen, los p valores de una cola de los diferentes experimentos se ha realizado el “metaanálisis de enriquecimiento” .

Tabla 7.18: Conjuntos de genes que rechazan alguna hipótesis nula tras el reanálisis y metaanálisis de enriquecimiento.

| Base de Datos | Método | nº de Conjuntos con un FDR <0.05(0.1) | | | | | | |
|---------------|--------|---------------------------------------|------------|------------|------------|------------|------------|--------------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis |
| KEGG | CAMERA | 143(174) | 169(195) | 123(173) | 158(206) | 141(175) | 117(153) | 233(257) |
| | ORA | 46(65) | 33(60) | 62(101) | 22(39) | 26(42) | 47(66) | 46(60) |
| GOBP | CAMERA | 958(1449) | 1414(1935) | 1005(1520) | 1130(1628) | 818(1198) | 904(1383) | 1723(2128) |
| | ORA | 452(794) | 248(532) | 366(674) | 55(174) | 137(402) | 393(737) | 121(210) |
| GOMF | CAMERA | 323(461) | 449(619) | 347(518) | 407(584) | 294(411) | 282(437) | 604(735) |
| | ORA | 144(228) | 70(158) | 109(210) | 14(52) | 47(127) | 107(210) | 24(48) |
| MH | CAMERA | 26(31) | 28(32) | 21(27) | 23(27) | 21(25) | 18(24) | 33(40) |
| | ORA | 13(15) | 9(13) | 11(14) | 9(12) | 11(12) | 13(16) | 13(22) |
| MCDB | CAMERA | 2274(2906) | 2591(3239) | 1906(2562) | 2131(2797) | 1951(2581) | 1813(2452) | 3419(3885) |
| | ORA | 1525(1945) | 990(1532) | 1202(1742) | 415(803) | 619(1119) | 1149(1787) | 1355(1911) |
| MCC | CAMERA | 524(584) | 482(556) | 311(390) | 450(529) | 378(460) | 318(392) | 578(645) |
| | ORA | 264(299) | 209(269) | 161(212) | 110(171) | 200(280) | 144(218) | 214(273) |
| MOG | CAMERA | 85(103) | 93(111) | 52(67) | 61(91) | 60(84) | 52(74) | 106(125) |
| | ORA | 35(44) | 22(36) | 40(47) | 19(29) | 21(42) | 27(38) | 37(46) |
| MIM | CAMERA | 1897(2436) | 2220(2781) | 1073(1672) | 1596(2183) | 1346(1944) | 1263(1857) | 2419(2885) |
| | ORA | 1891(2297) | 1816(2423) | 1060(1596) | 734(1314) | 596(1066) | 962(1414) | 1778(2361) |
| Total | CAMERA | 6227(8141) | 7444(9466) | 4838(6927) | 5956(8045) | 5008(6876) | 4765(6770) | 9113(10697) |
| | ORA | 4370(5687) | 3397(5023) | 3011(4596) | 1378(2594) | 1657(3090) | 2842(4486) | 3588(4931) |

Abrebiaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Onncogenic; MIM: Msigdb Immunology.

Tabla 7.19: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *calcium*.

| Base de Datos | Método | nº de Conjuntos con un FDR <0.05(0.1) | | | | | | |
|---------------|--------|---------------------------------------|---------|----------|----------|----------|----------|--------------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis |
| KEGG | CAMERA | 2(2) | 2(2) | 1(2) | 2(2) | 2(2) | 2(2) | 2(2) |
| | ORA | 0(0) | 0(1) | 0(1) | 0(0) | 1(1) | 1(2) | 1(1) |
| GOBP | CAMERA | 28(37) | 31(37) | 20(27) | 23(29) | 19(22) | 22(28) | 30(34) |
| | ORA | 1(5) | 3(7) | 6(9) | 0(2) | 1(3) | 11(15) | 2(2) |
| GOMF | CAMERA | 8(10) | 7(10) | 3(6) | 4(8) | 7(9) | 5(6) | 11(13) |
| | ORA | 0(3) | 0(1) | 1(2) | 0(0) | 0(1) | 3(3) | 0(0) |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MCDB | CAMERA | 7(8) | 4(8) | 5(5) | 5(6) | 5(5) | 4(5) | 7(10) |
| | ORA | 1(1) | 0(1) | 1(1) | 0(0) | 0(0) | 2(3) | 0(0) |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| Total | CAMERA | 45(57) | 44(57) | 29(40) | 34(45) | 33(38) | 33(41) | 50(59) |
| | ORA | 2(9) | 3(10) | 8(13) | 0(2) | 2(5) | 17(23) | 3(3) |

Tabla 7.20: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *cancer*.

| Base de Datos | Método | nº de Conjuntos con un FDR <0.05(0.1) | | | | | | |
|---------------|--------|---------------------------------------|----------|----------|----------|----------|----------|--------------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis |
| KEGG | CAMERA | 0(3) | 4(6) | 1(6) | 1(3) | 1(3) | 1(1) | 5(8) |
| | ORA | 0(2) | 0(0) | 4(5) | 0(0) | 0(2) | 1(1) | 1(4) |
| GOBP | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 10(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| GOMF | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MCDB | CAMERA | 214(267) | 212(261) | 168(217) | 187(241) | 202(252) | 172(226) | 291(327) |
| | ORA | 143(184) | 92(133) | 117(165) | 44(86) | 78(126) | 122(167) | 146(196) |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| Total | CAMERA | 214(270) | 216(267) | 169(223) | 188(244) | 203(255) | 173(227) | 296(335) |
| | ORA | 143(186) | 92(133) | 121(170) | 44(86) | 78(128) | 123(168) | 147(200) |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Onncogetic; MIM: Msigdb Immunology.

Tabla 7.21: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *WNT*.

| Base de Datos | Método | nº de Conjuntos con un FDR <0.05(0.1) | | | | | | |
|---------------|--------|---------------------------------------|---------|----------|----------|----------|----------|--------------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis |
| KEGG | CAMERA | 0(0) | 0(0) | 0(1) | 0(1) | 0(0) | 0(0) | 0(1) |
| | ORA | 0(1) | 1(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| GOBP | CAMERA | 3(5) | 5(6) | 3(3) | 5(6) | 2(3) | 3(5) | 5(7) |
| | ORA | 3(4) | 3(5) | 1(2) | 0(0) | 1(4) | 0(1) | 0(0) |
| GOMF | CAMERA | 0(0) | 1(1) | 0(1) | 0(0) | 0(1) | 0(2) | 0(0) |
| | ORA | 0(0) | 0(1) | 0(0) | 0(0) | 1(2) | 0(0) | 0(0) |
| MH | CAMERA | 1(1) | 0(0) | 1(1) | 1(1) | 0(1) | 0(0) | 1(1) |
| | ORA | 1(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MCDB | CAMERA | 10(16) | 11(15) | 15(16) | 11(18) | 13(17) | 7(12) | 17(22) |
| | ORA | 11(17) | 13(16) | 2(4) | 2(6) | 1(5) | 4(5) | 4(9) |
| MCC | CAMERA | 0(1) | 0(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MOG | CAMERA | 0(0) | 1(1) | 0(0) | 0(1) | 0(1) | 0(1) | 1(1) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MIM | CAMERA | 0(3) | 1(1) | 0(0) | 0(0) | 0(1) | 0(1) | 2(2) |
| | ORA | 1(2) | 0(1) | 1(3) | 1(2) | 0(1) | 0(0) | 1(2) |
| Total | CAMERA | 14(26) | 19(25) | 19(22) | 17(27) | 15(24) | 10(21) | 26(34) |
| | ORA | 16(25) | 17(24) | 4(9) | 3(8) | 3(12) | 4(6) | 5(11) |

Tabla 7.22: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *mitochondria*.

| Base de Datos | Método | nº de Conjuntos con un FDR <0.05(0.1) | | | | | | |
|---------------|--------|---------------------------------------|---------|----------|----------|----------|----------|--------------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis |
| KEGG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| GOBP | CAMERA | 16(29) | 24(29) | 21(29) | 20(29) | 16(19) | 19(24) | 32(36) |
| | ORA | 14(27) | 7(19) | 5(8) | 1(5) | 1(6) | 4(8) | 0(2) |
| GOMF | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MCDB | CAMERA | 13(16) | 14(14) | 19(22) | 16(21) | 14(16) | 7(11) | 24(28) |
| | ORA | 13(17) | 6(10) | 8(12) | 2(6) | 1(5) | 4(9) | 9(13) |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| Total | CAMERA | 29(45) | 38(43) | 40(51) | 36(50) | 30(35) | 26(35) | 56(64) |
| | ORA | 27(44) | 13(29) | 13(20) | 3(11) | 2(11) | 8(17) | 9(15) |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Tabla 7.23: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *reticulum*.

| Base de Datos | Método | nº de Conjuntos con un FDR <0.05(0.1) | | | | | | |
|---------------|--------|---------------------------------------|---------|----------|----------|----------|----------|--------------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis |
| KEGG | CAMERA | 1(1) | 0(0) | 0(0) | 1(1) | 0(1) | 0(1) | 1(1) |
| | ORA | 1(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| GOBP | CAMERA | 9(15) | 3(7) | 5(6) | 6(8) | 4(5) | 7(9) | 8(12) |
| | ORA | 6(10) | 2(3) | 2(3) | 1(2) | 1(2) | 4(5) | 1(1) |
| GOMF | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MCDB | CAMERA | 1(1) | 1(1) | 2(3) | 2(2) | 1(2) | 2(2) | 3(3) |
| | ORA | 1(1) | 0(1) | 0(0) | 0(0) | 0(1) | 1(1) | 0(1) |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| Total | CAMERA | 11(17) | 4(8) | 7(9) | 9(11) | 5(8) | 9(12) | 12(16) |
| | ORA | 8(12) | 2(4) | 2(3) | 1(2) | 1(3) | 5(6) | 1(2) |

Tabla 7.24: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *polyamine*.

| Base de Datos | Método | nº de Conjuntos con un FDR <0.05(0.1) | | | | | | |
|---------------|--------|---------------------------------------|---------|----------|----------|----------|----------|--------------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis |
| KEGG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| GOBP | CAMERA | 0(0) | 1(1) | 1(1) | 1(1) | 2(2) | 2(2) | 2(2) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| GOMF | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 1(1) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MCDB | CAMERA | 1(1) | 1(1) | 1(1) | 1(1) | 1(1) | 1(1) | 1(1) |
| | ORA | 1(1) | 1(1) | 0(0) | 0(0) | 0(0) | 0(1) | 0(1) |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| Total | CAMERA | 1(1) | 2(2) | 2(2) | 2(2) | 3(3) | 4(4) | 3(3) |
| | ORA | 1(1) | 1(1) | 0(0) | 0(0) | 0(0) | 0(1) | 0(1) |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Onncogenic; MIM: Msigdb Immunology.

Como resultado del metaanálisis de enriquecimiento, de los 9115/3588 conjuntos de genes que rechazan las hipótesis autocontenida/competitiva en el metaanálisis, se identificaron 578/214 conjuntos pertenecientes a MCC y 106/37 pertenecientes MOG que rechazaron las hipótesis nulas *autocontenida/competitiva*, 50/3 que contienen la palabra *calcium*, 296/147 que contienen *cancer*, 26/5 *wnt*, 56/9 *mitochondria*, 12/1 *reticulum* y 3/0 *polyamine*. En consecuencia, aparte de lo cabría esperar en cuanto a la identificación de conjuntos de genes involucrados en la señalización *Wnt* en la síntesis de *poliaminas* o en el cáncer, también se han encontrado fuertes evidencias de la existencia de un remodelado del Ca^{2+} intracelular en el CCR, así como de alteraciones a nivel mitocondrial y del RE, lo que apoya la hipótesis planteada en la presente tesis. No obstante, como se refleja en las tablas 7.18-7.24, el número de conjuntos de interés es elevado. Por ello, para su consulta, en el apéndice “B.4.2. Metaanálisis de enriquecimiento de genes” y en las tablas B.64-B.74 pueden consultarse, dividido en función de la base de datos a la que pertenecen, los resultados del análisis de enriquecimiento de aquellos conjuntos con un $\text{FDR} < 0.1$ cuyos nombres contienen alguna de las palabras clave.

Parte V

Discusión

“Es un error capital el teorizar antes de poseer datos. Insensiblemente uno comienza a alterar los hechos para encajarlos en las teorías, en lugar de encajar las teorías en los hechos”.

Ser Arthur Conan Doyle

Capítulo 8

Discusión

Actualmente, en la literatura científica existe una gran evidencia de la implicación de los sistemas de transporte del calcio intracelular en el cáncer. Por ejemplo, realizando una simple búsqueda en PubMed de *“Calcium AND Cancer”* se pueden encontrar artículos ya en el año 1918, como el de Cramer (1918) en el que estudió el efecto inhibitorio transitorio del calcio en el crecimiento tumoral de células de carcinoma de ratón, o el de Gill (1935) donde una carta al editor de este autor advierte, ya en el año 1935, de la evidente relación entre las deficiencias de calcio y las neoplasias en casos de hipoparatiroidismo y del efecto de diferentes compuestos de magnesio y calcio en el tratamiento antitumoral, y propone literalmente: *“Para un químico bastante alejado de su campo, surge la idea de usar proporciones definidas de sales de calcio y magnesio en la terapia del cáncer en lugar de solo magnesio.”*

En la actualidad, la comunidad científica ha aportado numerosas evidencias que sugieren que los sistemas de transporte de calcio intracelular están implicados en diferentes tipos de cáncer. No obstante, la forma en la que estos sistemas están alterados y sus efectos en la fisiopatología del cáncer pueden llegar a ser muy diferentes e incluso contrarios entre los diferentes tipos de cáncer. Ejemplo de ello es el efecto protector frente a la apoptosis que tiene la inhibición de SOCE en líneas celulares de CCR mediante la inhibición farmacológica de ORAI1 o STIM1 (Sobradillo et al. (2014)), mientras que el silenciamiento de ORAI1 en células B de linfoma No Hodgking mostró el efecto contrario; es decir, el aumento de la muerte celular (Vacher et al. (2015)). Por todo ello, nuestro grupo ha dedicado gran parte de su trabajo al estudio del remodelado de la homeostasis del calcio intracelular en diferentes tipos de cáncer, entre ellos el de colon. Sin embargo, el elevado número de sistemas de transporte y genes reguladores demanda un nuevo abordaje experimental que requiere el uso de nuevas herramientas de análisis masivo, tratamiento estadístico de los datos y modelos cercanos al cáncer colorrectal humano.

En este trabajo de investigación que se presenta como Tesis Doctoral, se ha estudiado, en primer lugar, el remodelado de los sistemas de transporte de calcio intracelular, tanto a nivel funcional como transcriptómico, utilizando para ello modelos celulares de colon humano normal y tumoral. En concreto las células NCM460 que representan células normales de colon humano,

y las células HT29, un modelo de cáncer de colon humano bien conocido. En segundo lugar, puesto que se ha propuesto que la ruta oncogénica de Wnt/c-Myc y la síntesis de poliaminas podría inducir este remodelado en CCR, se ha estudiado el efecto de la inhibición de la síntesis de poliaminas para investigar si este produce una reversión del mismo. Para ello, se han utilizado las mismas células humanas tanto tumorales como normales de colon. En concreto, en estos modelos celulares se ha realizado diferentes experimentos de análisis transcriptómico mediante técnicas de alto rendimiento basadas en las tecnologías tanto de microarrays como de *RNA-seq* y análisis del efecto de la inhibición de la síntesis de poliaminas por DFMO en ambas líneas celulares a nivel tanto funcional como transcriptómico.

En tercer lugar, se han empleado muestras pareadas de tejido humano de CCR y normal, en las que se ha analizado su transcriptoma mediante microarrays junto con la expresión de algunas proteínas claves ensayada mediante *Western-blot*: ORAI1, ORAI2, STIM1, STIM2 y TRPC1. Finalmente, en cuarto lugar, se han empleado datos transcriptómicos crudos (sin procesar) de tejido de colon humano normal y tumoral tanto extraídos de repositorios públicos como de las muestras pareadas anteriores, con el objeto de, por vez primera, realizar un metaanálisis global del remodelado del calcio intracelular en CCR. En concreto, se reanalizaron cada uno de los experimentos extraídos y se juntaron los resultados alcanzados con los obtenidos de las muestras de tejido humano pareadas de CRC procesadas por nuestro grupo. Así, a partir de estos resultados se obtuvo un efecto global, más preciso y fiable tanto de la expresión diferencial para cada uno de los genes expresados en estos tejidos como de los conjuntos de genes, como rutas de señalización, alterados y que, por tanto, podrían guardar relación con la fisiopatología del CCR humano.

Para facilitar la discusión de los resultados, el presente apartado se divide en cuatro partes. En las dos primeras se discutirán los resultados obtenidos a partir de los modelos celulares en presencia y ausencia de poliaminas endógenas (células control y células tratadas con DFMO, respectivamente). En la tercera parte se discutirán los resultados generados a partir de las muestras pareadas de tejido normal y tumoral de colon humano. Finalmente, en la cuarta y última parte, se discutirán los resultados del metaanálisis.

8.1. Remodelado del calcio intracelular en cáncer colorrectal en células modelo de colon humano normal (NCM460) y tumoral (HT29).

En primer lugar, con el fin de comprobar que los modelos celulares empleados a lo largo de la tesis se comportan según lo esperado, en las líneas celulares HT29 y NCM460 se estudió, mediante imagen de calcio intracelular, la concentración de Ca^{2+} citosólica en condiciones basales y el tamaño de los depósitos de calcio intracelular mediante la perfusión de CPA en medio libre de calcio para, a continuación, inducir la activación de SOCE mediante la perfusión con medio con calcio 1 mM. A partir de este experimento se obtuvo que la línea celular HT29 presenta una

mayor SOCE y un menor tamaño de los depósitos en comparación con las células NCM460. Estos resultados son similares a los descritos previamente por nuestro grupo (Sobradillo et al. (2014)). Los datos son también consistentes con trabajos previos en muestras de pacientes de CCR en las que existen mutaciones de ORAI1 con ganancia de función (Frischauf et al. (2017)). Los resultados son también consistentes con resultados previos obtenidos en otros tipos de cáncer incluyendo cáncer de pulmón donde es común encontrar un aumento significativo de la SOCE que correlaciona con mayor proliferación celular y peor pronóstico (Zhan et al. (2015)). Del mismo modo, el menor tamaño de los depósitos de Ca^{2+} observado en las células tumorales es similar al descrito previamente por nuestro grupo (Sobradillo et al. (2014)). En este trabajo, los datos de *Western blot* revelaron que la mayor entrada capacitativa observada en las células tumorales podría ser debida a la mayor expresión de proteínas de la familia ORAI, incluyendo ORAI1, ORAI2 y ORAI3, la mayor expresión del canal TRPC1 y la mayor expresión del sensor de Ca^{2+} STIM1, acompañada de la pérdida de expresión del sensor STIM2. Esta caída de los niveles de STIM2 descrita previamente también podría explicar la disminución del Ca^{2+} contenido en los depósitos ya que este vendría determinado en exclusiva por STIM1, un sensor de mayor afinidad por Ca^{2+} que STIM2 y que requiere, por tanto, de un vaciamiento extensivo del Ca^{2+} de los depósitos para su activación (Sobradillo et al. (2014)).

Sin embargo, otras muchas proteínas regulan SOCE y, además, otros muchos canales, bombas y transportadores, tanto en la membrana plasmática como en el retículo endoplásmico y las mitocondrias, están implicados en la homeostasis del calcio intracelular hasta llegar a un total en torno al centenar de proteínas. Por ello, para estudiar en toda su extensión el posible remodelado de la homeostasis del calcio intracelular en cáncer de colon, se hace necesario investigar la totalidad de genes implicados en el transporte de calcio intracelular.

Por ello, tras “validar” las líneas celulares HT29 y NCM460 usadas en el presente trabajo, se procedió a llevar a cabo un análisis transcriptómico, especialmente de los genes implicados en el transporte y la homeostasis del Ca^{2+} intracelular. Con este fin se emplearon tres tecnologías transcriptómicas de alto rendimiento diferentes, dos de ellas basadas en *RNA-seq* -IonTorrent e Illumina- y la restante basada en microarrays -Clariom D Human de Affymetrix-. Cabe destacar que el transcriptoma analizado mediante los microarrays Clariom D Human proviene del mismo cultivo empleado en los experimentos de imagen anteriores. A partir de los datos transcriptómicos, se realizó el análisis de expresión diferencial y de enriquecimiento de conjuntos de genes con el fin de encontrar cuáles de estos conjuntos podrían estar relacionados con el fenotipo tumoral; es decir, con las diferencias entre HT29 y NCM460. Así, de los cerca de 30.000 conjuntos de genes analizados se obtuvo que, con Illumina, el número de conjuntos de genes que muestran tal relación (que rechazan las hipótesis nulas autocontenida/competitiva) es muy bajo, 13/87. Aun así, se identifican conjuntos relacionados con el CCR y con la señalización de Wnt. En cuanto a los datos obtenidos mediante el experimento de microarrays, se obtiene un número más elevado de conjuntos relacionados con el genotipo de CCR. En concreto, 967/1823 conjuntos de genes rechazaron la hipótesis nula autocontenida/competitiva y, en consecuencia, estarían relacionados con el fenotipo de CCR. De estos, 89/62 y 4/6 pertenecen a las bases de datos que contienen

conjuntos de genes relacionados con cáncer MCC y MOG, respectivamente. Además, del total de conjuntos relacionados, 28/46 contienen la palabra *cancer*, 1/5 *wnt*, 6/6 *mitochondria* y 1/0 *reticulum*. Por tanto, como era de esperar, los datos indican que hay una evidente alteración de los conjuntos de genes relacionados con el cáncer. También se observa una alteración de dos orgánulos clave en la homeostasis del calcio intracelular: el RE y la mitocondria.

Para el análisis transcriptómico en líneas celulares modelo se han utilizado tres metodologías diferentes, en tres periodos diferentes. Inicialmente se utilizó la tecnología Ion Torrent basada en *RNA-seq* que, aunque reportó resultados muy importantes, los datos estaban limitados a un panel de cerca de 80 genes. Quisimos comprobar si la tecnología de Illumina, también basada en *RNA-seq* pero que abarca, esencialmente, el transcriptoma completo, podría reportar los mismos resultados en las líneas celulares. Los resultados muestran que los genes diferencialmente expresados por las dos metodologías son muy similares aunque no idénticos encontrándose algún gen en cada familia de genes que aparece diferencialmente expresado por una tecnología y no por la otra y/o viceversa. Ion Torrent detectó hasta 35 genes diferencialmente expresados en el panel, mientras que fueron 38 los genes diferencialmente expresados en el mismo set de genes según Illumina. Además, quisimos testar la tecnología de microarrays por varios motivos. En primer lugar el precio, ya que la metodología de microarrays es sensiblemente más barata que las basadas en *RNA-seq* aunque aporta esencialmente la misma información (salvo la secuencia de los transcritos). Además, y este es un motivo esencial, la mayor parte de los estudios publicados de análisis transcriptómico diferencial en cáncer de colon depositados en repositorios públicos utilizan esta metodología lo que la hace necesaria si queremos comparar el remodelado del transcriptoma de las líneas celulares con las muestras pareadas y con el metaanálisis de los datos procedentes de los repositorios. El método de microarrays arrojó un resultado de 32 genes diferencialmente expresados para el mismo panel. Por tanto, este método es el más restrictivo de los tres y el que menos genes diferencialmente expresados detecta aunque las diferencias globales son muy pequeñas, de solo entre 5 y 10 % respecto de los resultados de las tecnologías basadas en *RNA-seq*.

Teniendo en cuenta las observaciones anteriores, se ha considerado que un gen está diferencialmente expresado (DEG) en las siguientes circunstancias: i) cuando el gen muestra el mismo patrón de expresión diferencial con, al menos, dos de las tres tecnologías empleadas, ii) cuando el gen está diferencialmente expresado en una de ellas pero que el logFC en las demás tenga el mismo signo, iii) que lo esté únicamente en los microarrays pues es el método más restrictivo y los resultados posteriores en muestras de tejido se han realizado en estos mismos microarrays. Además, en caso de que en los microarrays se observe un DEG con un $FDR < 0.1$ y se observe de igual modo con alguna de las otras dos tecnologías, también se considerará como diferencialmente expresado. No obstante, en aquellos casos en los que se observen contradicciones habrá que tener cautela, si bien es cierto que, con el fin de poder comparar con los resultados posteriores en tejidos humanos y en los experimentos con DFMO, que han sido realizados con microarrays, primará lo observado en los microarrays. En la siguiente figura 8.1 se muestra un resumen del análisis de expresión diferencial entre las células HT29 y NCM460 para los genes relacionados con la homeostasis del calcio intracelular.

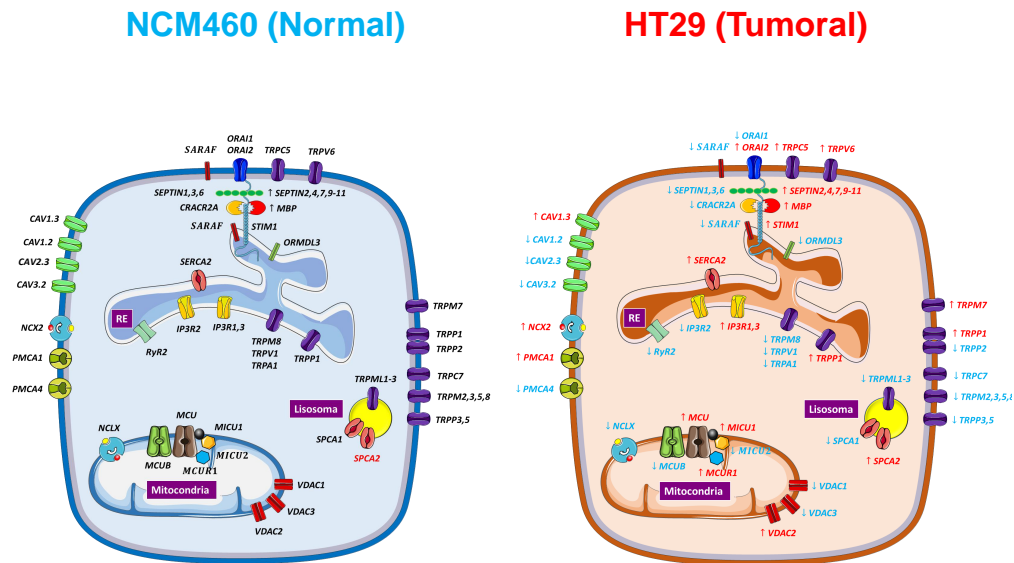


Figura 8.1: Expresión diferencial de genes relacionados con la homeostasis del Ca^{2+} intracelular entre las células HT29 y NCM460. En la figura se muestra la expresión diferencial en la comparación entre las células HT29 (derecha) y NCM460 (izquierda) como modelos de células de cáncer de colon y normales, respectivamente. Sobre la figura correspondiente a HT29 se muestran únicamente los genes cuya expresión se ve alterada respecto de las NCM460, donde el color rojo indica sobreexpresión y el azul infraexpresión.

8.1.1. Remodelado de los canales de calcio dependientes de voltaje

Respecto a los canales de Ca^{2+} dependientes de voltaje (VOCC), los datos sugieren un aumento en la expresión de *CAVI.3* y la disminución de la expresión de *CAV2.2* y *CAV3.2* en las células tumorales. Hay que aclarar que los datos transcriptómicos muestran diferencias en los niveles de ARNm o genes transcritos que no tienen porqué estar asociados necesariamente con diferencias similares en la expresión a nivel de proteínas. Este puede ser el caso de este tipo de canales que no parecen estar operando funcionalmente en las células. Sabemos que la despolarización con medio que contiene un alta concentración de K^+ activa la entrada de Ca^{2+} y el incremento de la $[\text{Ca}^{2+}]_{\text{cit}}$ en células excitables. Sin embargo, sabemos que la despolarización no induce incrementos de $[\text{Ca}^{2+}]_{\text{cit}}$ en estas células (datos no mostrados), lo que sugiere que los VOCC no son funcionales y, por tanto, diferencias en la expresión de los genes que codifican para estos VOCC no deberían contribuir a las diferencias en la homeostasis del Ca^{2+} intracelular entre las células HT29 y las NCM460. Esto no significa que los cambios no puedan contribuir de un modo no canónico a la tumorigénesis. De hecho, los resultados obtenidos son consistentes con la sobreexpresión de canales tipo *CAVI.3* descrita previamente en cáncer de colon y que podría contribuir a la adquisición de capacidad para migración celular a través de un mecanismo no canónico (Fourbon et al. (2017)). Por ejemplo, la escisión de su extremo C-terminal podría participar en el proceso tumorigénico mediante la modulación de la actividad de canales de K^+ y regular procesos tumorales, como por ejemplo la

proliferación, mitogénesis, diferenciación, apoptosis y metástasis (X.-T. Wang et al. (2000), Phan et al. (2017)).

Es interesante resaltar que, de las tres subfamilias de VOCC, CAV1 es la que muestra una mejor señalización hacia el núcleo y es capaz de inducir la expresión de diferentes genes regulados por CREB, los cuales están estrechamente relacionados con la fisiopatología del cáncer (Sakamoto and Frank (2009)), mientras que CAV2 no es capaz de activar dicha señalización. Por tanto, la sobreexpresión de *CAV1.3* e infraexpresión *CAV2.2* podría favorecer el fenotipo tumoral de las células HT29.

8.1.2. Remodelado de la entrada capacitativa de calcio (SOCE) y los canales TRPs

Respecto a SOCE, en este y otros trabajos previos se ha demostrado que las células HT29 muestran una SOCE mucho mayor que las NCM460. Además, las corrientes que subyacen son mayores y más complejas e incluyen posiblemente no solo canales de la familia ORAI, especialmente ORAI1, sino también otros canales como TRPC1. Los resultados del análisis transcriptómico revelan el aumento de la expresión de *STIM1* y *ORAI2* que podrían generar corrientes tipo Icrac o Icrac like. No obstante, aunque la proteína ORAI2 se sobreexpresa en las células HT29 respecto a las NCM460, se ha descrito que su silenciamiento no inhibe las corrientes inducidas por el vaciamiento de los depósitos (Sobradillo et al. (2014)). Sin embargo, la sobreexpresión de la proteína STIM1 sí podría ser determinante en el aumento de SOCE en las células tumorales. Además, el aumento de la expresión de *TRPC5* podría contribuir a generar complejos de canales con TRPC1 activados directamente por STIM1 o por la entrada de Ca^{2+} inducida por ORAI1, que podrían generar corrientes tipo Isoc y contribuir a una mayor SOCE. Por otra parte, los datos revelan cambios en la expresión de genes que codifican para proteínas reguladoras de SOCE que también podrían contribuir al remodelado de SOCE en las células tumorales. Por ejemplo, la sobreexpresión de *TRPM7*, un regulador positivo de ORAI (Kádár et al. (2021)) y de las septinas *SEPTIN2,4,8,9,10,11* podría favorecer SOCE en las células tumorales. Además, los genes que codifican para algunos de los reguladores negativos de SOCE, como *SARAF* y *ORMDL3*, se expresan menos en las células tumorales lo que contribuiría a una mayor SOCE en dichas células. No obstante, la menor expresión de *ORAI1*, el canal canónico de la entrada capacitativa, y de los genes *CRACR2A* y las septinas *SEPTIN1,3,6* que regulan positivamente SOCE, a la par del aumento de los reguladores negativos *SEPTIN7* y *MBP* podrían tener efectos opuestos y reducir SOCE. En cualquier caso, la disminución de la expresión de *ORAI1* podría ser simplemente el reflejo especular debido a un efecto compensatorio de la mayor expresión de ORAI1 a nivel de proteína descrita previamente en estas células (Sobradillo et al. (2014)).

En cuanto a la entrada de calcio desde el medio extracelular independiente de depósitos, los datos sugieren que la entrada a través de los canales ARC podría ser similar e, incluso, estar favorecida en las HT29, pues hay un aumento de *STIM1* y una disminución de *SARAF*, mientras

que *ORAI3* no cambia, si bien es cierto que *ORAI1* está disminuido. De hecho, se ha descrito que el ácido araquidónico induce la proliferación celular a través de una entrada que aumenta el calcio intracelular, lo que podría explicar la mayor concentración de calcio basal encontrada en las HT29 (Alessandra Fiorio Pla et al. (2008)). Otro posible mecanismo de entrada de calcio independiente tanto de los depósitos como de STIM es a través de la interacción entre *ORAI1* y *SPCA2*, pero, al igual que en los casos anteriores, mientras que *SPCA2* está sobreexpresado, *ORAI1* está infraexpresado, por lo que tampoco puede inferirse el comportamiento únicamente con estos datos.

Mención especial merece la sobreexpresión de *TRPV6* en las células tumorales. TRPV6 no es un canal operado por depósitos aunque es muy selectivo para Ca^{2+} . En realidad TRPV6 es un canal de Ca^{2+} activado por Ca^{2+} cuyo papel fisiológico es el transporte de calcio en el tracto gastrointestinal desde la luz del intestino hacia los vasos. Sus niveles de expresión están regulados por el receptor de vitamina D que a su vez depende de los niveles de calcio plasmático. La sobreexpresión de *TRPV6* podría contribuir a la mayor permeabilidad a Ca^{2+} de las células tumorales. De hecho, la sobreexpresión de *TRPV6* ha sido identificada en el CCR desde estadios tempranos de la enfermedad, pues su silenciamiento ha mostrado tanto la inhibición de la proliferación como la inducción de la apoptosis en células de CCR. En este sentido, la supresión de TRPV6 debida a una dieta rica en calcio ejerce un efecto protector frente al CCR.

La mayoría del resto de canales TRP de la membrana plasmática incluyendo TRPC7, TRPM2,8, TRPML1-3, TRPA1, TRPP2,3,5 están, en realidad, regulados negativamente en las células tumorales. La reducción de los canales no selectivos de calcio TRPM2, TRPA1, TRPML1 podría prevenir la muerte celular por apoptosis debida al estrés oxidativo, pues estos se activan por ROS (Hayes, Dinkova-Kostova, and Tew (2020)).

Al contrario que el TRPV6, la disminución del TRPP5 podría conferir resistencia a la muerte, pues su sobreexpresión en células Hela ha mostrado la inhibición de la proliferación y la inducción de la apoptosis (Xiao et al. (2010)). De igual modo, el TRPV1 se considera un supresor tumoral en el CRC (Rizopoulos and Assimakopoulou (2021)), por lo que su disminución podría favorecer la progresión tumoral, pues este es un regulador negativo importante de la expresión de EGFR (Mesquita et al. (2021)). En este sentido, se ha considerado el uso de la capsazepine como estrategia terapéutica para el CRC. De hecho, se ha observado que este compuesto puede inducir la apoptosis en células CRC (Rizopoulos and Assimakopoulou (2021)).

Otro canal de la superfamilia de los TRP que podría favorecer la metástasis es el TRPA1, pues se ha relacionado su disminución con un aumento de la capacidad migratoria de las células cancerígenas (Cojocarú et al. (2021)) y con un peor pronóstico del CRC (datos disponibles en <https://www.proteinatlas.org/ENSG00000104321-TRPA1/pathology>, v20.1proteinatlas.org).

8.1.3. Remodelado de los Sistemas de Extrusión de Calcio

Los datos funcionales indican que las células tumorales muestran unos niveles de calcio intracelular basal significativamente mayores que las células normales. Estos datos son similares a los descritos previamente y podrían explicar el mayor potencial proliferativo de las células tumorales. De hecho, se sabe que el simple aumento del nivel de calcio intracelular basal al aumentar el calcio extracelular incrementa la tasa de proliferación en estas células (Valero et al. (2008)). Los mayores niveles de calcio intracelular pueden ser debidos a una mayor permeabilidad basal a calcio por expresión de canales activos en situación de reposo, y ese podría ser el caso de los canales TRPV6 que se sobreexpresan en estas células, como se ha comentado arriba. También, la diferencia podría ser debida a una menor extrusión de calcio desde el citosol al exterior celular o a los depósitos intracelulares de calcio a través de bombas y/o transportadores. Los datos transcriptómicos muestran que la expresión de PMCA4 es significativamente menor en células tumorales. Estos datos son consistentes con resultados previos donde se demuestra que la expresión de esta bomba disminuye en cáncer de colon lo que se correlaciona con el aumento del nivel basal del calcio intracelular y la tasa proliferativa (Aung et al. (2009)). Sin embargo, los datos muestran también sobreexpresión de PMCA1 en las células tumorales, lo que sugiere que en cáncer de colon podría haber un intercambio de bombas de PMCA4 a PMCA1 en la transformación tumoral.

Además, se observa aumento de la expresión de NCX2 y, en menor medida, de NCX3. Se ha descrito que NCX1 podría unirse a TRPC1 y TRPC6 en cáncer de páncreas y modular su actividad canales (J. Xu et al. (2018)). Quizás la sobre expresión de NCX2 podría contribuir al remodelado de la homeostasis del Ca^{2+} intracelular a través de interacciones con TRPC5 que está también sobreexpresado en células tumorales u otros canales de esta superfamilia.

A nivel de depósitos intracelulares de Ca^{2+} , se observan aumentos significativos de *SERCA2* y *SPCA2*, y una disminución de *SPCA1*. Consistentemente, en muestras de tejido de pacientes de CCR se ha observado un aumento de *SERCA2* (Chung et al. (2006)). Además, en líneas celulares de cáncer de colon se ha visto que la sobreexpresión de esta isoforma aumenta la tasa proliferativa y su capacidad para metastatizar (Fan et al. (2014)). Los datos sugieren un intercambio de *SPCA1* por la *SPCA2* en cáncer de colon podría regular el ciclo celular en situaciones de hipoxia (Jenkins et al., (2016)). Consistentemente, M. M. Feng et al. (2010) encontraron que *SPCA2* está sobreexpresada en cáncer de mama, próstata y colon (M. M. Feng et al. (2010), Dang and Rao (2016)). De hecho, encontraron que su supresión disminuye la $[\text{Ca}^{2+}]_{\text{cit}}$ y la tumorigenicidad. Como resultado, la formación del complejo ORAI1/*SPCA2* provoca una señalización constitutiva de Ca^{2+} independiente de los depósitos que promueve la tumorigénesis. Por tanto, la sobreexpresión de *SPCA2* en cáncer de colon podría contribuir a la formación de canales constitutivos hechos de ORAI1/*SPCA2*, mayores niveles de Ca^{2+} intracelular y tumorigénesis. En este punto se desconoce si los cambios en la expresión de *SERCA2* podrían contribuir al relativo vaciamiento de los depósitos intracelulares de Ca^{2+} observado en la células tumorales respecto a las normales.

8.1.4. Remodelado de la liberación de Ca^{2+} de los depósitos

La diferencia relativa en el tamaño de los depósitos podría deberse más bien a cambios en la expresión de canales de Ca^{2+} operados por ligando en el RE de las familias IP3R y RyR. Los resultados muestran que en las células tumorales hay un intercambio de *IP3R2* (que disminuye) por *IP3R1* y *3* (que aumentan). Además, se observa una disminución de *RyR2*. Sakakura et al. (2003) ya observó el aumento de *IP3R3* en muestras de tejido humano y líneas celulares de CCR, al igual que Shibao et al. (2010). También se ha observado que una mayor expresión de esta isoforma podría estar relacionada con una mayor agresividad. Cabe destacar que durante la EMT, la expresión tanto de las isoformas *IP3R1* como de la *3* aumenta en líneas celulares de cáncer de mama (Davis et al. (2013)). Si estos canales tuviesen actividad basal, entonces una sobreexpresión de ellos podría contribuir a explicar el relativo vaciamiento de los depósitos intracelulares de Ca^{2+} en las células tumorales y su resistencia a la apoptosis. Este efecto podría también ser favorecido por el aumento del ratio STIM1/STIM2, sensores de los niveles de Ca^{2+} del RE que impedirían la activación de SOCE hasta que se produce un vaciamiento muy grande (Sobradillo et al. (2014)). Finalmente, la sobreexpresión de *TRPP1*, un canal que se localiza también en la membrana del RE, podría contribuir al vaciamiento parcial de los depósitos de las células tumorales.

La disminución de la expresión de *RyR2*, junto con la pérdida de canales TRP asociados al retículo como *TRPM8*, *TRPA1* y *TRPV1* puede tener que ver la desdiferenciación de las células que prescinde de sensores celulares de temperatura y estrés oxidativo y cuya activación podría generar el fenómeno de liberación de Ca^{2+} inducida por Ca^{2+} (CICR) que quedaría fuertemente afectado cuando no prácticamente desmantelado en las células tumorales. Además, la localización celular aberrante de *TRPM8* se ha relacionado con la progresión tumoral (Monteith et al. (2007), Bidaux et al. (2007)).

Finalmente, hay que indicar que el remodelado del nivel de llenado de los depósitos de calcio influye decisivamente en la posibilidad de la activación de la entrada capacitativa simplemente al acercarse o no el nivel de llenado de los depósitos al umbral de activación de este mecanismo.

8.1.5. Remodelado del transporte de Ca^{2+} en la mitocondria

Finalmente, los datos muestran cambios significativos de genes implicados en el transporte de Ca^{2+} mitocondrial. Se observa el aumento significativo del poro del complejo uniportador de calcio mitocondrial (MCU) y su regulador *MICU1* junto con *VDAC2*, el canal aniónico de la membrana externa mitocondrial que es sobreexpresado a la par de la infraexpresión de *VDAC1* y *VDAC3* junto con la inhibición de *MICU2* y *MCUb*, dos reguladores negativos de MCU.

Como se comentó en la introducción, *MICU1* es un modulador de MCU que impide la captación de Ca^{2+} mitocondrial a bajas concentraciones de Ca^{2+} citosólico pero a altas concentraciones se disocia para permitir la captación. Su actividad está regulada por *MICU2*, que inhibe la captación de Ca^{2+} mediante la unión a *MICU1* a través de un puente disulfuro. Además, su perfil de expresión

depende de la expresión de MICU1. Cabe destacar que la expresión de MICU2 es necesaria para el funcionamiento de MCU para poder inhibir la captación de Ca^{2+} a grandes concentraciones. Por otro lado, MCUB es un regulador negativo de MCU, por lo que su menor expresión favorecería la captación de Ca^{2+} mitocondrial por MCU. Por tanto, los cambios encontrados sugieren una mayor capacidad de captación de Ca^{2+} mitocondrial por parte de las mitocondrias en las células tumorales.

Al aumento de esta capacidad de Ca^{2+} mitocondrial le ha de seguir su salida de vuelta al citosol a través del intercambiador NCLX que, paradójicamente, está infraexpresado en las células tumorales. No obstante, se ha descrito que la disminución del NCLX podría jugar un papel clave en el aumento de la capacidad metastásica de las células de cáncer de colon. En concreto, en términos mecanísticos, el aumento de calcio mitocondrial conlleva el aumento de ROS que, a su vez, activa la señalización por HIF1 α la cual favorece la metástasis de aquellas células con baja expresión de NCLX (Pathak et al. (2020)). La infraexpresión del NCLX podría jugar también un papel clave en la quimiorresistencia y expresión de rutas de señalización epitelio-mesenquimal, hipoxia y células madre. Esta baja expresión se ha asociado con mutaciones en TP53 y tumores BRAF wild type. Cabe destacar que se ha observado una reducción de NCLX desde el estadio I al IV. Debido a esta expresión paradójica, se hace relevante determinar cuáles de los cambios observados se reflejan a nivel de proteínas, pues podría no coincidir con lo encontrado a nivel transcripcional.

Finalmente, hay que indicar que el aumento de la captación de calcio mitocondrial regula tanto la liberación de calcio (al favorecer la activación/inactivación de los receptores de IP3) como la entrada capacitativa de calcio, al prevenir o no la inactivación dependiente de calcio de los canales ORAI1.

En resumen, la expresión diferencial de 56 genes implicados en el transporte de calcio en las células normales y tumorales de colon humano puede contribuir a explicar la mayor entrada de calcio a través de la entrada capacitativa (sobreexpresión de *STIM1* y mayor captación de Ca^{2+} mitocondrial), el mayor nivel de calcio intracelular basal (infraexpresión de *PMCA4* y sobreexpresión de *TRPV6* y *SPCA2*), menor contenido de Ca^{2+} de los depósitos (aumento del cociente *STIM1/STIM2* y sobreexpresión de *IP3R1,3*) y mayor captación de Ca^{2+} mitocondrial (sobreexpresión de *MCU* y cambios en proteínas moduladoras).

8.2. Efecto de la depleción de poliaminas sobre el remodelado del Calcio Intracelular en Cáncer Colorrectal observado en células modelo de colon humano normal y tumoral

Como se describe en detalle en la Introducción, gran parte de los tumores están asociados a mutaciones en APC que originan la activación de la ruta de señalización Wnt y de la expresión del oncogén *c-MYC*. Como resultado, tiene lugar la sobreexpresión de ODC, el enzima limitante en la biosíntesis de poliaminas. Se ha propuesto que las poliaminas podrían contribuir al remodelado de la homeostasis del calcio intracelular y a la aparición del fenotipo tumoral por lo que, en este

trabajo, nos propusimos estudiar el efecto de la depleción de poliaminas sobre la homeostasis del calcio intracelular, tanto a nivel funcional como a nivel transcriptómico. Para ello, tratamos los dos modelos celulares con DFMO, un inhibidor suicida de ODC. Hay que destacar que utilizamos el mismo cultivo celular para los experimentos de imagen de calcio que para los estudios de transcriptómica.

Los resultados muestran que el tratamiento con DFMO revierte parcialmente el remodelado de la homeostasis del calcio intracelular a varios niveles que tiene lugar en el cáncer de colon. En primer lugar, disminuye los niveles basales de calcio intracelular. Asimismo, el DFMO disminuye la elevada entrada capacitativa característica de las células tumorales y, finalmente, aumenta el tamaño de los depósitos de calcio intracelular que está disminuido en las células tumorales. Este inhibidor también produce efectos en las células normales aunque de menor magnitud. Así, DFMO también disminuye los niveles basales de Ca^{2+} intracelular en las células normales así como el Ca^{2+} de los depósitos, este último contrario al efecto observado en células tumorales.

Aunque estos resultados son interesantes, los datos más relevantes fueron obtenidos en el análisis transcriptómico. En primer lugar, destaca el hecho de que el DFMO ha mostrado un efecto selectivo sobre las células HT29 en relación con las NCM460. En concreto, mientras que en las células normales únicamente se vieron afectados por el DFMO 62/165 transcritos (0.48/1.27 % de los que se expresan) con un para un $\text{FDR} < 0.05/0.1$, en las células tumorales el número de transcritos diferencialmente expresado fue del orden de 50 veces superior, 3394/4402 (26.16/33.93 %) para los mismos FDR. Esta enorme diferencia en cuanto a la sensibilidad de unas células y otras a la depleción de poliaminas podría deberse a que las células tumorales sobreexpresan ODC y, por tanto, crecen en un ambiente enriquecido en poliaminas por lo que serían muy sensibles a su depleción. Sin embargo, las células normales, donde la ODC no debería estar activada o inducirse de forma muy transitoria, el ambiente en poliaminas es pobre y su depleción no tendría mayores influencias a nivel transcriptómico. En condiciones normales, la ODC se activa de manera transitoria durante los procesos de restitución epitelial donde promueve el cambio de un fenotipo diferenciado a otro con capacidad proliferativa y migratoria esencial para la restitución o sellado epitelial. Es interesante que la ODC tiene la vida más corta conocida para un enzima, de solo 5 min, y, además, esta se degrada de un modo independiente a su ubiquitinación lo que limita de modo notable la síntesis de poliaminas en células normales.

En células normales, el único gen del panel de genes implicados en transporte y homeostasis de Ca^{2+} intracelular afectado por la depleción de poliaminas fue *IP3R1*, con un $\text{FDR} = 0.098$, que codifica para el receptor de IP3 tipo 1. Se conoce que el receptor IP3R1 regula la transferencia de Ca^{2+} directa desde el RE a la mitocondria. Por tanto, su sobreexpresión podría facilitar una mayor transferencia, en condiciones basales, de Ca^{2+} desde el RE a la mitocondria, lo que podría afectar tanto al nivel de llenado de los depósitos de Ca^{2+} como a la concentración basal de Ca^{2+} intracelular. Sin embargo, en las células tumorales, DFMO no induce efecto alguno sobre los canales de liberación de Ca^{2+} de los depósitos, por lo que el efecto de DFMO sobre el relleno de los depósitos en células tumorales debe explicarse de otra manera. En cuanto a la disminución observada en los niveles basales de Ca^{2+} en las células NCM460 tratadas con DFMO, puesto que

ningún gen del transporte de calcio ha mostrado expresión diferencial, dicha reducción podría deberse al efecto que tienen las poliaminas sobre ciertos canales como el TRPM4 y TRPM7, pues a pH fisiológico los grupos amino se unen a ciertas proteínas, como los receptores de glutamato, o los TRPM4 y TRPM7, regulando negativa y positivamente su actividad, respectivamente. Por ello, la disminución en la síntesis de poliaminas en las NCM460 podría estar evitando esta regulación positiva de TRPM7 y, en consecuencia, disminuyendo los niveles basales de calcio intracelular (Lange and Koomoa (2014)). Además, puesto que TRPM4 es un regulador negativo de SOCE (Kappel, Borgström, et al. (2019)), la disminución de poliaminas podría favorecer la actividad de TRPM4 y este, a su vez, regular negativamente.

El tratamiento con DFMO en células tumorales no provoca cambio alguno en la expresión de canales de calcio dependientes de voltaje, pero si aumenta la expresión de *STIM1* y *STIM2* así como otros genes que modulan SOCE incluyendo *CRACR2A*, *ORMDL3*, *SEPTIN6*, 8 y 9, y el canal *TRPP2*. Sin embargo, inhibe la expresión de los canales *TRPC1* y 5, *TRPV6* y *TRPP1*. A nivel de bombas, DFMO aumenta la expresión de *PMCA4* e inhibe la de *SPCA2*. Finalmente, a nivel mitocondrial, DFMO aumenta la expresión de *MCU* y *VDAC3*. Por tanto, DFMO modula la expresión de solo un gen del panel de genes implicados en transporte de Ca^{2+} en células normales y de 16 genes en células tumorales. Es de destacar que en 10 de los 16 genes modulados por DFMO en las células tumorales, el efecto de DFMO revierte el remodelado inducido por la transformación celular. Es decir, por ejemplo, la expresión de *PMCA4* disminuye en las células tumorales y DFMO revierte esa disminución. Igualmente, *SPCA2*, que está aumentado en las células tumorales, su expresión disminuye tras el tratamiento con DFMO. Lo mismo ocurre con los moduladores de SOCE *CRACR2A*, *ORMDL3* y *SEPTIN6*, los canales *TRPC5*, *TRPV6*, *TRPP1*, *TRPP2* y *VDAC3*. Por tanto, la depleción de poliaminas revierte en buena medida el remodelado de la homeostasis del Ca^{2+} intracelular en células tumorales, y solo en células tumorales.

La cuestión que se plantea ahora es cómo los cambios transcriptómicos podrían contribuir a los cambios funcionales observados en las células tumorales. En primer lugar, la disminución de los niveles basales de Ca^{2+} intracelular inducida por DFMO podría explicarse fácilmente por la sobreexpresión de *PMCA4* que extruye el Ca^{2+} citosólico al exterior celular. Además, la inhibición del *TRPV6* inducida por DFMO, un canal sobreexpresado en las células tumorales e implicado en el transporte basal de Ca^{2+} intracelular, podría colaborar en la disminución de los niveles basales de Ca^{2+} intracelular tras la depleción de poliaminas. La disminución de *TRPV6* inducida por DFMO podría tener un efecto protector, pues su sobreexpresión comienza desde los inicios de la enfermedad y se correlaciona con la patogenia de los tumores. Además, como se ha dicho anteriormente, puede favorecer la proliferación celular. De hecho, su silenciamiento también es capaz de inhibir la proliferación (Rizopoulos and Assimakopoulou (2021)). Por lo tanto, la disminución de TRPV6 tras el tratamiento con DFMO únicamente en las HT29 apoyaría la evidencia encontrada por Gutiérrez et al. (2019), pues observaron que el tratamiento con DFMO disminuye la proliferación de las células HT29. En este sentido, la supresión de la expresión de *TRPV6* asociada a una dieta rica en calcio ejerce un efecto protector frente al CCR (Rizopoulos and Assimakopoulou (2021)). Por lo tanto, el tratamiento con DFMO podría ejercer un efecto protector

contra el CCR a través de la disminución de la expresión de *TRPV6*.

En segundo lugar, el mayor contenido de Ca^{2+} de los depósitos podría explicarse por el aumento del cociente *STIM2/STIM1*. El sensor de calcio del RE *STIM2*, de menor afinidad por Ca^{2+} , tiende a activar la entrada capacitativa de Ca^{2+} a depleciones muy limitadas lo que favorece el relleno de los depósitos hasta niveles mayores de Ca^{2+} . De hecho, el simple silenciamiento de *STIM2* en las células HT29 conlleva la depleción parcial de los depósitos (Sobradillo et al. (2014)). Adicionalmente, los cambios en el canal *TRPP1*, que puede localizarse en el RE, podría también tener un papel en el relleno de los depósitos. En este sentido, el canal *TRPP1* también se ha relacionado con un mal pronóstico de CRC (Rizopoulos and Assimakopoulou (2021)) y con el aumento de la proliferación a través de la regulación de la expresión de *EGFR*, por lo que el DFMO, que induce la infraexpresión de este canal, también podría ejercer un papel protector a través de esta vía.

Finalmente, la menor entrada capacitativa observada en células tumorales tratadas con DFMO podría ser debida, en primer lugar, al aumento del cociente *STIM2/STIM1*, más característico de las células normales (Sobradillo et al., 2014). Además, la disminución de *TRPC5* y, sobretodo, de *TRPC1* podría ser esencial en esta reversión. Nuestro grupo ha descrito que en las células tumorales de colon SOCE está mediada no solo por canales *ORAI1* sino también por *TRPC1*. Además, puesto que *TRPC1* suele formar heterómeros con canales *TRPC5*, la inhibición de ambos sugieren la pérdida de la participación de estos canales en SOCE debida a la depleción de poliaminas. Finalmente, los cambios en la expresión de genes reguladores de SOCE como *CRACR2A*, un regulador positivo de SOCE, *ORMDL3*, y las septinas *SEPTIN6* y *8* podrían contribuir también al efecto de DFMO sobre SOCE en las células tumorales. Incluso, puesto que *TRPM4* es un regulador negativo de SOCE y que este canal se regula negativamente por poliaminas, la disminución de las poliaminas tras el tratamiento con DFMO podría favorecer la regulación negativa de SOCE por *TRPM4*.

Otro de los efectos del DFMO es el aumento de la susceptibilidad a la muerte celular, normalmente disminuida en las células tumorales. Pues bien, los datos muestran que DFMO aumenta la expresión de *MCU* y de *VDAC3* en las células tumorales. De los dos, el primero está sobreexpresado en las HT29 control respecto de las NCM460 control, mientras que el segundo está disminuido. En cualquier caso, el aumento de ambos podría dar lugar a una mayor entrada de calcio a la mitocondria y, en consecuencia, a una mayor sobrecarga mitocondrial la cual facilitaría la muerte celular. Nuestro grupo describió (Gutiérrez et al. (2019)) que las células HT29 tratadas con DFMO eran más susceptibles a la apoptosis temprana inducida por estrés oxidativo que, se sabe, depende de la sobrecarga de calcio mitocondrial. Por lo tanto, la mayor susceptibilidad a la apoptosis podría explicarse por el aumento de la expresión de *MCU* y de *VDAC3*.

En resumen, la depleción de poliaminas inducida por DFMO induce cambios en la expresión de 16 genes implicados en el transporte y la homeostasis del calcio intracelular. En 10 de ellos los cambios consisten en revertir el efecto de los cambios producidos durante la transformación celular observado en las líneas celulares. Estos cambios permiten explicar fácilmente la disminución de los

niveles basales de calcio y de la entrada capacitativa de calcio así como el aumento de los depósitos de calcio intracelular inducidos por DFMO así como los cambios en el fenotipo celular. En cambio, la depleción de poliaminas solo induce el cambio en la expresión de un único gen implicado en transporte, el IP3R1, lo que permite explicar la disminución de los depósitos de calcio en estas células. Evidentemente, los cambios en la expresión génica no tienen porqué correlacionarse con los cambios similares a nivel de proteínas o en la actividad funcional. Por tanto, es necesario estudiar en detalle los efectos de DFMO sobre la expresión de proteínas y la actividad de estas en condiciones fisiopatológicas para establecer los efectos de la depleción de poliaminas sobre la homeostasis del calcio intracelular.

Finalmente, puesto que la depleción de poliaminas induce los cambios discutidos arriba, se hace evidente que el exceso de poliaminas debido a mutaciones endógenas o adquiridas en APC, c-Myc u ODC, o incluso una dieta y/o microbiota abundante en poliaminas, podría generar los cambios opuestos en el transcriptoma que favorecerían el fenotipo tumoral, incluyendo los siguientes:

- Las poliaminas podrían aumentar la expresión de los canales *TRPV6*, *TRPC1* y *5*, *TRPP1* y la bomba *SPCA2*.
- Las poliaminas podrían inhibir la expresión del canal *TRPP2*, la bomba *PMCA4*, los reguladores de SOCE *STIM1*, *STIM2*, *CRACR2A*, *ORMLD3*, *Septinas 6,8,9* y los transportadores de calcio mitocondrial *MCU* y *VDAC3*.

A continuación, se muestra un resumen del análisis de los resultados de los experimentos tanto de imagen de Ca^{2+} como transcriptómicos realizados sobre las líneas celulares tratadas y sin tratar con DFMO (figura 8.2).

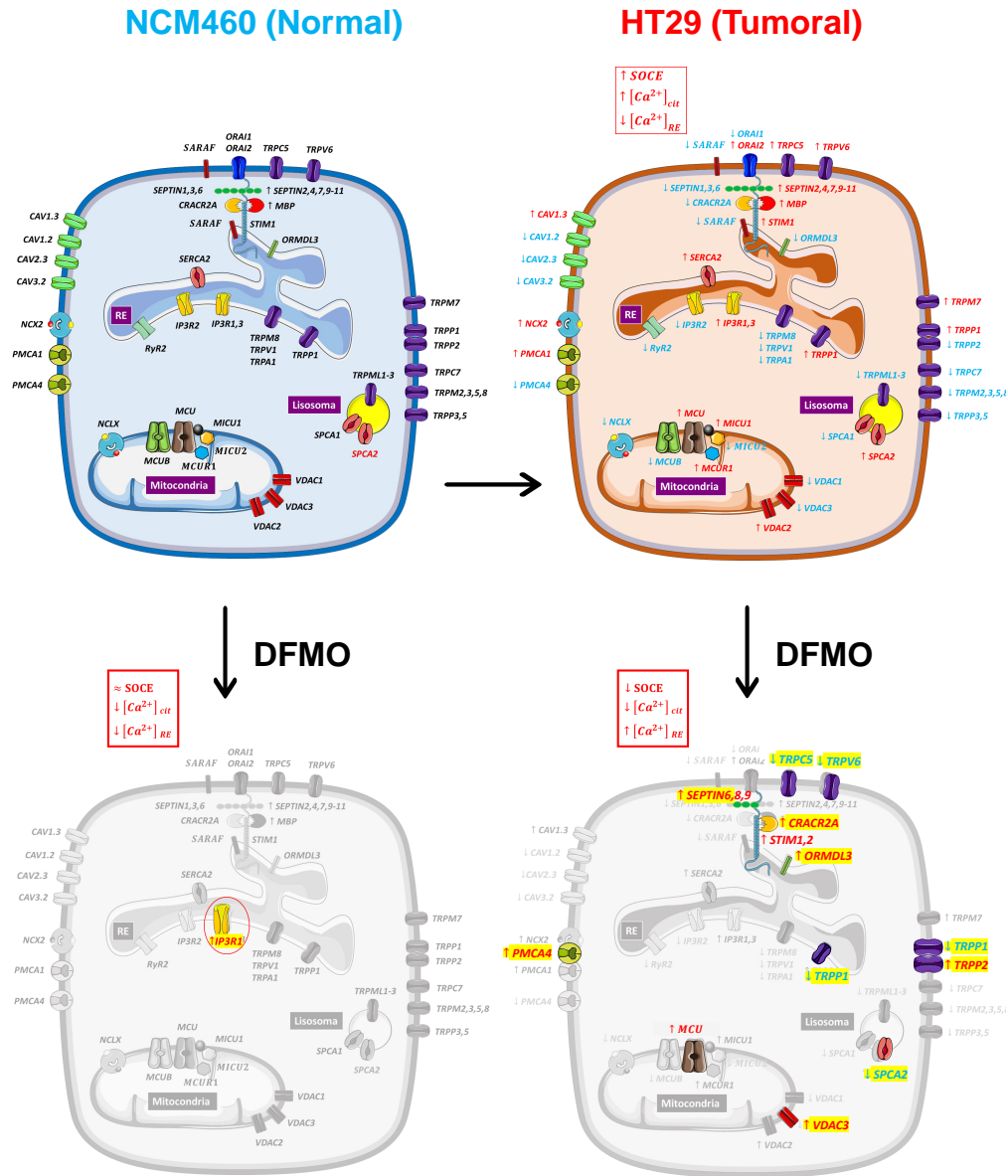


Figura 8.2: Diferencias en la homeostasis del Ca^{2+} intracelular entre las células HT29 y NCM460 tratadas y sin tratar con DFMO. En la parte superior se muestra la expresión diferencial en la comparación entre las células HT29 (derecha) y NCM460 (izquierda) como modelos de células de cáncer de colon y normales, respectivamente. Sobre la figura correspondiente a HT29 se muestran únicamente los genes cuya expresión se ve alterada respecto de las NCM460, donde el color rojo indica sobreexpresión y el azul infraexpresión. El recuadro en rojo muestran las diferencias entre HT29 y NCM460 observadas mediante los experimentos de imagen de Ca^{2+} . En la parte inferior, sobre el resumen de expresión diferencial entre HT29 y NCM460 (en blanco y negro) se superponen los cambios inducidos por el tratamiento con DFMO en cada una de las líneas celulares color los cambios que se producen dentro de cada una de las líneas celulares tras el tratamiento con DFMO. Los genes subrayados en amarillo son aquellos cuya expresión diferencial observada entre las células HT29 y NCM460 en condiciones control se revierte tras el tratamiento con DFMO. Los recuadros en rojo de la parte inferior indican los cambios observados mediante experimentos de imagen de Ca^{2+} que se inducen en cada línea celular tras el tratamiento con DFMO.

8.3. Remodelado del Calcio Intracelular en muestras pareadas de tejido de colon humano normal y tumoral

Los datos discutidos anteriormente han aportado información muy valiosa sobre el transcriptoma del remodelado del calcio intracelular en cáncer colorrectal y la contribución de las poliaminas a estos cambios. Sin embargo, no podemos obviar el hecho de que los datos obtenidos derivan de la comparación de dos líneas celulares modelo: una línea celular normal, aunque inmortalizada, y otra línea tumoral. Ambas líneas proceden originalmente de pacientes de distinta edad, sexo y raza, por lo que, a las diferencias transcriptómicas asociadas al cáncer debemos añadir estos tres factores de confusión. Además, la comparación reflejaría, en todo caso, las diferencias debidas al tipo de tumor que padeció el paciente donante de las células tumorales, su estadio y evolución. Por tanto, a pesar de que el uso de líneas celulares permite obtener información muy relevante al poder comparar el análisis transcriptómico con el funcional, se hace necesario extender el análisis a un modelo más cercano al fisiopatológico del cáncer colorrectal humano. Para ello, solicitamos y obtuvimos una serie de muestras pareadas de colon humano de diferentes paciente, de tal manera que de cada paciente se dispone tanto de tejido tumoral como normal, con el objeto de llevar a cabo tanto el análisis transcriptómico de todo exhaustivo como de la expresión de cinco proteínas implicadas en SOCE. De todas las muestras recibidas, se seleccionaron únicamente muestras pareadas de adenocarcinoma de colon humano.

En el estudio transcriptómico observacional de las muestras de tejido humano pareadas se obtuvo, a grandes rasgos, una pequeña cantidad de genes diferencialmente expresados relacionados con la homeostasis del calcio intracelular, seguramente debido a la gran variabilidad de inherente al factor aleatorio *Paciente*. No obstante, puesto que estos resultados se han empleado en un posterior metaanálisis, aportan una información importante a la hora de obtener un efecto global de expresión diferencial para cada gen a partir de esta y más evidencia obtenida del posterior reanálisis de experimentos transcriptómicos extraídos de repositorios públicos (se verá más adelante).

De forma general, el análisis global de la lista de genes diferencialmente expresados, análisis de enriquecimiento de conjuntos de genes, revela una clara alteración en los conjuntos de genes relacionados con el cáncer, con la señalización de calcio intracelular, con la homeostasis mitocondrial y del retículo y, además, con la ruta de señalización Wnt y la síntesis de poliaminas, ambas características del CRC. El estudio transcriptómico en las muestras pareadas se llevó a cabo mediante la tecnología de microarrays, pues de las tres usadas en las líneas celulares, es la más asequible económicamente, el más utilizado en los estudios llevados a cabo hasta la fecha y, además, menos costoso en términos computacionales lo cual permite poder analizar dichos datos sin necesidad de recurrir a *cloud computing* o a servicios de supercomputación. De esta manera, los resultados obtenidos en el análisis transcriptómico de estas muestras de tejido podrán ser comparados con los resultados obtenidos en las líneas celulares mediante el método de microarrays. A continuación, se enumeran los cambios más relevantes en relación con la homeostasis del calcio intracelular y se comparan con lo observado en el experimento con

microarrays en las líneas celulares:

- En primer lugar, no se encontraron diferencias estadísticamente significativas en cuanto a la expresión diferencial de canales de calcio dependientes de voltaje. Esto contrasta con las diferencias encontradas en líneas celulares donde se obtuvo, mediante microarrays, un aumento de *CAVI.3* y disminución de *CAVI.2* en las células tumorales.
- Respecto a los genes implicados en SOCE, en las muestras pareadas se encontró una disminución del gen *MS4A12* que codifica para un canal operado por depósitos específico de colon y en las septinas *SEPTIN1* y *6* en el tejido tumoral junto con el aumento de la expresión de *ORMDL3* y las *SEPTIN8* y *11*. Los datos discrepan con los resultados obtenidos en líneas celulares donde se observó el aumento de *STIM1*, *MBP* y *SEPTIN9* y *10* en las células tumorales junto con la disminución de *ORAI1*, *CRACR2A*, *ORMDL3*, *SARAF* y *SEPTIN6* y *8*. Por tanto, existe una gran discrepancia entre los genes de SOCE diferencialmente expresados entre las líneas celulares y los de las muestras pareadas de tejido humano de cáncer de colon. *ORMDL3* es una proteína de RE que inhibe la captación de Ca^{2+} por el RE a través de la inactivación de *SERCA2b* y promueve la inactivación dependiente de Ca^{2+} de los canales implicados en SOCE (James, Milstien, and Spiegel (2019)). Por tanto, la sobreexpresión de *ORMDL3* podría conducir a una limitación de SOCE.
- En cuanto a los canales TRP, los resultados en muestras pareadas muestran que en tejido tumoral hay una sobreexpresión de canales *TRPV2* y una disminución de los canales *TRPM2* y *TRPP5*. Estos datos contrastan con los cambios observados en las líneas celulares donde encontramos sobreexpresión de *TRPC4* y *5*, *TRPV5* y *6*, y *TRPP1* en las células tumorales junto con la disminución de la expresión de *TRPC7*, *TRPM2*, *TRPML1*, *2* y *3*, *TRPV1* y *TRPP2*. De nuevo, existe una gran discrepancia entre los cambios observados en líneas celulares y los datos obtenidos en muestras pareadas, pues, únicamente, coincide la disminución de *TRPM2*.
- Respecto a la expresión diferencial de bombas y transportadores, en las muestras tumorales encontramos aumentos significativos en la expresión de las bombas *SERCA2*, *SPCA1* y *SPCA2*. En cambio, en las líneas celulares habíamos observado aumento en la expresión de las bombas *PMCA1* y *SPCA2* y los transportadores *NCX2* y *3* en las células tumorales junto con la infraexpresión de *PMCA4* y *SPCA1*. En este caso, el único gen que coincide en ambos casos es el gen de *SPCA2*. Es interesante el hecho de que este es uno de los pocos genes regulados positivamente por las poliaminas.
- En lo concerniente a los canales de liberación de Ca^{2+} , el estudio en muestras pareadas revela que el tejido tumoral muestra sobreexpresión de *IP3R3* e infraexpresión de *IP3R1*. En el estudio con células, los datos mostraban sobreexpresión de *IP3R3* en células tumorales e infraexpresión de *IP3R2* y *RyR2*. En ambos casos, la sobreexpresión de *IP3R3* junto con la disminución de alguna de las otras dos isoformas de IP3R podría indicar una menor

inactivación de la liberación de calcio desde el RE dependiente de calcio. Por tanto, en este caso, solo la sobreexpresión de *IP3R3* en el fenotipo tumoral es común a ambos modelos.

- Finalmente, en relación a los sistemas de transporte de Ca^{2+} mitocondrial, es muy notable que en las muestras pareadas se observa sobre expresión de prácticamente todos los sistemas de transporte de Ca^{2+} mitocondrial en el tejido tumoral incluyendo sobreexpresión de *MCU*, *MICUI*, *MICU2*, *MCURI*, *VDAC1*, *VDAC3* y *NCLX*. Contrastan los resultados en las líneas celulares donde solo se observa sobreexpresión de *MCU* en las células tumorales junto con infraexpresión de *MICU2*, *MCUb*, *VDAC1*, *VDAC3* y *NCLX*. En cualquier caso, dado el valor de las muestras pareadas, los resultados sugieren que el CCR está asociado a un aumento de la captación de Ca^{2+} mitocondrial.

El análisis de los datos en las muestras pareadas no se limitó al análisis de expresión diferencial convencional, sino que también se aplicaron métodos de machine learning con el fin de encontrar qué más genes de calcio intracelular podrían estar más relacionados con el tejido tumoral y sano. En concreto, se redujo la dimensión de los datos mediante el análisis de componentes principales y, posteriormente, se aplicó el algoritmo de árbol de clasificación C5.0 para encontrar qué componente principal explica la diferencia entre ambos tipos de tejidos. Finalmente, tras encontrar dicha componente, se buscaron los genes que aportan más peso o influencia en ella a través de la estimación de la correlación de Pearson de cada uno de ellos con dicha componente. Por último, se aplica un contraste de hipótesis sobre estos coeficientes de correlación estimados y se corrigen los p valores estimando los valores de FDR mediante el método de Benjamini and Hochberg (1995). De esta manera, a parte de los genes diferencialmente expresados encontrados mediante el análisis de expresión diferencial anterior, se ha detectado, a mayores, algún gen adicional que podría estar alterado entre ambos fenotipos. A continuación se describen los genes que correlacionan, tanto positiva como negativamente, con el fenotipo tumoral:

- Para el caso de los canales de calcio dependientes de voltaje, el nuevo análisis tampoco reveló genes alterados entre ambos fenotipos, por lo que cambios en la expresión de estos genes no correlacionan con el fenotipo tumoral.
- Sin embargo, en los genes implicados en SOCE, el análisis adicional reveló nuevos genes que correlacionan con el fenotipo tumoral incluyendo una correlación positiva de *SARAF* y las *SETIN2* y *7* junto con una correlación negativa de *STIM1* con el fenotipo tumoral.
- El análisis de correlación con la componente principal también añadió nuevos genes de canales TRP relacionados con el fenotipo tumoral incluyendo la correlación positiva con el fenotipo tumoral de *TRPV2* y *TRPP1* y negativa de *TRPC7* y *TRPM1*.
- En cuanto a las bombas y transportadores, el análisis de correlación reveló que las tres bombas diferencialmente sobreexpresadas en tejido tumoral (*SERCA2*, *SPCA1* y *2*) también correlacionaban con el fenotipo tumoral lo que refuerza la importancia de estos 3 genes en el

cáncer colorrectal, especialmente el de *SPCA2* que es potenciado por poliaminas. Este último podría sugerir que, a pesar de haber una menor SOCE, la entrada de calcio desde el medio extracelular podría tener lugar a través de un mecanismo independiente de los depósitos y de STIM a través del complejo *SPCA2/ORAI1*.

- Respecto a los canales de liberación de calcio, el análisis de correlación coincide con el de expresión diferencial respecto a la sobreexpresión de *IP3R3*, aunque no en el caso de la infraexpresión de *IP3R1* en muestras tumorales. Los datos refuerzan, pues, la asociación de la sobreexpresión de *IP3R3* con el fenotipo tumoral.
- Finalmente, el análisis de correlación con el fenotipo tumoral de los genes asociados al transporte de calcio mitocondrial coincide, esencialmente, con el análisis de expresión diferencial y, además, añade el aumento de *VDAC2* en el tejido tumoral. Por tanto, el análisis de correlación refuerza la asociación de la sobreexpresión de casi todos los sistemas de transporte de calcio mitocondrial con el fenotipo tumoral.

Durante toda la discusión se ha puesto de manifiesto la idea de que los cambios observados en la expresión de genes no tienen por qué correlacionar positivamente con cambios en la expresión de las proteínas que codifican y/o en su actividad. Por eso, se decidió investigar los cambios en la expresión de las 5 proteínas más relevantes para SOCE. Para ello, se utilizaron las mismas muestras pareadas para, de forma paralela, aislar tanto el ARNm para el análisis transcriptómico como extractos de proteínas para medir los niveles de las proteínas seleccionadas por *Western blot*: los canales de la familia ORAI, ORAI1 y ORAI2, los dos sensores de calcio del RE STIM1 y STIM2, y el canal TRPC1. Se observaron, a diferencia del análisis transcriptómico, diferencias entre hombres y mujeres en varias proteínas, por lo que el análisis estadístico de los datos se ha hecho teniendo en cuenta el sexo de los donantes. En concreto:

- No se detectaron diferencias significativas en la expresión proteica de ORAI1, ni en varones, ni en mujeres. Este dato contrasta con datos publicados previamente por nuestro grupo que describió una mayor expresión de ORAI1 en las células tumorales respecto de las normales (Sobradillo et al. (2014)), si bien es cierto que en aquel caso sí había sobreexpresión del gen ORAI1 detectado por qRT-PCR.
- Los resultados de ORAI2 son sorprendentes. En primer lugar, encontramos que hay dimorfismo sexual en la expresión proteica de ORAI2, siendo esta isoforma más abundante en mujeres que en hombres, tanto en tejido normal como tumoral. Además, en ambos casos, el tejido tumoral muestra menor expresión proteica de ORAI2. Esto, de nuevo, contrasta con los datos previos publicados en las muestras celulares donde había sobreexpresión de ORAI2, tanto a nivel de ARNm como de proteína en las células tumorales, aunque este canal no parecía contribuir a SOCE en estas células tal y como se demostró mediante estudios de silenciamiento (Sobradillo et al. (2014)). En cualquier caso, en las mismas muestras celulares no se observaron diferencias de ORAI1 ni de ORAI2 de a nivel de ARNm entre el tejido normal y el tumoral.

- Respecto el sensor de Ca^{2+} del RE STIM1, encontramos que, al igual que había ocurrido a nivel de ARNm, la expresión de STIM1 a nivel de proteína estaba disminuida en el tejido tumoral. Sin embargo, este fenómeno solo es estadísticamente significativo en varones pero no en mujeres.
- Los resultados con STIM2 son opuestos a los de STIM1. La expresión proteica de STIM2 es mayor en el tejido tumoral pero este fenómeno solo parece ocurrir en las mujeres y no en los hombres. En este caso, no parece haber correlación entre los cambios a nivel de ARNm y los cambios a nivel de proteína.
- Finalmente, analizamos en detalle la expresión de TRPC1. Encontramos diferencias significativas en el comportamiento de varones y mujeres. Mientras en algunos varones solo se observaba expresión proteica de TRPC1 en el tejido tumoral, en algunas mujeres se observaban niveles similares de TRPC1 en el tejido normal y tumoral. Sin embargo, si se agrupan los datos de varones y mujeres, no se observan diferencias significativas en la expresión de TRPC1 entre tejido normal y tumoral. Estos resultados sugieren que podría haber una gran heterogeneidad en la contribución de TRPC1 al fenotipo tumoral, probablemente relacionada con el estadio de los tumores a lo que se podría sumar un posible dimorfismo sexual. Por lo tanto, sería necesario aumentar el tamaño muestral para poder encontrar, en caso de que existan, diferencias en la expresión de TRPC1 tanto entre ambos sexos como entre el tejido tumoral y normal, así como un posible diferente comportamiento entre sexos en la variación de la expresión de TRPC1 durante la transformación tumoral.

En resumen, los resultados del análisis transcriptómico en muestras pareadas de tejido normal y tumoral arrojan una serie de genes que correlacionan muy bien con el fenotipo celular. Además, de los 30 genes diferencialmente expresados en el tejido relacionados con la homeostasis del calcio intracelular, alrededor de la mitad, 16, coinciden con la expresión diferencial observada en los modelos celulares. En concreto, coincide el aumento de los genes *SEPTIN2,7,11*, *TRPP1*, *SECA2*, *SPCA2*, *IP3R3*, *MCU*, *MICU1*, *MCUR1* y *VDAC2*, y la disminución de los genes *SEPTIN1,6*, *TRPC7*, *TRPM2*, *TRPP5* en el fenotipo tumoral. Además, el análisis a nivel de proteínas de SOCE sugiere la existencia de cierto dimorfismo sexual en la expresión de algunos de estos canales.

A continuación, se muestra un resumen del análisis de expresión diferencial tanto a nivel transcriptómico como de proteína (figura 8.3).

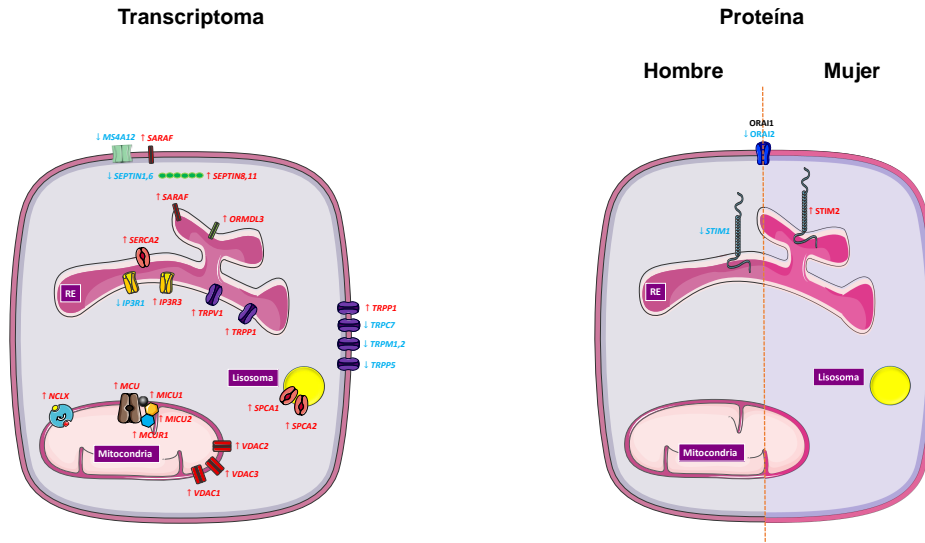


Figura 8.3: Expresión diferencial de genes implicados en la homeostasis del Ca^{2+} intracelular en muestras pareadas de tejido de cáncer de colon humano vs normal. En la izquierda se muestra el análisis de expresión diferencial a nivel transcripcional y los genes que más peso tienen en la clasificación entre muestras tumorales y normales. Para los primeros, los genes en rojo son aquellos cuya expresión en el tejido tumoral es mayor y en azul aquellos que es menor. Para los segundos, el rojo indica que un gen está correlacionado positivamente con la componente principal capaz de agrupar los datos en muestras tumorales y normales, mientras que el rojo indica una correlación negativa. En la figura de la derecha se muestra la expresión diferencial a nivel de cuatro proteínas implicadas en SOCE, de tal manera que en la mitad izquierda se muestran las proteínas diferencialmente expresados solo en hombres, en la derecha las que sólo están diferencialmente expresadas en mujeres, y el medio las que muestran el mismo patrón de expresión diferencial tanto en hombres como en mujeres.

8.4. Metaanálisis del Remodelado del Calcio Intracelular en Cáncer Colorrectal

El análisis de muestras pareadas discutido en el apartado anterior ha aportado información relevante por ser muestras de los mismos pacientes y por haberse llevado a cabo en ellas un doble análisis, pues de cada paciente se ha estudiado tanto su transcriptoma como la expresión de algunas proteínas esenciales. Sin embargo, el número de pacientes es demasiado pequeño para concluir que los datos obtenidos reflejan toda la enorme heterogeneidad de tipos de CCR que pueden diferir en su base genética, localización, estadio y evolución. Para incluir toda esta variedad en nuestro análisis, decidimos reanalizar diferentes estudios transcriptómicos publicados en repositorios públicos realizados con microarrays de Affymetrix, al igual que en la mayoría de análisis realizados en esta tesis, con el fin de, posteriormente, realizar un metaanálisis en el cual también se incluye nuestro propio estudio transcriptómico de muestras pareadas.

Tras el reanálisis de los diferentes experimentos extraídos de GEO y de las muestras pareadas del propio laboratorio, se realizó un metaanálisis de expresión diferencial y otro de enriquecimiento de conjuntos de genes. De cerca de 30000 conjuntos de genes analizados, se obtuvo que casi una

tercera/décima parte de ellos (hipótesis nulas autocontenida/competitiva), 9115/3588, mostraron estar relacionados con el fenotipo de CCR, de los que 578/214 y 106/37 pertenecen a dos bases de datos que contienen conjuntos de genes relacionados con cáncer, *Msigdb Computational Cancer* y *Msigdb Onncogenic*, respectivamente. Además, de los 9115/3588 conjuntos identificados, 50/3 que contienen la palabra *calcium*, 296/147 contienen *cancer*, 26/5 *wnt*, 56/9 *mitochondria*, 12/1 *reticulum* y 3/0 *polyamine*. Por un lado, los datos quedan “validados” en el sentido de que, como era de esperar, hay una evidente alteración debida al cáncer *per se*, como muestra el elevado número de conjuntos de genes relacionados con el fenotipo tumoral identificado en las bases de datos *Msigdb Computational Cancer* y *Msigdb Onncogenic*, e, incluso, de los conjuntos que contienen la palabra cáncer en otras bases de datos. Además, también aparecen la señalización de *wnt* y de la síntesis de poliaminas, cuya alteración ha sido ampliamente descrita en el CCR. Por otro lado, también se hace evidente que la homeostasis de la mitocondria, del RE y del calcio intracelular está alterada en el tejido humano de CCR respecto del normal.

A continuación se compara brevemente los resultados del metaanálisis con los datos obtenidos por microarrays en las muestras pareadas y las líneas celulares:

- Respecto a los canales de calcio dependientes de voltaje, el metaanálisis revela que las nuestras tumorales sobreexpresan *CAV1.3* e infraexpresan *CAV2.1* y *CAV3.2*. (tabla 8.1). Los datos coinciden con la sobreexpresión de *CAV1.3* y la disminución de *CAV3.2* encontrada en células tumorales, si bien es cierto que, en ambos casos, también se observa la disminución de un miembro de *CAV2*, por lo que el modelo de líneas celulares refleja parcialmente el remodelado de los canales de calcio dependientes de voltaje encontrado en el metaanálisis (tabla 8.1).

Tabla 8.1: Comparación entre modelos del análisis de expresión diferencial para VOCC.

| Gen | Modelos | |
|---------------|------------------|--------------|
| | Líneas Celulares | Metaanálisis |
| <i>CAV1.2</i> | ↓ | |
| <i>CAV1.3</i> | ↑ | ↑ |
| <i>CAV2.1</i> | | ↓ |
| <i>CAV2.3</i> | ↓ | |
| <i>CAV3.2</i> | ↓ | ↓ |

- En cuanto a los genes implicados en SOCE, el metaanálisis reveló un gran número de cambios (tabla 8.2). Por un lado, la sobreexpresión de *ORAI1*, *ORAI2* y *ORAI3*, así como de los genes reguladores de SOCE *CRACR2A* y las *SEPTIN2,4,7* y *9* junto con la infraexpresión de *STIM1* y el canal *MS4A12*, y los moduladores de SOCE *STIMATE*, *SARAF*, y las *SEPTIN6,8,10* y *11*. En comparación con las líneas celulares, solo se observó la coincidencia en la infraexpresión de *SARAF*, *SEPTIN1* y *6*, y la sobreexpresión de *ORAI2*, *SEPTIN2,4,7* y *9*. Por tanto, el remodelado de SOCE en las líneas celulares difiere notablemente del

encontrado en el metaanálisis (tabla 8.2). En cualquier caso, los cambios observados en los genes implicados en SOCE serían consistentes con el aumento de la entrada capacitativa de Ca^{2+} fundamentalmente por la sobreexpresión de los 3 ORAI, hecho que también se ha observado en estudios previos del laboratorio en las células tumorales a nivel proteico (Sobradillo et al. (2014)).

Tabla 8.2: Comparación entre modelos del análisis de expresión diferencial para SOCE.

| Gen | Modelos | |
|-----------------|------------------|--------------|
| | Líneas Celulares | Metaanálisis |
| <i>ORA1</i> | ↓ | ↑ |
| <i>ORA12</i> | ↑ | ↑ |
| <i>ORA13</i> | | ↑ |
| <i>STIM1</i> | ↑ | ↓ |
| <i>STIM2</i> | | |
| <i>MS4A12</i> | | ↓ |
| <i>CRACR2A</i> | ↓ | ↑ |
| <i>STIMATE</i> | | ↓ |
| <i>ORMDL3</i> | ↓ | |
| <i>SARAF</i> | ↓ | ↓ |
| <i>MBP</i> | ↑ | ↓ |
| <i>SEPTIN1</i> | ↓ | ↓ |
| <i>SEPTIN2</i> | ↑ | ↑ |
| <i>SEPTIN3</i> | ↓ | |
| <i>SEPTIN4</i> | ↑ | ↑ |
| <i>SEPTIN6</i> | ↓ | ↓ |
| <i>SEPTIN7</i> | ↑ | ↑ |
| <i>SEPTIN8</i> | | ↓ |
| <i>SEPTIN9</i> | ↑ | ↑ |
| <i>SEPTIN10</i> | ↑ | ↓ |
| <i>SEPTIN11</i> | ↑ | ↓ |

- Para los miembros de la superfamilia de canales TRP, el metaanálisis muestra que hay una disminución generalizada en la expresión de estos en las muestras de tejido tumorales incluyendo *TRPC1*, *TRPC6*, *TRPM1*, *TRPM4,5,6,7,8*, *TRPML1* y *2*, *TRPV1* y *3*, *TRPA1* y *TRPP1* junto con el aumento en la expresión de *TRPC7*, *TRPV4* y *TRPP2*. El resultado es conceptualmente similar al obtenido en los modelos celulares, pues hay una clara disminución de los miembros de esta superfamilia (tabla 8.3). Sin embargo, sorprende encontrar que pocos de los cambios observados en las líneas celulares coinciden con los encontrados en el metaanálisis. En concreto, de los 17 TRP diferencialmente expresados según el metaanálisis, únicamente coinciden 6 que, además, están también infraexpresados: *TRPM5*, *TRPM8*, *TRPML1,2*, *TRPV1* y *TRPA1*.

Tabla 8.3: Comparación entre modelos del análisis de expresión diferencial para TRP.

| Gen | Modelos | |
|---------------|------------------|--------------|
| | Líneas Celulares | Metaanálisis |
| <i>TRPC1</i> | | ↓ |
| <i>TRPC5</i> | ↑ | |
| <i>TRPC6</i> | | ↓ |
| <i>TRPC7</i> | ↓ | ↑ |
| <i>TRPM1</i> | | ↓ |
| <i>TRPM2</i> | ↓ | |
| <i>TRPM3</i> | ↓ | |
| <i>TRPM4</i> | | ↓ |
| <i>TRPM5</i> | ↓ | ↓ |
| <i>TRPM6</i> | | ↓ |
| <i>TRPM7</i> | ↑ | ↓ |
| <i>TRPM8</i> | ↓ | ↓ |
| <i>TRPML1</i> | ↓ | ↓ |
| <i>TRPML2</i> | ↓ | ↓ |
| <i>TRPML3</i> | ↓ | |
| <i>TRPV1</i> | ↓ | ↓ |
| <i>TRPV2</i> | | |
| <i>TRPV3</i> | | ↓ |
| <i>TRPV4</i> | | ↑ |
| <i>TRPV6</i> | ↑ | |
| <i>TRPA1</i> | ↓ | ↓ |
| <i>TRPP1</i> | ↑ | ↓ |
| <i>TRPP2</i> | ↓ | |
| <i>TRPP3</i> | ↓ | ↑ |
| <i>TRPP5</i> | ↓ | |

- En cuanto a la expresión de bombas y transportadores, el metaanálisis revela también una disminución generalizada en la expresión de estos genes incluyendo la infraexpresión de *PMCA1*, *PMCA3* y *4*, *SERCA2* y *3*, *NCX2* y *SPCA2* en las muestras tumorales junto con la sobreexpresión de *SPCA1* (tabla 8.4). Los resultados son también diferentes de los encontrados en las líneas celulares con la excepción de la infraexpresión de *PMCA4*. De hecho, el resto de estos genes diferencialmente expresados en las líneas celulares muestran un comportamiento inverso en comparación con el observado en el metaanálisis (tabla 8.4).

Tabla 8.4: Comparación entre modelos del análisis de expresión diferencial para CPT.

| Gen | Modelos | |
|---------------|------------------|--------------|
| | Líneas Celulares | Metaanálisis |
| <i>PMCA1</i> | ↑ | ↓ |
| <i>PMCA3</i> | | ↓ |
| <i>PMCA4</i> | ↓ | ↓ |
| <i>SERCA2</i> | ↑ | ↓ |
| <i>SERCA3</i> | | ↓ |
| <i>NCX2</i> | ↑ | ↓ |
| <i>SPCA1</i> | ↓ | ↑ |
| <i>SPCA2</i> | ↑ | ↓ |

- Respecto a los canales de liberación de calcio, el metaanálisis revela la sobre expresión de *IP3R2* en muestras tumorales junto con la infraexpresión de *IP3R1* y *RYR3* (tabla 8.5). Los datos difieren también de los resultados en líneas celulares donde el fenotipo tumoral estaba asociado a sobreexpresión de *IP3R1* e *IP3R3*, e infraexpresión de *IP3R2* y *RYR2*. Así, mientras que tanto *IP3R1* como *IP3R3* tienen un comportamiento totalmente opuesto en ambos modelos, también se observa que, en ambos, hay una disminución de uno de los miembros de RyR, *RyR2* y *RyR3*, respectivamente (tabla 8.5).

Tabla 8.5: Comparación entre modelos del análisis de expresión diferencial para CLC.

| Gen | Modelos | |
|--------------|------------------|--------------|
| | Líneas Celulares | Metaanálisis |
| <i>IP3R1</i> | ↑ | ↓ |
| <i>IP3R2</i> | ↓ | ↑ |
| <i>IP3R3</i> | ↑ | |
| <i>RYR2</i> | ↓ | |
| <i>RYR3</i> | | ↓ |

- Finalmente, los sistemas de transporte de calcio mitocondrial mostraron una inhibición generalizada en las muestras tumorales a excepción de *MCUb* y *VDAC1*, pues aparecen sobreexpresados en el tejido tumoral (tabla 8.6). El perfil de expresión, en este caso, también difiere al observado en los modelos celulares, pues en los modelos celulares se observó la sobreexpresión de *MCU*, *MICU1*, *MCUR1* y *VDAC2*, y la disminución de *MICU2*, *MCUb* y *VDAC1* (tabla 8.6).

Tabla 8.6: Comparación entre modelos del análisis de expresión diferencial para MCT.

| Gen | Modelos | |
|--------------|------------------|--------------|
| | Líneas Celulares | Metaanálisis |
| <i>MCU</i> | ↑ | ↓ |
| <i>MICU1</i> | ↑ | ↓ |
| <i>MICU2</i> | ↓ | ↓ |
| <i>MICU3</i> | | ↓ |
| <i>MCUR1</i> | ↑ | |
| <i>EMRE</i> | | ↓ |
| <i>MCUb</i> | ↓ | ↑ |
| <i>VDAC1</i> | ↓ | ↑ |
| <i>VDAC2</i> | ↑ | ↓ |
| <i>VDAC3</i> | ↓ | |
| <i>NCLX</i> | ↓ | ↓ |

Los resultados, en principio, sugieren que los modelos celulares no reflejan adecuadamente el remodelado de la homeostasis del calcio intracelular que resulta del metaanálisis transcriptómico. En particular, de los 58 de estos genes que muestran expresión diferencial en el metaanálisis, únicamente coincide la expresión diferencial de 19 de ellos: sobreexpresión de *CAV1.3*, *ORAI2*, *SARAF*, *SEPTIN2,4,7* y *9*, e infraexpresión de *CAV3.2*, *SEPTIN1,6*, *TRPM5*, *TRPM8*, *TRPML1,2*, *TRPV1*, *TRPA1*, *PMCA4*, *MICU2* y *NCLX*. Por el contrario, la expresión diferencial de 22 genes muestra un comportamiento totalmente contrario entre las muestras de tejido analizadas en el metaanálisis y los modelos celulares: *ORAI1*, *STIM1*, *CRACR2A*, *MBP*, *SEPTIN10,11*, *TRPC7*, *TRPM7*, *TRPP1,3*, *PMCA1*, *SERCA2*, *NCAX2*, *SPCA1,2*, *IP3R1,2*, *MCU*, *MICU1*, *MCUb*, *VDAC1* y *VDAC2*. Además, otros 16 genes que han resultado diferencialmente expresados en el metaanálisis no mostraron cambio alguno en los modelos celulares. Ello puede ser debido a varios factores. En primer lugar, los modelos celulares pueden reflejar un único tipo de cáncer colorrectal que estaría poco representado en el metaanálisis. Además, los modelos celulares incluyen diferencias no relacionadas con el cáncer, como por ejemplo las diferencias debidas al sexo, a la edad y a la raza de los donantes de los modelos celulares. Finalmente, el análisis transcriptómico de los tejidos incluye células de vasos y del estroma tumoral que pueden contribuir a general diferencias adicionales a las estrictamente debidas a la transformación de células epiteliales en células tumorales. Sin embargo, a pesar de las diferencias moleculares, los cambios transcriptómicos podrían dar lugar a fenotipos funcionales que, al menos en algunos aspectos, podrían ser similares a los descritos en el modelo celular. Por tanto, procede discutir qué cambios en la homeostasis del calcio intracelular podrían inferirse de los cambios descritos a nivel transcriptómico. Este análisis no es, en absoluto, obvio, pues no disponemos de la información a nivel proteómico ni funcional (medidas de corriente, imagen de calcio...).

En vista de estos resultados, no parece tan evidente, por lo menos en un primer análisis, que en las muestras de tejido humano de CCR, respecto de las normales, analizadas en el metaanálisis están favorecidos o no SOCE y otros mecanismos de entrada de calcio a través del medio extracelular. En primer lugar, la sobreexpresión de las tres isoformas de *ORAI*, *CRACR2A*, *SEPTIN2,4*, *TRPV4* y

la infraexpresión de *SARAF*, *MBP* y *TRPM4* indicarían una mayor SOCE. No obstante, el aumento de *SEPTIN7* y la disminución de *TRPC1,6*, *TRPM7*, *MS4A12*, *STIM1* y *STIMATE* sugieren una disminución de SOCE en el tejido humano de CCR. Además, la disminución de *TRPC1* no solo podría disminuir la entrada no selectiva de calcio activada por el vaciamiento de los depósitos a través de estos canales, sino también a través de *TRPC3*, pues el *TRPC1*, a parte de formar heterodímeros con *ORAI1*, también los forma con *TRPC3/6/7*. Además, la activación de *TRPC3* por *STIM1* depende de *TRPC1*, pues por sí solo no puede ser activado por esta vía (J. P. Yuan et al. (2007)). Además, también se ha descrito que en el CCR se produce un cambio de hacia *TRPC1-STIM2* (Sobradillo et. al (2014)). En cuanto a la disminución de *TRPM4* y *TRPM7*, podría afectar a SOCE, pues, por un lado, el *TRPM4* actúa como regulador negativo de SOCE y, por otro lado el *TRPM7* lo hace como regulador positivo (Kappel, Borgström, et al. (2019), Kádár et al. (2021)). Sin embargo, puesto que la síntesis de poliaminas está potenciada en las células de CCR y que la actividad de *TRPM7* se potencia por estas, su disminución podría ser simplemente un mecanismo compensatorio debido a una excesiva activación (Lange and Koomoa (2014)). Además, la pérdida de *TRPM4* podría también favorecer SOCE (Kappel, Borgström, et al. (2019)).

Puesto que los resultados anteriores parecen ser algo contradictorios, aunque apuntan hacia un aumento de SOCE, esta información podría complementarse con evidencia ya existente. En concreto, podría cruzarse el perfil de expresión anterior con los resultados obtenidos tanto de los análisis transcriptómicos de las células HT29 y NCM460, tratadas y no tratadas con DFMO. Por un lado, de los experimentos de imagen en las células no tratadas se sabe que SOCE aumenta y que su perfil de expresión muestra un aumento de *ORAI2*, *STIM1*, *TRPC5*, *TRPV6*, *SEPTIN2,4,8,9,10,11* y *TRPM7* y la disminución de los reguladores negativos *SARAF* y *ORMDL3*, lo que inclina la balanza hacia una mayor SOCE. Sin embargo, también disminuye *ORAI1*, *TRPC7*, los reguladores positivos de SOCE *CRACR2A*, *SEPTIN1* y aumentan los reguladores negativos *SEPTIN7* y *MBP*. Por otro lado, de los experimentos en las células HT29 tratadas con DFMO, en las que se observa la disminución de SOCE, se deduce que la sobreexpresión del regulador negativo de SOCE *ORMDL3* junto con la disminución de *TRPC1,5* y *TRPV6* es suficiente para compensar la sobreexpresión de *STIM1,2*, de las *SEPTIN6,8,9* y del regulador positivo *CRACR2A*.

Poniendo todos estos datos en conjunto con los resultados del metaanálisis, el aumento de SOCE podría estar apoyado por diferentes patrones de expresión. En primer lugar, se observa que en las HT29 tanto cuando disminuye como cuando aumenta SOCE, *STIM1* aumenta su expresión. Además, también se observa un aumento de *SEPTIN2,4,7,9* en las células con mayor SOCE. Por lo tanto, esto apoyaría el aumento de SOCE en las muestras de tejido tumoral del metaanálisis, pues muestran este mismo patrón de expresión. En cuanto a *ORAI*, se observa que aumenta *ORAI2* y disminuye *ORAI1* en las HT29 control (mayor SOCE) respecto de las NCM460 control, por lo que el aumento de las tres isoformas en el metaanálisis podría indicar un aumento de SOCE en el tejido tumoral. En cuanto a los reguladores de SOCE, en las HT29, respecto de las NCM460, se ha observado que disminuyen los reguladores negativos *SARAF* y *ORMDL3* pero aumentan otros como *MBP*, mientras que en las tratadas con DFMO, en las que disminuye SOCE, aumenta la expresión del regulador positivo *CRACR2A* pero también aumenta la del regulador negativo *ORMDL3*. Así,

tanto en el aumento de SOCE como en su disminución, coexisten diferentes patrones de aumento y disminución en la expresión de reguladores tanto positivos como negativos, por lo que el cómputo de los efectos de cada uno de ellos podrá determinar las características de este mecanismo. Por lo tanto, los resultados anteriores podrían indicar, en su cómputo global, un aumento de SOCE en las células de tejido tumoral analizadas en el metaanálisis. En concreto, el aumento de las tres isoformas de ORAI, del regulador positivo *CRACR2A*, la disminución de los reguladores negativos *SARAF*, *MBP* y *TRPM4*, e la disminución de *SETPINI,6* y el aumento de *SEPTIN2,4,7,9* que también se observa en los modelos celulares podría indicar una mayor SOCE en las células de los tejidos de cáncer de colon analizados en el metaanálisis, a pesar de mostrar la disminución de *STIMI* (que aumenta en las líneas celulares tras tratarlas con DFMO y que muestran menor SOCE), de los *TRPCI,6*, o del regulador positivo del SOCE *STIMATE*.

No obstante, como se verá más adelante, los depósitos de las células del tejido tumoral podrían tener un menor contenido de calcio en condiciones basales, al igual que sucede en las líneas celulares estudiadas HT29 vs NCM460. Además, puesto que *STIM2* no está infraexpresado, este sensor podría estar detectando este menor tamaño de los depósitos y acabar activando SOCE en las células del tejido tumoral que, debido a la menor expresión de *SERCA2-3*, su llenado estaría desfavorecido y, por ende, SOCE podría permanecer activada más tiempo. Además, a esto habría que sumarle la sobreexpresión del regulador positivo de SOCE *CRACR2A*, cuya unión a ORAI está conservado en las tres isoformas. De hecho, Sobradillo et al. (2014) observó que en diferentes líneas celulares de CRC (HT29, SW480-ADH y SW480-R) vs. normales (NCM460 y NCM456) SOCE está aumentada. Además, también observó que la expresión diferencial entre HT29 y NCM460 identificó, a nivel proteico, el aumento de las tres isoformas de ORAI, el aumento de *STIM1* y la disminución de *STIM2*, pero a nivel transcriptómico ambas isoformas de *STIM* aumentaron. Por lo tanto, el perfil de expresión encontrado podría estar, en parte, de acuerdo con un aumento de SOCE en el tejido de CRC estudiado, pues a pesar de la disminución de *STIMI*, *STIM2* no disminuye y, a la par, aumentan las tres isoformas de *ORAI*. Además, a esto anterior habría que sumarle la sobreexpresión del regulador positivo de SOCE *CRACR2A*, cuya unión a ORAI está conservado en las tres isoformas. En cuanto a las septinas, la sobreexpresión tanto de *SEPTIN2,4* como de *SEPTIN7* y la infraexpresión de *SEPTINI*, a pesar de que por separado puedan parecer resultados contradictorios, en conjunto podría sugerir un aumento de los complejos filamentosos que estos forman e, incluso, un aumento de los complejos filamentosos más cortos, pues aumentan dos miembros pertenecientes al subgrupo *SEPTIN2* (*SEPTIN2,4*) y *SEPTIN7* al a par que disminuye otro miembro del *SEPTIN2* (*SEPTINI*) y, como resultado, el ratio miembros *SEPTIN2/SEPTIN7* estaría favorecido. En consecuencia, los filamentos más cortos de septinas ayudan a la organización y oligomerización de ORAI en la membrana plasmática lo cual podría indicar que SOCE está favorecida (Deb and Hasan (2016)). Además, la sobreexpresión de *TRPV4* podría poner de manifiesto la participación no solo de la corriente CRAC sino otras tipo SOC descritas en líneas celulares de CCR (Sobradillo et al. (2014)).

En segundo lugar, en lo que se refiere a la entrada de calcio desde el medio extracelular independiente de depósitos, al igual que en el caso de SOCE, también se encuentran resultados

contradictorios difíciles de interpretar. Por un lado, la entrada a través de los canales ARC estaría favorecida por la sobreexpresión de *ORAI1,3* junto con la infraexpresión de *SARAF*, pero desfavorecida por la infraexpresión de *STIM1*, por lo que sería interesante comprobar, por ejemplo mediante ensayos de inmunistoquímica, si la infraexpresión de *STIM1* tiene lugar en toda la célula, en la membrana plasmática o en la del RE. Otro posible mecanismo de entrada de calcio independiente de depósitos y, además, también de STIM, es a través de la interacción entre *ORAI1* y *SPCA2*, pero, al igual que en los casos anteriores, mientras que *ORAI1* está sobreexpresado, *SPCA2* está infraexpresado, por lo que tampoco puede inferirse el comportamiento únicamente con estos datos. También se conoce otra entrada independiente de depósitos a través de *ORAI1* que tiene lugar cuando *SEPTIN7* está disminuida junto con *IP3R* y/o *STIM1* también infraexpresados, pero sucede exactamente lo mismo; es decir, mientras que *ORAI1* está sobreexpresado y *STIM1* e *IP3R* disminuidos, *SEPTIN7* está sobreexpresado. También, la entrada no selectiva de Ca^{2+} podría verse disminuida, tal y como indicaría la infraexpresión de los *TRPM1,6-8*. En cuanto a la sobreexpresión de *CAVI.3* junto con la disminución de *CAV2.1* y *CAV3.2*, posiblemente no tenga gran efecto sobre la concentración de calcio citosólico, pues la disminución de las isoformas de miembros de las familias Cav2 y Cav3 podría estar induciendo, mediante mecanismos compensatorios, la sobreexpresión de un miembro de la familia Cav1 -Cav1.3 en este caso- con el fin de intentar mantener la homeostasis del Ca^{2+} intracelular y en los procesos en los que participen los VOCC (Aldea et al. (2002)).

Por lo tanto, es evidente que la información obtenida del perfil de expresión de estas muestras proporciona información contradictoria en algunos casos y, en consecuencia, estos datos no son suficientes para predecir qué sucede con estos mecanismos de entrada de calcio desde el medio extracelular. Por ello, sería necesario realizar estudios funcionales de estos mecanismos, ya sea mediante técnicas electrofisiológicas como *patch-clamp* u otras como imagen de Ca^{2+} para aclarar estas contradicciones. No obstante, las muestras empleadas en el laboratorio no eran adecuadas para realizar estos estudios funcionales, pues estaban congeladas y preservadas en OCT. Afortunadamente, a pesar de que los datos anteriores respecto de la entrada independiente de depósitos no parecen clarificar si están o no favorecidos, sí que se han encontrado datos que indicarían una mayor entrada de calcio extracelular. Por un lado, podría tener lugar la entrada de calcio no selectiva a través del *TRPC7* independiente de los depósitos, pues la sobreexpresión de *TRPC7* junto con la disminución de *SERCA2* indica que esta entrada estaría potenciada (DeHaven et al. (2009)). Por otro lado, la sobreexpresión de *CAVI.3* podría estar potenciando la actividad del *TRPV4* (Tajada and Villalobos (2020)) que, además, también está sobreexpresado y, en consecuencia, se estaría aumentando aún más la entrada de Ca^{2+} desde el medio extracelular por estos mecanismos independientes de depósitos.

El perfil de expresión de los genes relacionados con la homeostasis del Ca^{2+} del RE, como es la disminución de *SERCA2,3* y *CAV2.1* (H. Ma et al. (2012)), podría indicar un menor llenado de estos. En cuanto a su liberación, la sobreexpresión de *IP3R2* y *TRPV4*, pues este último también se expresa en la membrana del RE, podría indicar una mayor liberación de Ca^{2+} desde el RE que junto con la disminución de *IP3R1* sugiere que dicha liberación presenta una menor inactivación

dependiente de Ca^{2+} . No obstante, se observa una disminución de *TRPA1*, *TRPP1* y *RyR3*, lo que podría deberse a un mecanismo compensatorio para evitar una liberación excesiva de calcio desde el RE, por ejemplo, inducida por la una mayor liberación de calcio a través de los IP3R2 que pudiera acabar por vaciar los depósitos y provocar un excesivo estrés del RE que diera lugar a la apoptosis. Además, la disminución de *TRPA1* también podría ser un mecanismo de defensa frente a la muerte por ROS (Shimizu, Takahashi, and Mori (2014)). Además, la disminución del tamaño de los depósitos *per se* podría resultar en un mecanismo de resistencia a la muerte celular (Sobradillo et al. (2014)). En este sentido, la sobreexpresión de *SPCA1* podría evitar una excesiva sensibilización de las células de estos tejidos al estrés del RE. En cuanto a la salida de Ca^{2+} al medio extracelular, la disminución de las *PMCA1,3,4* y del *NCX2* sugiere que hay un aumento del calcio intracelular, lo que podría suponer un problema para la supervivencia de la célula, pues dicho aumento podría provocar una captación excesiva por la mitocondria. Sin embargo, como se verá más adelante, el transporte de calcio hacia la mitocondria está claramente atenuado.

Algunos de los genes relacionados con la homeostasis del calcio intracelular estudiados también podrían estar relacionados con diferentes características fenotípicas de las células tumorales tales como la resistencia a la muerte, la capacidad metastásica, la autonomía proliferativa e, incluso, con el peor pronóstico de la enfermedad y/o la resistencia a tratamientos anticancerígenos. Respecto a la resistencia a la muerte, la reducción de los canales no selectivos de calcio *TRPM4,7*, *TRPA1*, *TRPML1* podría prevenir la muerte celular por apoptosis debido al estrés oxidativo, pues estos se activan por ROS (Hayes, Dinkova-Kostova, and Tew (2020)). Además, la disminución de *TRPM1,6-8*, que presentan una permeabilidad no selectiva por calcio, podría evitar la entrada excesiva de calcio a la célula y, en consecuencia, evitar la apoptosis mitocondrial asociada a la sobrecarga de calcio. Incluso la simple disminución de *TRPM7*, el cual puede tanto actuar como modulador positivo de ORAI como introducir Ca^{2+} al disminuir el pH (Kádár et al. (2021)), podría contribuir a evitar la apoptosis mitocondrial evitando la entrada excesiva de calcio desde el medio extracelular tanto dependiente de los depósitos como por la disminución del pH característico de las células cancerígenas en su medio intracelular debido a la reprogramación metabólica. También, la disminución de *TRPC1* podría favorecer la resistencia a la muerte de las células de estos tejidos tumorales, pues se ha descrito que la inducción de este gen promueve la apoptosis del epitelio intestinal al inhibir NF- κ B. Otro TRP importante en la regulación de la muerte celular es el TRPV4, de tal manera que su sobreexpresión podría proteger a las células cancerígenas frente a la apoptosis. Además, el perfil de expresión de los genes que participan en el transporte de calcio mitocondrial también evidencian la resistencia a la muerte de estos tejidos cancerígenos. Por un lado, la infraexpresión de *VDAC2* junto con el aumento de *VDAC1* podría ejercer un papel protector frente a la apoptosis en las células de estos tejidos cancerígenos, pues, por un lado, *VDAC2* participa activamente en la apoptosis a través de la unión a ceramida, pues su unión desencadena la apoptosis mitocondrial (Dadsena et al. (2019)). Por otro lado, *VDAC1* es capaz de unirse a las proteínas antiapoptóticas HKI y HKII para proteger a las células tumorales de la apoptosis, ya que así disminuye la producción de mROS, y, además, también es capaz de favorecer su proliferación (Camara et al. (2017)). Por el contrario, si HK se separa de *VDAC1* entonces se induce la liberación

de citocromo c. En cuanto al transporte mitocondrial de calcio, la sobreexpresión del inhibidor negativo de MCU *MCUb* junto con la infraexpresión de *MCU* evidencia la menor capacidad de la mitocondria para actuar como *buffer* de calcio y, en consecuencia, se evita que la supuesta mayor concentración de Ca^{2+} que podría haber diera lugar a la sobrecarga mitocondrial y la inevitable muerte por apoptosis. Evidentemente, dada la menor capacidad de captar calcio por la mitocondria, la menor expresión de *NCLX* no tendría, en principio, importancia en este sentido. Además, puesto que hay menos expresión de *MCU*, tampoco parece relevante que sus reguladores negativos *MICU1-2*, el adaptador *EMRE*, e incluso los positivos como *MICU3* estén infraexpresados.

En cuanto a la capacidad metastásica de las muestras analizadas, se ha descrito que la disminución del *NCLX* podría jugar un papel clave en el aumento de esta en las células de cáncer de colon a través del aumento de calcio mitocondrial, pues este conlleva el aumento de ROS. Así, este aumento de ROS podría activar la señalización por HIF1 α la cual favorece la metástasis de aquellas células con baja expresión de *NCLX* (Pathak et al. (2020)). Esta baja expresión se ha asociado con mutaciones en TP53 y tumores BRAF wild type. Cabe destacar que se ha observado una reducción de *NCLX* desde el estadio I al IV. En este sentido, la reducción en el tejido de CRC de *TRPM1* y del miR-211 codificado por uno de sus intrones podría favorecer su capacidad metastásica, pues este miR es capaz de reducir la expresión de genes relacionados con la capacidad metastásica como IGF2R, TFGBR2 o NFAT5 (Fiorio Pla and Gkika (2013)) (Fiorio Pla and Gkika (2013)). Otro canal de la superfamilia de los TRP que podría favorecer la metástasis es el TRPA1, pues se ha relacionado su disminución con un aumento de la capacidad migratoria de las células cancerígenas (Cojocararu et al. (2021)) y con un peor pronóstico del CCR (datos disponibles en <https://www.proteinatlas.org/ENSG00000104321-TRPA1/pathology>, v20.1proteinatlas.org). El TRPV4, ha mostrado ser relevante en el proceso tumoral, pues su sobreexpresión podría proteger a las células cancerígenas frente a la apoptosis o la autofagia, favorecer la progresión tumoral y la transición de estas células hacia un fenotipo más invasivo (A. Fiorio Pla et al. (2012), Arbabian et al. (2020), D. Yang and Kim (2020), Rizopoulos and Assimakopoulou (2021)). En este sentido, la sobreexpresión del *CAVI.3*, que ya ha sido descrita con anterioridad en el CCR (C.-Y. Wang et al. (2015)), participa en el proceso tumorigénico mediante la modulación de la actividad de canales como los de potasio o el TRPV4 (Tajada and Villalobos (2020)). De hecho, se ha descrito que los VOCC están implicados en procesos tumorales, como por ejemplo la proliferación, mitogénesis, diferenciación, apoptosis y metástasis (X.-T. Wang et al. (2000), Phan et al. (2017)).

Otro gen sumamente importante como marcador del pronóstico del CCR es *MS4A12*, proteína implicada no solo en SOCE, sino también en la proliferación, diferenciación y regulación del ciclo celular, pues se ha relacionado su disminución con un peor pronóstico del CCR (L. He, Deng, and others (2017), J. Han et al. (2020), J. W. Han et al. (2021)). De hecho, su disminución ha mostrado una pérdida de la diferenciación de las células cancerígenas en el CCR, lo que es característico de las células cancerígenas y que, además, esta falta de diferenciación da cuenta de la proliferación de estas células. Otro miembro de SOCE que también ha sido descrito como indicador de mal pronóstico en el CRC es el *ORAI3* (Ibrahim et al. (2019)), el cual también está sobreexpresado en los tejidos de CCR analizados. Otros genes relevantes en el proceso tumoral relacionados con

el transporte del calcio son *VDAC1* y *SPCA1*, cuya sobreexpresión podría indicar un aumento de la autonomía proliferativa. También, el TRPV1 es capaz de actuar como supresor tumoral en el CCR (Rizopoulos and Assimakopoulou (2021)), por lo que la infraexpresión observada de este canal también podría favorecer la progresión tumoral. También se ha detectado la alteración en la expresión de TRPM8, cuya localización celular aberrante se ha relacionado con la progresión tumoral (Monteith et al. (2007), Bidaux et al. (2007)). En cuanto a los tratamientos anticancerígenos, la reducción de *TRPM6-7* favorece la resistencia a quimioterápicos como doxorubicina (Castiglioni et al. (2015), Locatelli et al. (2021)). En este sentido, la infraexpresión del *NCLX* podría jugar también un papel clave en la quimiorresistencia (Pathak et al. (2020)).

Por último, atendiendo al perfil de expresión obtenido del metaanálisis centrado en los genes del transporte de calcio intracelular y a los cambios observados en las células HT29 tras el tratamiento con DFMO, bajo el supuesto de que el comportamiento de estas células sea similar que las células cancerígenas de los tejidos cancerígenos evaluados en el metaanálisis, sería posible revertir parcialmente el fenotipo tumoral modificando la expresión de los siguientes 5 genes: *STIM1*, *SEPTIN6,8*, *PMCA4* y *MCU*. Por lo tanto, al igual que sucede en las líneas celulares, el tratamiento con DFMO podría alterar SOCE, la $[Ca^{2+}]_{cit}$ y, además, aumentar la entrada de Ca^{2+} a la mitocondria.

En resumen, el metaanálisis transcriptómico del cáncer colorrectal revela numerosos cambios en la expresión de genes asociados al fenotipo tumoral que discrepan en buena medida no solo de los encontrados en los modelos celulares sino también de nuestro propio estudio transcriptómico en muestras pareadas. Del estudio transcriptómico no se puede derivar cambios claros en la homeostasis del calcio intracelular al desconocer los cambios a nivel proteico y funcional. Sin embargo, los datos sugieren cambios similares, al menos parcialmente, en la homeostasis del calcio intracelular a los encontrados en los modelos celulares que podrían explicar algunas características fenotípicas de las células tumorales. Se hace necesario, pues, un estudio proteómico y funcional en muestras pareadas de cáncer colorrectal humano.

A continuación se presenta un resumen esquemático del análisis de expresión diferencial obtenido a través del metaanálisis (8.4).

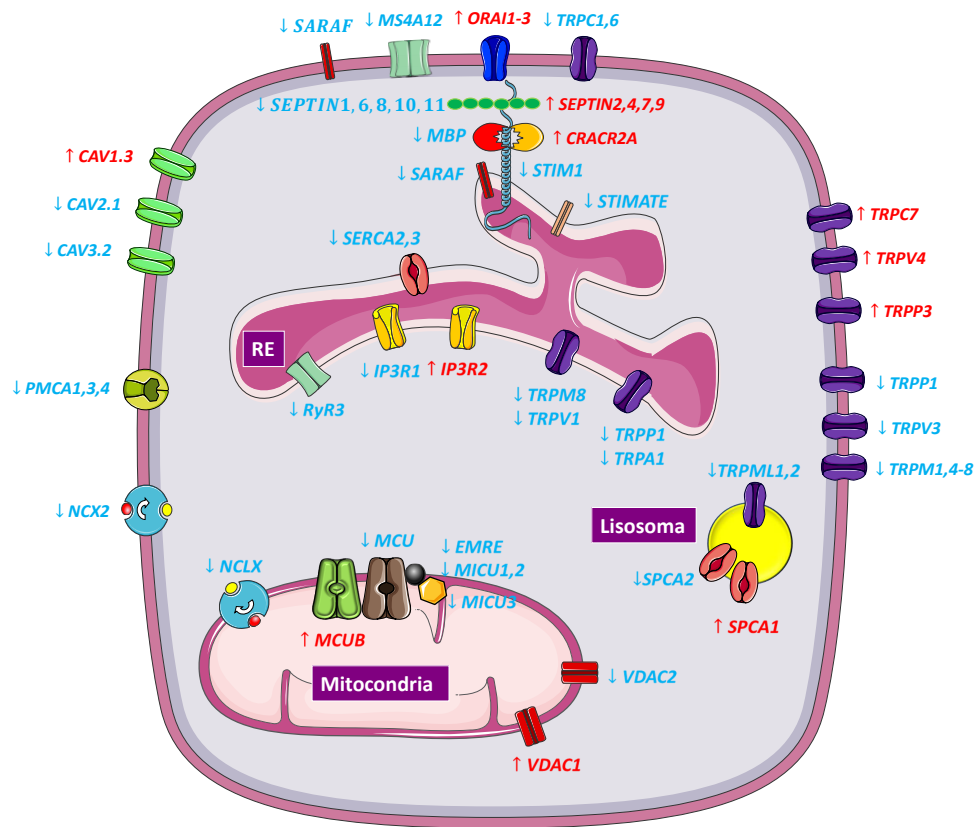


Figura 8.4: Metaanálisis de expresión diferencial entre el tejido de cáncer de colon y normal de los genes relacionados con el transporte de Ca^{2+} intracelular. En la representación esquemática de la célula tumoral se muestran aquellos genes que están diferencialmente expresados entre el tejido tumoral y normal de colon. En rojo se representan los que están sobreexpresados y en azul aquellos infraexpresados

Parte VI

Conclusiones

“Falta de conocimiento acerca de todos los factores y el hecho de no incluirlos en nuestro razonamiento impone falsas conclusiones.”

Richard Bunckminster Fuller

Capítulo 9

Conclusiones

1. Las células modelo de cáncer de colon humano HT29 presentan una mayor concentración de Ca^{2+} citosólico basal, una mayor entrada capacitativa de Ca^{2+} y un menor nivel de llenado de los depósitos intracelulares que las células NCM460, modelo de epitelio normal de colon humano. Estas diferencias podrían reflejar un remodelado de la homeostasis del Ca^{2+} intracelular asociado al cáncer colorrectal que contribuiría al fenotipo tumoral caracterizado por la mayor tasa de proliferación celular, capacidad de invasión y resistencia a la apoptosis.
2. El análisis transcriptómico de las células tumorales vs. las células normales identificó un amplio abanico de conjuntos de genes relacionados con el fenotipo tumoral, entre los que destacan numerosos conjuntos de genes relacionados con el cáncer, con la señalización por Wnt y con los orgánulos mitocondria y retículo endoplasmático.
3. El análisis transcriptómico triple realizado ha detectado la sobreexpresión de 24 genes implicados en transporte de calcio en las células tumorales: los canales *CAVI.3*, *ORAI2* (y sus reguladores *STIM1*, *MBP*, y las *SEPTIN2,4,7,9-11*), *TRPC5*, *TRPV6*, *TRPM7* y *TRPP1*, las bombas *PMCA1*, *SERCA2*, *SPCA2*, el intercambiador *NCX2*, los canales intracelulares *IP3R1* e *IP3R3*, el uniportador de calcio mitocondrial *MCU*, su modulador *MICU1*, *MCUR1*, y el canal *VDAC2* de la membrana mitocondrial externa, junto con la infraexpresión de 32 genes incluyendo los canales *CAVI.2*, *CAV2.2* y *CAV2.3*, *ORAI1* (y sus moduladores *ORMDL3*, *SARAF*, *CRACR2A* y las *SEPTIN1,3* y 6), *TRPC7*, *TRPP2,3* y 5, *TRPM2,3,5* y 8, *TRPV1*, *TRPA1* y *TRPML1,2* y 3, las bombas *PMCA4* y *SPCA1*, los canales intracelulares *IP3R2* y *RyR2*, los reguladores negativos de la captación de Ca^{2+} mitocondrial *MICU2* y *MCUb*, los canales de la membrana mitocondrial externa *VDAC1* y 3 y el intercambiador mitocondrial *NCLX*.
4. La depleción de poliaminas con DFMO revierte, al menos parcialmente, tanto los cambios en la homeostasis del Ca^{2+} intracelular como los del transcriptoma de las células tumorales. DFMO modula hasta 4402 genes diferencialmente expresados en las células tumorales frente a solo 162 genes en las normales. DFMO disminuye la expresión de *TRPC5*, *TRPV6*,

- TRPP1* y *SPCA2* que estaban sobreexpresados en las células tumorales, a la vez que aumenta la expresión de *CRACR2A*, *ORMDL3*, *SEPTIN6*, *TRPP2*, *PMCA4* y *VDAC3* que estaban infraexpresados en las células tumorales. En cambio, en las células normales, DFMO solo modula la expresión de un único gen, el canal intracelular *IP3R1*.
5. El análisis transcriptómico por microarrays llevado a cabo en las muestras pareadas de tejido de colon humano normal y tumoral revela que el fenotipo tumoral correlaciona con la sobreexpresión de 21 genes incluyendo los reguladores de SOCE (*SARAF*, *ORMLD3* y *SEPTIN2,7,8* y *11*), los canales *TRPV1* y *2* y *TRPP2*, las bombas *SERCA2*, *SPCA1* y *SPCA2*, el canal intracelular *IP3R3* y los sistemas de transporte de Ca^{2+} mitocondriales *MCU*, *MICU1*, *MICU2*, *MICUR1*, *VDAC1,2* y *3* y *NCLX*. El fenotipo tumoral correlaciona además con la infraexpresión génica de solo 9 genes incluyendo el canal específico de colon *MS4A12*, los reguladores de SOCE *STIM1* y *SEPTIN1* y *6*, y los canales *TRPC7*, *TRPM1* y *2*, *TRPP5* e *IP3R1*. Hay que destacar el previsible y remarcable aumento de la capacidad de captación de calcio mitocondrial en el fenotipo tumoral. Estos cambios difieren de forma notable de los encontrados en los modelos celulares. En concreto, respecto de los genes relacionados con la homeostasis del calcio intracelular diferencialmente expresados en los modelos celulares (56) y en los tejidos (30), únicamente coincide la expresión diferencial de 16 de ellos: sobreexpresión de los genes *SEPTIN2,7,11*, *TRPP1*, *SECA2*, *SPCA2*, *IP3R3*, *MCU*, *MICU1*, *MCUR1* y *VDAC2*, y disminución de los genes *SEPTIN1,6*, *TRPC7*, *TRPM2*, *TRPP5* en el fenotipo tumoral.
 6. El análisis de la expresión de proteínas en las muestras pareadas de tejido humano de colon normal y tumoral reveló un sorprendente dimorfismo sexual en la expresión de *ORAI2*, mayor en las mujeres, pero no en *ORAI1*. Además, las muestras tumorales mostraban una disminución de la expresión de *ORAI2*, pero no de *ORAI1*. Los niveles de expresión de las proteínas *STIM1* y *STIM2* eran similares en ambos sexos. Sin embargo, las muestras tumorales mostraban una menor expresión de *STIM1*, aunque solo en varones, y una mayor expresión de *STIM2*, aunque solo en mujeres, que el tejido normal. La expresión de *TRPC1* era similar en varones y mujeres y no cambiaba, en promedio, en las muestras tumorales. Sin embargo, para algunos pacientes no se observaba expresión alguna en el tejido normal mientras que en tejido tumoral este canal se expresaba abundantemente.
 7. El metaanálisis de microarrays reveló una elevada alteración de diferentes conjuntos de genes (entre 3000 y 9000) relacionados con el cáncer de colon. Entre estos se observa la alteración de conjuntos relacionados tanto con el cáncer de colon (señalización Wnt y poliaminas) como con la homeostasis del calcio intracelular de la mitocondria y el retículo endoplasmático.

8. El metaanálisis indica la sobreexpresión de 16 genes en el tejido tumoral que codifican para los canales *CAVI.3*, *ORAI1,2* y *3* (y sus moduladores *CRACR2A* y *SEPTIN2,4,7,9*, *TRPC7*, *TRPV4*, *TRPP3*, la bomba *SPCA1*, el canal de liberación de calcio *IP3R2*, el modulador del transporte mitocondrial *MCUb* y el canal de la membrana externa mitocondrial *VDAC1*, junto con la infraexpresión de 41 genes incluyendo los de los canales *CAV2.1*, *CAV3.2*, el canal específico de tejido *MS4A12*, los moduladores de SOCE *STIMI*, *SARAF*, *STIMATE*, *MBP* y las *SEPTIN1,6,8,10* y *11*, los canales *TRPC1* y *6*, *TRPV1* y *3*, *TRPM1,4-8*, *TRPP1*, *TRPML1* y *2*, las bombas *PMCA1,3,4*, *SERCA2* y *3*, *SPCA2*, el transportador *NCX2*, los canales intracelulares *IP3R1* y *RyR3* así como los transportadores mitocondriales *MCU*, *EMRE*, *MICU1,2,3*, *VDAC2* y *NCLX*.
9. El patrón de genes diferencialmente expresados encontrado en el metaanálisis difiere de los cambios en los modelos celulares aunque coinciden en 19 genes incluyendo *CAVI.3*, *CAV3.2*, *ORAI2*, *SARAF*, las *SEPTIN1,2,4,6,7,9*, *TRPV1*, *TRPML1,2*, *TRPM5*, *TRPM8*, *TRPA1*, *PMCA4*, *MICU2* y *NCLX*. Respecto a las muestras pareadas, los datos del metaanálisis coinciden en solo 10 genes incluyendo el canal específico de tejido *MS4A12*, los moduladores de SOCE *STIMI* y las *SEPTIN1,2,6* y *7*, los canales *TRPM1*, la bomba *SPCA1*, el canal del retículo *IP3R1* y el canal mitocondrial *VDAC1*. Finalmente, en los tres análisis llevados a cabo solo coinciden 4 genes, las *SEPTIN1,2,6* y *7*.
10. El perfil de expresión transcriptómico revela una elevada severidad del tejido tumoral estudiado en el metaanálisis, pues hay evidencias de que el proceso tumoral está favorecido (disminución de *TRPV1* y *TRPM1*) y que podría tener lugar el aumento de la autonomía proliferativa (aumento de *VDAC1*, *SPCA1*), el aumento de su capacidad metastásica (disminución de *NCLX*, *TRPM1* y *TRPA1*), una mayor resistencia a la muerte (aumento de *TRPV4* y *CAVI.3*), pérdida de diferenciación (disminución de *MS4A12*), peor pronóstico (disminución de *MS4A12* y aumento de *ORAI3*), y la resistencia a tratamientos quimioterápicos como la doxorrubicina (disminución de *TRPM6,7* y *NCLX*).

Parte VII

Apéndices

“Hay un granjero que tiene gallinas, pero no ponen huevos, así que pide a un físico que le ayude. El físico realiza unos cuantos cálculos y le dice: -¡Tengo la solución!, pero sólo funciona con gallinas esféricas y en el vacío”.

Leonard Hofstadter, serie “The Big Bang Theory”

Apéndice A

Fundamentos básicos para el análisis de datos

En la presente tesis se han usado diferentes técnicas y metodologías de análisis de datos muchas de las cuales no son de uso común. Por ello, con el fin de no entorpecer la lectura del resto de capítulos, se ha creído conveniente incluir este apéndice para exponer brevemente algunas de ellas así como ciertos conceptos y terminología que se han empleado a lo largo de la presente tesis. Algunos de estos conceptos, en especial los modelos lineales, son la base de los métodos empleados en el análisis de los datos de *RNA-seq* y *microarrays*. No obstante, puesto que los métodos específicos de cada una de estas tecnologías transcriptómicas no serán explicados al detalle en este apéndice debido a su complejidad y a la extensión que ello supondría, se recomendarán diferentes textos para el lector que desee profundizar en ellos.

A.1. Estadística descriptiva

Antes de comenzar con cualquier análisis, es necesario entender ciertos aspectos sobre la variable o variables que se van a estudiar tales como qué tipo de variable se manejan, cómo se distribuyen, como se relacionan entre sí... Por ello, el primer paso es aplicar la estadística descriptiva la cual se caracteriza por no hacer suposiciones sobre la variable estudiada. No obstante, las técnicas de estadística descriptiva varían según el tipo de variable que se está analizando. Una de las formas más habituales de clasificar las variables es la propuesta por Stevens and others (1946):

- *Variables no métricas*: son aquellas de naturaleza *cualitativa* y, por ello, es necesario codificarlas como números para poder aplicar muchas de las técnicas de análisis de datos. Dentro de esta categoría pueden distinguir dos tipos:
 - *Variables de escala nominal*: son aquellas cuyas categorías no tienen un orden (por ejemplo, el color de los ojos), son exhaustivas (la observación ha de pertenecer a una de las categorías) y mutuamente excluyentes (pero sólo a una). En este tipo de variables pueden usar estadísticos como la *moda* o la *frecuencia* y pueden representarse mediante tablas de contingencia. Respecto de la codificación numérica, puesto que no existe un orden entre sus categorías, estas se codifican como *dummy variable*: se convierte la variable original con k categorías en $k - 1$ (a veces k) variables binarias (0 o 1); es decir, que cada una de las nuevas k variables tienen todos sus valores igual a 0 excepto los correspondientes a las observaciones que tienen la categoría a la que representan, que tendrán el valor 1.
 - *Variables de escala ordinal*: son aquellas cuyas categorías tienen una relación de *orden total* (por ejemplo delgado, normopeso y obeso). A parte de poder usar los mismos estadísticos que para las variables de escala nominal, también se pueden emplear la *mediana* y los *percentiles*. En cuanto a su codificación numérica, es posible asignar un número a cada una de las categorías de tal manera que se mantenga el orden de la variable original.
- *Variables métricas*: estas variables, cuyos valores son numéricos, pueden dividirse en continuas (entre dos valores cualquiera hay infinitos valores) o discretas (hay dos valores diferentes entre los cuales no hay ningún valor) y, además entre datos agrupados o no. En este tipo de variables es posible diferenciar entre datos agrupados o sin agrupar. Además, también pueden agruparse en:
 - *Escala intervalo*: a parte de los estadísticos que se emplean en las variables no métricas, también se pueden aplicar diferencias entre sus valores. Por ejemplo, si se mide la temperatura en °C, la diferencia entre 5 °C y 10 °C es de 5 °C y es la misma que entre 50 °C y 55 °C. Sin embargo, en este tipo de variables el 0 no es absoluto y, por ende, 50 °C no son el doble que 25 °C.
 - *Escala de razón*: en estas variables el 0 si tiene un significado absoluto y, para asemejarlo al caso anterior con la medición de la temperatura, la variable temperatura con escala de razón es la temperatura medida en K, puesto que en esta escala el 0 es el valor mínimo y, por ello, 50 K es el doble de 25 K.

A.1.1. Estadística descriptiva univariante

Cuando se está interesado en una variable, el primer paso es conocerla y, para ello, se emplean medidas descriptivas univariantes. No obstante, hay que tener especial cuidado cuando la variable no es homogénea, como por ejemplo en el caso de la variable *altura de las personas de un país*, puesto que, en realidad, sus datos no son homogéneos ya que provienen de dos poblaciones: hombres y mujeres. A continuación, se presentan las medidas descriptivas univariantes más usadas:

- *Medidas de centralización*: son aquellas que dan información sobre el valor alrededor del cual se localizan el resto de datos, como por ejemplo la media -variables cuantitativas- o frecuencia -cualitativas-.
- *Medidas de dispersión*: dan idea de cómo los datos de la variable se alejan o se acercan entre sí, como por ejemplo la desviación estándar o el coeficientes de variación.
- *Medidas de asimetría*: miden cuán simétrica es la distribución de la variable, como por ejemplo el coeficiente de asimetría (ecuación A.1):

$$A_j = \frac{1}{n} \frac{\sum_i^n (x_{ij} - \bar{x}_j)^3}{s_j^3} \quad (\text{A.1})$$

donde j es la variable, n las observaciones e $i \in \{1, 2, \dots, n\}$. Así, cuando la distribución es simétrica, $A_j = 0$, cuando tiene cola por la derecha $A_j > 0$ y por la izquierda si $A_j < 0$. No obstante, puede ser que una distribución asimétrica de lugar a $A_j = 0$. Además, también es interesante contrastar la *hipótesis nula de simetría*; es decir, que la variable ha sido generada por una distribución simétrica, para lo que se emplea el estadístico $G_j \sim N(0, \text{Var}(G_j))$ o a $G_j/(\sqrt{6/n}) \sim N(0, 1)$ con $\text{Var}(G_j) \approx 6/n$ si $n > 50$ (Peña (2002a))(ecuación A.2).

$$G_j = \frac{\sqrt{n(n-1)}}{n-2} A_j ; \quad \text{Var}(G_j) = \frac{6n(n-1)}{(n-2)(n+1)(n+3)} \quad (\text{A.2})$$

- *Medidas de homogeneidad*: proporcionan información sobre la generación de los datos en el sentido de si proceden de la misma población -homogéneos-, si hay presencia de datos atípicos o si han sido generados por diferentes distribuciones -heterogéneos-, en cuyo último caso habrá que estudiar por separado las diferentes poblaciones. Para medir la homogeneidad los coeficientes más usados son el *coeficiente de kurtosis* y el de *variación*. Así, a partir del *coeficiente de kurtosis* para la variable j , K_j , se tiene el *coeficiente de homogeneidad* (ecuación A.3):

$$H_j = \frac{1}{n} \frac{\sum_i^n (x_{ij} - \bar{x}_j)^4}{s_j^4} - 1 = K_j - 1 \geq 0 \quad (\text{A.3})$$

Al igual que con la simetría, también es interesante contrastar la *hipótesis nula de población homogénea*, para lo que se emplea el estadístico $K_j \sim N(3, 24/n)$ para $n > 200$. El coeficiente de kurtosis es muy interesante para evaluar la posible presencia de datos atípicos ($K_j > 6$) o mixtura de poblaciones ($K_j < 2$). Otra medida descriptiva muy útil para evaluar la homogeneidad es el *coeficiente de variación (CV)* que da idea del porcentaje del error de medición respecto de la media de la variable j (ecuación A.4)

$$CV_j = \frac{s}{|\bar{x}|} \quad (\text{A.4})$$

de tal manera que cuando los datos son homogéneos $CV < 1.5$, por lo que valores superiores son indicativo de la necesidad de evaluar fuentes de variación no consideradas que estén haciendo que los datos sean heterogéneos, como por ejemplo no considerar el sexo y que este si que tenga efecto sobre la variable en cuestión.

Ante la presencia casi inevitable de heterogeneidad o de valores atípicos es conveniente, en muchas ocasiones (no en todas) recurrir al uso de estadísticos descriptivos más robustos los cuales son menos sensibles a la presencia de datos atípicos. Entre ellos, están la *mediana* como medida de centralización, o la *desviación absoluta mediana (MAD o MEDA)* como medida de dispersión. No obstante, muchas veces es aconsejable, sobretodo cuando se hacen análisis exploratorios de muchas variables, comparar los estadísticos descriptivos observados clásicos con los robustos.

- *Normalidad*: de sobra es conocida la importancia de la *distribución normal*, pero lo que no se suele saber es que el culpable de ello son algunos teoremas como el *Teorema Central del Límite* el cual dice que cuando se tienen n variables aleatorias independientes de media μ_i y varianza σ_i^2 de una distribución cualquiera y que, además, la distribución de cada una de las n variables aleatorias puede ser diferente, la suma de todas ellas, $Z \sim N\left(\sum \mu_i, \sqrt{\sum \sigma_i^2}\right)$. Así, cuando la variable aleatoria j observada es el resultado de sumar muchas variables independientes entonces la distribución de dicha variable se aproximará a una normal.

A.1.2. Análisis descriptivo multivariante

Lo anterior es ideal cuando sólo se tiene una variable, lo cual no suele ser muy habitual. Lo habitual suele ser tener numerosas variables, sobretodo en técnicas transcriptómicas de alto rendimiento, pues, precisamente, *alto rendimiento* se refiere a que son capaces de medir muchos genes (variables) simultáneamente. Por ello, a parte de estudiar cada una de las variables de forma aislada, también es conveniente estudiarlas en su conjunto. Es más, muchas técnicas de análisis que se usarán a lo largo de la tesis se basan en técnicas multivariantes. Entre las diferentes medidas descriptivas multivariantes que existen, las más comunes son:

- *Medidas de Centralización Multivariante:* las más comunes son *vector de medias* y el *medianas*:

- *Vector de medias:* esta medida proporciona el valor medio de cada una de las variables y lo almacena en un vector (ecuación A.5).

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \mathbf{X}' \mathbf{1} = (\bar{x}_1, \dots, \bar{x}_p)' \quad (\text{A.5})$$

- *Vector de medianas:* en los datos multivariantes, debido a la falta de un orden natural de estos, el vector de medianas no tiene porqué coincidir con el centro de gravedad (vector de medias).

- *Medidas de variabilidad:* cuando se tiene una variable, para medir la dispersión de los datos se usa la varianza. No obstante, cuando se tiene más de una variable para cada observación es posible medir no solo dicha variabilidad para cada una de las variables, $Var(j)$, sino también la covarianza entre ellas la cual da una idea de la relación lineal entre ellas, $Cov(k,j)$ la cual es simétrica. Es más, la varianza es un caso especial de la covarianza $Cov(x_k, x_j)$ para $j = k$. En particular, a partir de estos valores para cada una de las variables es posible obtener la matriz de varianzas y covarianzas S , la cual es cuadrada (de orden $p \times p$), simétrica ($Cov(x_k, x_j) = Cov(x_j, x_k)$) y semidefinida positiva. Así, en las posiciones kj o jk están los valores $Cov(k,j)$ y en las posiciones de la diagonal principal jj $Cov(x_k, x_k) = Var(x_k)$ (ecuación A.6).

$$\begin{aligned} \mathbf{S} &= \frac{1}{n} \sum_i^n (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})' = \frac{1}{n} \tilde{\mathbf{X}}' \tilde{\mathbf{X}} = \\ &= \begin{pmatrix} Var(1) & \dots & Cov(1p) \\ \vdots & \dots & \vdots \\ Cov(p1) & \dots & Var(p) \end{pmatrix} = \begin{pmatrix} s_1^2 & \dots & s_{1p} \\ \vdots & \dots & \vdots \\ s_{p1} & \dots & s_p^2 \end{pmatrix} \end{aligned} \quad (\text{A.6})$$

donde la matrix $\tilde{\mathbf{X}}$ es la matriz de datos centrados (a cada valor se le resta el valor medio de la variable. La matriz anterior tiene, a su vez, dos formas: sesgada \mathbf{S} , cuando en el cálculo

se divide entre n , o corregida \hat{S} , cuando se divide entre los grados de libertad $n - 1$. Es más, de las dos, la forma corregida se considera un estimador insesgado de la matriz de varianzas-covarianzas poblacional. Esta matriz es muy informativa sobretodo en un análisis exploratorio, puesto que:

- Si una variable escalar tiene una varianza elevada entonces sus valores están muy dispersos y, por ende, la información que aporta cada una de las observaciones de dicha variable es superior a aquella aportada por otras variables escalares con una menor varianza.
 - Si la covarianza procede de variables cuyos valores no han sido ni centrados ni divididos por su desviación típica, valores altos (en valor absoluto) de covarianza entre las variables k y j indican un elevado grado de asociación lineal entre ellas. Por lo tanto, proporcionan una medida del grado de relación lineal entre cada par de variables. De hecho, una elevada covarianza es indicativo de que ambas variables están aportando la misma información, por lo que son variables redundantes y dependiendo del tipo de análisis (modelos lineales por ejemplo) es necesario eliminar todas menos una.
- *Medidas de dependencia lineal:* la matriz de covarianzas anterior proporciona una información muy interesante. Sin embargo, los valores dependen de las unidades de cada una de las variables escalares y, por ello, no hay un valor a partir del cual considerar que hay mucha o poca asociación o que la dispersión es muy elevada o no. Así, parece lógico el interés por encontrar otras medidas que den una idea más clara sobre cuán asociadas están dos variables. Entre las diferentes medidas que existen, la más común es el coeficiente de *correlación de Pearson* que no es más que la $Cov(kj)$ cuando los valores de las variables escalares k y j han sido centradas (se les ha restado a cada una de ellas su media) y se han escalado (los valores de cada una de ellas se han dividido entre su desviación típica), de tal manera que ahora cada una de estas nuevas variables tienen media 0 y varianza igual a 1. Así, para cada par de variables se obtiene el siguiente valor de correlación lineal (ecuación A.7):

$$r_{j,k} = \frac{s_{j,k}}{s_j s_k} \quad (\text{A.7})$$

siendo $s_{j,k}$, s_j y s_k elementos de la matriz de varianzas y covarianzas. Como resultado, se tiene un valor $r_{j,k}$ que es adimensional, invariante frente a transformaciones lineales (como cambios de escala) y que toma valores entre -1 y $+1$ de tal manera que cuanto más cercano sea $r_{j,k}$ a uno de los valores extremos mayor será la dependencia lineal. Así, si en vez de construir una matriz cuyos valores sean $Cov(jk)$ ($S_{j,k}^2$) sean $r_{j,k}$ se construye la matriz de correlaciones \mathbf{R} , sumamente interesante para medir la correlación o dependencia dos a dos entre todas las variables escalares de la variable vectorial \mathbf{X} . Es más, esta matriz es la matriz de *varianzas-covarianzas* de los datos centrados y estandarizados, por lo que sus propiedades

son idénticas a las de la matriz \mathbf{S} ; es decir, es una matriz *cuadrada, simétrica y semidefinida positiva* (ecuación A.8).

$$\mathbf{R} = \begin{pmatrix} 1 & \cdots & r_{1,p} \\ \vdots & \ddots & \vdots \\ r_{p,1} & \cdots & 1 \end{pmatrix} \quad (\text{A.8})$$

Cabe destacar que la matriz anterior se puede obtener calculando los valores $r_{j,k}$ a partir de la matriz \mathbf{S} o tipificando la matriz de datos originales (ecuación A.9: a los datos de la variable j se les resta su media y se les divide entre la desviación estándar) y calculando la matriz \mathbf{S} de la matriz de datos tipificados, puesto que en tal caso $\mathbf{S} = \mathbf{R}$.

$$\mathbf{X}_{\text{tipificados}} = \begin{pmatrix} \frac{x_{11}-\bar{x}_1}{s_1} & \cdots & \frac{x_{1p}-\bar{x}_p}{s_p} \\ \vdots & \ddots & \vdots \\ \frac{x_{n1}-\bar{x}_1}{s_1} & \cdots & \frac{x_{np}-\bar{x}_p}{s_p} \end{pmatrix} \quad (\text{A.9})$$

de tal manera que ahora cada variable j tiene media $\bar{x}_j = 0$ y desviación estándar $s_j = 1$, por lo que (ecuación A.10):

$$r_{j,k} = \frac{s_{j,k}}{s_j s_k} = \frac{s_{j,k}}{1 \cdot 1} = s_{j,k} \quad (\text{A.10})$$

y, por ende, $\mathbf{R} = \mathbf{S}$.

- *Distancias*: todo el mundo tiene la idea intuitiva de una distancia en el espacio vectorial euclídeo \mathbb{R}^3 y, a partir de esta idea básica se puede comprender más fácilmente el concepto de distancia entre las diferentes observaciones de la matriz de datos, donde en vez de haber $p = 3$ espacios hay un p variables que generan el espacio vectorial real \mathbb{R}^p . Es más, aunque a primera vista no lo pueda parecer, la varianza es una medida de distancia para las observaciones de la variable escalar x ; en particular si la distancia de la observación i a la media de las observaciones es $d = \sqrt{(x_i - \bar{x})^2}$, el promedio de estas distancias es la varianza de la variable en cuestión. Por lo tanto, las distancias también se suelen emplear como medidas de variabilidad. Algunas de las diferentes distancias existentes, destacan:

- *Distancia de Minkowski*: esta es una de las familias de distancias más empleadas en el espacio vectorial real \mathbb{R}^p (ecuación A.11):

$$d_{\text{Min},iw} = \left(\sum_{j=1}^p |x_{ij} - x_{wj}| \right)^{1/r} \quad (\text{A.11})$$

a partir de la cual, variando el parámetro r es posible obtener distancias muy usadas en el análisis de datos, como por ejemplo la distancia *Euclídea*, $r = 2$ o la de *Manhattan*

($r = 1$). La *Distancia Euclídea* es una de las medidas de distancia más empleadas para variables cualitativas; no obstante, depende fuertemente de las unidades de medida de las diferentes variables, puesto que tendrán más peso en el valor de la distancia aquellas cuyas unidades sean mayores y, por ende, dominarán el valor de la distancia. Por lo tanto, a no ser que todas las variables tengan las mismas unidades, no tiene sentido su uso o si los datos han sido previamente transformados (normalizados, estandarizados...) como suele hacerse en el análisis de componentes principales, en la aplicación de redes neuronales artificiales... Sin embargo, esta estrategia podría esconder la variabilidad original de los datos y, en ocasiones, como en la detección de *outliers*, no suele interesar. Otra desventaja de esta distancia es que no considera las correlaciones entre las diferentes variables (por ejemplo, la que puede haber entre el peso y la altura), lo que conlleva una distorsión espacial.

Por ello, existen otras medidas de distancia que solventan estos problemas, como por ejemplo la distancia de Mahalanobis, ampliamente usada (o versiones robustas de esta) en diferentes metodologías como por ejemplo la detección de *outliers*. Esta medida de distancia, a diferencia de las anteriores, es invariante a cambios de escala y tiene en cuenta las posibles correlaciones entre las diferentes variables (ecuación A.12).

$$d_{M,jk} = (\mathbf{x}_j - \mathbf{x}_k)' \mathbf{S}^{-1} (\mathbf{x}_j - \mathbf{x}_k) \quad (\text{A.12})$$

Existen diferentes versiones de esta distancia. Por un lado, están las versiones clásicas en las que la informaciónn obtenida depende de sobre quién se considere la distancia. Por ejemplo, si se desea conocer la distancia de cualquier observación respecto del vector de medias o, por el contrario, entre dos poblaciones distintas a partir de los vectores de medias de cada una de ellas y como matriz \mathbf{S}^{-1} la media ponderada de las correspondientes matrices de varianzas y covarianzas de ambas poblaciones. Además, si se promedian las distancias de *Mahalanobis*, el promedio es p , es decir, el número de variables. Así, es evidente la complejidad de esta medida de distancia en comparación con las anteriores. Además, es importante tener en cuenta que esta es capaz de controlar el efecto que tiene que diferentes variables tengan diferentes escalas y unidades de medida o que estén más o menos correlacionadas entre sí; es más, esta medida de distancia actúa en un espacio invariante en el cual las correlaciones entre las diferentes variables están corregidas de tal manera que a mayor correlación entre dos variables menor será su distancia.

A.2. Variables compuestas o redundantes

A partir de las variables de la matriz de datos \mathbf{X} es posible crear nuevas variables, *variables compuesta* Y , que sean combinación lineal de las estas cuyos coeficientes sean $\mathbf{a} = (a_1, a_2, \dots, a_p)$ (ecuación A.13):

$$Y = \mathbf{X}\mathbf{a} = a_1X_1 + a_2X_2 + \dots + a_pX_p \quad (\text{A.13})$$

Algunos métodos, como por ejemplo en el análisis de componentes principales, se basan en crear variables compuestas Y las cuales reciben el nombre de *componentes principales (PC)* (ver apartado “A.3 Análisis de componentes principales”). No obstante, si en la matriz de datos \mathbf{X} hay variables r compuestas entonces existe información repetida que no aporta información nueva en dicha matriz, de ahí que a estas variables compuestas también se las conozca como *variables redundantes*. La presencia de estas puede dar lugar a problemas tales como que \mathbf{X} no sea de rango máximo, sino de rango h por contener $r = p - h$ variables redundantes. En consecuencia, \mathbf{S} tampoco lo será y, por ende, será singular -determinante igual a cero y no es invertible- (ecuación A.14).

$$\text{rg}(\mathbf{S}) = \text{rg}(\tilde{\mathbf{X}}) = \text{rg}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}) = \text{rg}(\tilde{\mathbf{X}}\tilde{\mathbf{X}}') \quad (\text{A.14})$$

El rango de una matriz simétrica, como es \mathbf{S} , es igual al número de valores propios diferentes de cero, siendo $r = p - h$ el número de valores propios nulos (hay que recordar que una matriz simétrica puede diagonalizarse, cuyos miembros de la diagonal serán los valores propios y el valor del determinante de ambas, que ha de ser igual, puede obtenerse multiplicando estos; al ser singular, el determinante es 0, por lo que hay valores propios iguales a 0). Afortunadamente, los autovectores asociados al valor 0 muestran, a partir de sus elementos, qué combinaciones lineales dan lugar a las r variables redundantes. Además, hay ocasiones en las que, debido a redondeos, una variable no es exactamente una combinación lineal del resto, por lo que da lugar a un valor propio diferente de cero pero muy cercano a él, lo que indica la existencia de una relación aproximadamente lineal entre esta variable y el resto.

Cuando aparecen este tipo de variables, ya sean aquellas que dan lugar a un autovalor igual a cero o cercano a él, es conveniente eliminarlas de la matriz de datos para emplear diferentes métodos para evitar posibles problemas derivados de esta situación. Por ejemplo, en regresión multivariante cuando hay variables que son combinaciones lineales perfectas en la matriz de datos se dice que hay *multicolinealidad perfecta*, lo que provoca que la matriz $\mathbf{X}'\mathbf{X}$ sea singular y, por ende, no invertible, lo que supone un problema ya que es necesario para el algoritmo que usa obtener su inversa. No obstante, suele presentarse con demasiada frecuencia, sobretodo cuando se tiene un número demasiado elevado de variables explicativas, que una variable no sea exactamente combinación lineal del resto pero que de lugar a unos valores propios muy cercanos a cero para la matriz \mathbf{S} , lo que recibe el nombre de *multicolinealidad no perfecta*, que también puede dar lugar a diferentes problemas a la hora de generar los modelos.

A.3. Análisis de componentes principales

En el análisis multivariante, como puede ser el análisis de los datos de experimentos transcriptómicos de alto rendimiento, hay un elevado número de variables -genes, sondas...-, lo que se conoce como datos de gran dimensión. Por ello, suele ser conveniente para ciertos análisis reducir la dimensionalidad de los mismos. Además, las variables originales suelen mostrar, en mayor o menor grado, correlación entre ellas, a veces demasiado elevada, por lo que es habitual encontrar variables redundantes -no aportan información adicional- y que, por ende, podrían ser eliminadas, puesto que en muchos análisis estas pueden ser perjudiciales.

Una posibilidad para reducir la dimensión de los datos sería escoger un pequeño subgrupo de variables, dos o tres de entre todas ellas; no obstante, a parte de que se podría estar cometiendo un considerable error, cuando se tratan datos con miles de variables no es tarea sencilla escoger 3 de ellas. Así, una de las metodologías más empleadas en la reducción de la dimensión es la que se conoce como *análisis de componentes principales (ACP)*, en la que se construyen una serie de nuevas variables, tantas como variables originales, a partir de combinaciones lineales de las p variables originales (ecuación A.15):

$$u_{ki} = a_{k1} \cdot x_{i1} + a_{k2} \cdot x_{i2} + \dots + a_{kp} \cdot x_{ip} \quad (\text{A.15})$$

de tal manera que u_{kj} es la proyección de los datos para la observación i sobre la nueva variable k , como se conocen normalmente, *Componente Principal (PC)* PC_k definida por los coeficientes $\{a_{k1}, \dots, a_{kp}\}$ contengan la mayor variabilidad posible con el fin de perder la menor cantidad de información tras reducir la dimensión y, además, que estén incorrelacionadas entre sí para que cada una aporte una información diferente y no redundante. Por lo tanto, está claro la importancia de cómo hallar dichos coeficientes. Así, antes de comenzar con la obtención de estos conviene recordar la estructura de la matriz de varianzas-covarianzas (\mathbf{S}) o la matriz de correlaciones \mathbf{R} , las cuales contienen en su diagonal principal los términos que indican la variabilidad para \mathbf{S} de cada una de las variables o la correlación consigo mismas para \mathbf{R} (diagonal de unos en \mathbf{R}); además, también se ha visto que cuando los datos se tipifican, entonces $\mathbf{S} = \mathbf{R}$.

De modo muy resumido, el ACP consiste en encontrar una base del espacio vectorial real tal que la matriz de covarianzas \mathbf{S} en dicha base sea diagonal, es decir, que las nuevas variables, llamadas componentes principales, estén incorrelacionadas entre sí. Para evitar que el análisis esté dominado por aquellas variables que tienen una mayor escala (unidades más grandes), conviene partir no de la matriz de covarianzas o de los datos originales, sino de la matriz de correlaciones o de los datos tipificados. Por lo tanto, puesto que la matriz de covarianzas (o de correlaciones) es una matriz cuadrada (tantas filas y columnas como p variables) y simétrica entonces es diagonalizable puesto que se puede descomponer en el producto de la ecuación A.16), donde \mathbf{D} es una matriz diagonal que resulta ser la matriz de covarianzas en el nuevo espacio:

$$\mathbf{S} = \mathbf{PDP}^{-1} \quad (\text{A.16})$$

donde \mathbf{P} es la matriz de cambio de base, \mathbf{P} , cuyas columnas son las coordenadas de los vectores propios respecto de la base de \mathbf{S} y, además, dichos vectores propios forman una base ortogonal del espacio vectorial \mathbb{R}^p en la que la matriz de covarianzas \mathbf{S} pasa a ser diagonal, \mathbf{D} , donde los elementos de la diagonal son los p valores propios λ_i asociados a los p vectores propios \mathbf{P} . Por lo tanto, los datos originales \mathbf{X} se proyectan sobre la nueva base mediante la matriz de cambio de base \mathbf{P} , o sea, que si la nueva matriz de datos proyectada sobre la nueva base se llama \mathbf{U} , entonces se obtiene como $\mathbf{U} = \mathbf{XP}$. Por lo tanto, las columnas de la matriz \mathbf{P} son las PC.

Además, puesto que \mathbf{S} y \mathbf{D} son semejantes estas tienen la misma traza -la suma de sus elementos de la diagonal- y, puesto que esta representa la variabilidad total de los datos o cantidad de información, tanto \mathbf{S} como \mathbf{D} tienen la misma variabilidad. No obstante, mientras que en \mathbf{S} cada elemento de la diagonal es igual a 1 -se usa la matriz de correlaciones para visualizarlo más fácil- cada una de las p variables originales retiene un $1/p \cdot 100$ % de la variabilidad total, en \mathbf{D} sus elementos de la diagonal principal no son unos, sino $\{\lambda_1, \lambda_2, \dots, \lambda_i, \dots, \lambda_p\}$ (ordenados de mayor a menor) y, por ende, cada una de las nuevas variables PC_i retiene una cantidad diferente de variabilidad determinada por el correspondiente valor propio λ_i al que esté asociada cada una, $\lambda_i / \sum_{i=1}^p \lambda_i \cdot 100$.

Por último, recordar que, por normal general, las PC se nombran de la 1 hasta p en orden decreciente de valor propio al que estén asociados, es decir, que la CP_1 tiene el valor propio λ_1 que es mayor al resto de valores propios, y así sucesivamente. Por lo tanto, la CP_1 es la nueva variable o PC que más variabilidad o información explica, la PC_2 la siguiente que más explica...

Por lo tanto, siguiendo este procedimiento es posible reducir la dimensión de los datos de p variables originales a $h < p$ PC que expliquen gran parte de esta variabilidad puesto que las primeras h PC explicarán más variabilidad que cualquier subconjunto de h variables originales. No obstante, todavía queda preguntarse cuántas PC seleccionar, para lo que existen diferentes criterios:

1. Criterio de Kaiser: elegir las primeras PC cuyos valores propios son superiores a 1.
2. Varianza explicada: elegir las PC que expliquen al menos el 70 u 80 % de la varianza total.
3. Scree plot de Cattell (1966): método gráfico en el que se representa en el eje X las PC en orden decreciente de varianza explicada y en el eje Y se representa la varianza explicada, y se eligen las componentes hasta el codo del gráfico. Existe un método automático propuesto por Zhu and Ghodsi (2006) para automatizar dicha selección basándose en este gráfico.
4. El criterio de la interpretabilidad: se ha de tener conocimiento previo sobre la materia estudiada, de tal manera que, en este caso, se eligen todas las componentes que mejor recojan la esencia del significado de las variables originales y que las PC tengan sentido en el contexto del problema investigado.
5. Autores como Jeffers (1967) sugieren que son suficientes sólo dos componentes para explicar la mayor parte de la varianza y otros como Jolliffe (1973) sugieren que este número hay que aumentarlo hasta cuatro o cinco PC.

No obstante, es evidente que cualquiera de las anteriores soluciones, en comparación con las decenas de miles de variables que hay en un experimento transcriptómico de alto rendimiento, la reducción de la dimensión es considerable. Por lo general, suelen bastar entre 2 y 5 componentes principales para acumular el 70 % de la variabilidad de los datos (aunque no siempre).

Una vez se tienen las componentes principales, en ocasiones es interesante poder dar una explicación a alguna de ellas. Por ejemplo, si se tienen dos poblaciones, tumoral y normal, y se realiza un experimento transcriptómico, es interesante encontrar qué componente principal es capaz de agrupar las diferentes muestras en cada una de las dos poblaciones. Así, una vez identificada dicha componente principal es posible analizar qué variables originales, genes, tienen más peso o influyen más en ella. Para ello, pueden emplearse dos estrategias. La primera es estudiar los coeficientes de la componente, a_i , de tal manera que tendrán más influencia aquellos genes cuyo coeficiente sea mayor en valor absoluto. Otra opción es estudiar la correlación entre cada una de las variables originales j y la componente principal PC_k en cuestión, $Cov(j, PC_k)$ y realizar un contraste de hipótesis sobre el parámetro poblacional de la correlación de Pearson ρ_{j,PC_k} , estimada, r_{j,PC_k} ; es decir, contrastar la hipótesis nula $H_0 = \rho_{j,PC_k} = 0$ frente a la alternativa $H_1 = \rho_{j,PC_k} \neq 0$ (ver apartado “A.5 Coeficiente de correlación de Pearson”).

A.4. Detección multivariante de outliers

la detección de datos atípicos es un campo bastante complejo en el que a menudo surgen nuevas metodologías o aplicaciones de otras ya existentes para llevar a cabo la detección de *outliers*. Hay criterios univariantes; es decir, tratar a cada una de las variables escalares por separado. No obstante, este enfoque no es eficiente a la hora de detectar datos atípicos de carácter multivariante. Esto se debe a que, por lo general, los datos atípicos multivariantes provienen de situaciones en los que el efecto sobre cada una de las variables es pequeño sobre todas las variables en lugar de mostrar un efecto notable en ellas. Entre los diferentes enfoques multivariantes, unos de los más usados son aquellos basados en la distancia de Mahalanobis de las observaciones al vector de medias.

$$d_M(\mathbf{x}_i, \bar{\mathbf{x}}) = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \quad (\text{A.17})$$

Así, es posible estimar la distancia de Mahalanobis mediante métodos clásicos, los cuales fallan ante la presencia de múltiples *outliers* y el efecto enmascarador que se produce al no tener todos los *outliers* unas distancias de Mahalanobis altas. Por ello, suelen emplearse métodos más robustos en los que se usan estimadores robustos del vector de medias \mathbf{x}_i y de la matriz de varianzas-covarianzas \mathbf{S}^{-1} , lo que se conoce como *distancia robusta de Mahalanobis*. Entre los diferentes métodos robustos para la estimación de estas distancias, uno de los más sencillos es el mostrado a continuación:

1. Eliminar de la muestra los datos sospechosos con el fin de evitar posibles enmascaramientos y obtener el vector de medias y la matriz de covarianzas sin la distorsión que estos pudieran llegar a ocasionar.
 - Buscar el 50 % de los datos con menor distancia de Mahalanobis al centro \bar{X} , media de todos los datos, bajo el supuesto de que el 50 % de los datos con menor distancia de Mahalanobis al centro no son datos atípicos.
 - Calcular la media “reducida” \bar{X}_R y la matriz de covarianzas reducida S_R , para ese conjunto reducido de datos.
2. Calcular las distancias de Mahalanobis d_M de todos los datos a la media \bar{X}_R con S_R y considerar como atípicos los datos tales que $d_M > p + 3p\sqrt{2p}$, donde p es el número de variables.

Aunque este método funciona considerablemente bien, existen otras estimaciones robustas más eficientes. Sin entrar en mucho detalle, los estimadores robustos más empleados son el del *Elipsoide de Mínimo Volumen (MVE)*, el del *Determinante de Mínima Covarianza (MCD)*, otros basados en S-estimadores y otros basados en algoritmos determinísticos (Ruppert (1992), Rocke (1996), Salibian-Barrera and Yohai (2006), Shieh and Hung (2009), Hubert, Rousseeuw, and Verdonck (2012)).

Por último, para decidir si una distancia es o no demasiado grande, partiendo de la base de que los datos siguen una distribución normal, entonces el cuadrado de las distancias robustas de Mahalanobis seguirá una χ_p^2 , por lo que el valor crítico por encima del cual se considerará una observación como un *outliers* será $\chi_{p,\alpha/2}^2$ donde $\alpha = 0.05$. Otra opción también puede ser considerar que la observación es atípica cuando su distancia sea superior a $p + 3p\sqrt{2p}$ (Shieh and Hung (2009), Peña (2002a)).

En particular, para la detección de *outliers* multivariantes en microarrays y en los datos de *RNA-seq* se ha empleado una metodología combinada en la que, en primer lugar, se calculan los componentes principales sobre todos los datos, posteriormente se dividen las muestras en función de la clase o tratamiento a la que pertenezcan y dentro de cada uno de los grupos se calculan las distancias de Mahalanobis robustas cuando el número de muestras en dicho grupo sea superior a $2p$ y la distancia de Mahalanobis clásica en el caso contrario.

A.5. Coeficiente de correlación de Pearson

En el apartado “A.1.2 Análisis descriptivo multivariante” se explicó cómo estimar el coeficiente de correlación de Pearson y se vio que, a diferencia del valor de la covarianza, el cual depende de las unidades de cada variable, el valor de la correlación es más intuitivo que el de la covarianza puesto que está acotado entre -1 y +1. No obstante, todavía queda por “encontrar” a partir de qué valor considerar que una correlación estimada es o no significativa; es decir, inferir si el valor observado del coeficiente de correlación diferente de cero es debido o no al azar. Para ello, se contrasta sobre el parámetro poblacional ρ_{jk} , el cual se estima mediante r_{jk} , la hipótesis nula $H_0 = \rho_{jk} = 0$ (no hay correlación) frente a la alternativa $H_1 = \rho_{jk} \neq 0$ (sí hay correlación) bajo los siguientes supuestos:

- La distribución en el muestreo de una función del coeficiente de correlación de Pearson (ecuación A.18) sigue una distribución *t de Students* con $n-2$ grados de libertad bajo la hipótesis nula.
- El par (k,j) sigue una distribución bivariada y, por ende, cada una de ellas por separado seguirá una distribución normal.
- Las variables j y k sean métricas -de intervalo o de razón-.

En consecuencia, el estadístico de contraste bajo la hipótesis nula es (ecuación A.18):

$$t = \frac{r_{jk}}{\text{sqrt} \frac{1-r_{jk}^2}{n-2}} \quad (\text{A.18})$$

Además, aunque también es posible obtener unos intervalos de confianza a partir de la expresión anterior, se suele usar otro método para ello. En concreto, el conocido como la *Transformación Z de Fisher*, en el que supone que Z (ecuación A.19):

$$Z_{jk} = \frac{1}{2} \ln \left(\frac{1 + r_{jk}}{1 - r_{jk}} \right) = \text{artanh}(r_{jk}) \quad (\text{A.19})$$

sigue una distribución normal $N(\text{artanh}(\rho_{jk}), 1/(\sqrt{n-3}))$. Por lo tanto, se tiene que el intervalo de confianza para el parámetro $\text{artanh}(\rho_{jk})$, para un nivel de significación α , es (ecuación A.20):

$$\text{artanh}(\rho_{jk}) \in \left[\text{artanh}(\rho_{jk}) \pm z_{\alpha/2} 1/(\sqrt{n-3}) \right] \quad (\text{A.20})$$

por lo que el intervalo de confianza para el parámetro de interés, ρ_{jk} , es (ecuación A.21):

$$\rho_{jk} \in \left[\tanh(\text{artanh}(\rho_{jk}) \pm z_{\alpha/2} 1/(\sqrt{n-3})) \right] \quad (\text{A.21})$$

A.6. Estudios, diseño de experimentos y modelos

En la investigación, cuando un investigador se plantea una serie de preguntas, y tras la adecuada búsqueda bibliográfica de la evidencia existente, fija una serie de objetivos con el fin de dar respuesta a esta. Para ello, es necesario diseñar el estudio que se llevará a cabo. No obstante, existen diferentes tipos de estudios y clasificaciones de estos, por lo que la elección del mismo es una de las etapas determinantes de este. Así, un estudio puede definirse o caracterizarse en función de si el investigador interviene o no sobre la población en estudio, de la finalidad del estudio, si existe un factor temporal en la toma de mediciones... En primer lugar, en función de si el investigador va a intervenir o no en el desarrollo del estudio este puede ser *experimental* u *observacional* (Manterola et al. (2019), Oehlert (2010), Vásquez and Pérez (2016)).

Por un lado, en los estudios observacionales el investigador se limita meramente a registrar una serie de eventos de la población estudiada sobre la cual no tiene ningún control, sin que el investigador manipule ninguna de las variables. Por otro lado, en los estudios experimentales el investigador (experimentador) interviene en el estudio para evaluar el efecto de dicha intervención (factor de interés, como un tratamiento) sobre la población estudiada. Dentro de los estudios observacionales, se pueden clasificar en función de la finalidad del estudio: *descriptivo* en los que la finalidad no es encontrar una relación causa-efecto sino describir los datos, o *analítico* en los que el interés es buscar, encontrar y evaluar una relación causa-efecto (Manterola et al. (2019), Vásquez and Pérez (2016)).

Los estudios también pueden clasificarse en función de la existencia o no de un factor temporal; es decir, si las mediciones se realizan a lo largo de un periodo temporal, *longitudinales*, o se realiza una única medición en un momento en concreto, *transversales*. En el caso de estudios observacionales, considerando el factor temporal pueden distinguirse, por ejemplo, *estudios de corte transversal* (transversal) o de *cohortes* (longitudinal), los cuales hay que tener en cuenta que pueden presentar sesgos al no haber tenido control o influencia sobre la población en estudio. En cuanto a los estudios experimentales, en función del tiempo pueden distinguirse, por ejemplo, estudios *longitudinales*, *ensayos clínicos* o estudios *transversales* en los que a diferencia de los anteriores, la población está bajo control y si se ha diseñado adecuadamente el experimento, no habrá sesgos (Manterola et al. (2019), Vásquez and Pérez (2016)).

Es obvio que hay una infinidad de tipos de estudios que pueden emplearse para contestar a las posibles preguntas que se plantea el investigador. Sin embargo, un investigador podría preguntarse por qué realizar un estudio experimental, con todo lo que ello conlleva, en lugar de uno observacional que pueda contestar a la misma pregunta y resultar menos tedioso y barato (no siempre). Para ello, aunque normalmente se da por hecho el conocimiento del término experimento por estar interiorizado en la mayoría de los investigadores, es conveniente saber qué lo caracteriza. En concreto, un experimento se caracteriza por los tratamientos y las unidades experimentales usadas, cómo se asignan los tratamientos a las diferentes unidades experimentales y qué respuestas son medidas.

No obstante, aunque la anterior caracterización es adecuada, todavía queda por definir adecuadamente otros términos usados en esta, como tratamientos o unidad experimental, puesto que se usarán a lo largo de la tesis y, por ello, es conveniente concretar el significado de cada uno de ellos en este contexto. Por ello, a continuación se definen estos:

- *Tratamientos*: son las diferentes condiciones que se aplican sobre las diferentes unidades experimentales y que son susceptibles de ser comparadas entre sí. Por ejemplo, en un experimento en el que se desea comparar el efecto del DFMO en una línea celular usando dos o más concentraciones diferentes, los tratamientos serían dichas concentraciones. Incluso, podrían ser las diferentes líneas celulares de las que se desea comparar, por ejemplo, la expresión diferencial de diferentes genes.
- *Tratamiento Control*: es el tratamiento respecto del cual se comparan el resto de tratamientos y que se entiende que son las condiciones basales. Por ejemplo, puede tratarse de un fármaco de uso común y respecto del cual se quieran comparar nuevos fármacos... No obstante, cuando se usan diferentes tratamientos, el control suele ser la unidad experimental a la que se le da, por ejemplo, el medio en el que se disuelve el tratamiento en el resto de condiciones experimentales pero, en este caso, sin el tratamiento en cuestión, lo que se conoce como *placebo*.
- *Variable independiente, predictora o explicativa*: son las variables a partir de las cuales se explica una variable que depende de ellas. En un modelo lineal típico, como puede ser el de la regresión simple $y = \beta_0 + \beta_1 \cdot x$, la variable y es la variable dependiente cuyos valores pueden explicarse por la variable x , que es la independiente y se considera que es controlable (en modelos más complejos como los modelos mixtos habrá variables explicativas aleatorias y no controlables).
- *Factor fijo*: son las variables independientes de carácter cualitativo cuyas categorías, a las que se denomina *niveles*, dan lugar a los diferentes tratamientos. En el caso del estudio del efecto del DFMO, el factor *Tratamiento con DFMO* tiene dos niveles: *Control* y *DFMO*. Otro ejemplo es el factor *Línea Celular*, que también tiene dos: *HT29* y *NCM460*. Obviamente, un factor puede tener más de dos niveles.
- *Combinación de factores*: en numerosas ocasiones se desarrollan diseños complejos, como por ejemplo en el caso del estudio del efecto del DFMO en las líneas celulares HT29 y NCM460, en el que hay dos factores fijos: *Línea Celular* y *Tratamiento con DFMO*. Así combinando cada uno de sus niveles se obtiene un total de 4 tratamientos: *HT29.Control*, *NCM460.Control*, *HT29.DFMO* y *NCM460.DFMO*.
- *Unidad experimental*: son las muestras sobre las que se aplica un tratamiento. Por ejemplo, en los experimentos del efecto del DFMO en las líneas celulares HT29 hay 2 posibles tratamientos (*HT29.Control*, *HT29.DFMO*), la unidad experimental es el cristal en el que se han sembrado las células y sobre el cual se aplica el tratamiento.

- *Unidades de medida o de respuesta*: es ligeramente diferente a las unidades experimentales, aunque en ocasiones se refieren a lo mismo. En un experimento de imagen, mientras que la unidad experimental es el cristal sobre el que se aplica un tratamiento, la unidad de medida son cada una de las diferentes células sembradas en dicho cristal que han sido registradas por el equipo de imagen de Ca^{2+} , pues este no mide la señal de Ca^{2+} de un cristal, sino de cada una de sus células.
- *Respuesta*: son los registros observados tras aplicar el tratamiento; es decir, la medida tomada con la que se evalúa el experimento. En el experimento de imagen, la respuesta es la señal registrada para cada una de las células de un cristal, por lo que en una misma unidad experimental puede haber más de una respuesta, pero todas ellas para la misma condición experimental. Además, como es el caso de los experimentos en los que se evalúa el efecto del DFMO, para una misma célula es posible obtener más de una respuesta: máximo de la señal de los depósitos, área de la señal de los depósitos, máximo de la señal de SOCE...
- *Aleatorización*: es la aplicación de un mecanismo probabilístico conocido a la hora de asignar a cada una de las muestras (unidad experimental) un tratamiento. También puede ser aplicado para el orden en el que se registra cada una de las diferentes unidades experimentales (el orden en el que se miden los diferentes cristales de un experimento de imagen de Ca^{2+})...
- *Confusión*: tiene lugar cuando parte del efecto de un factor o de un tratamiento, o la totalidad de dicho efecto, es confundido con la de otro factor o tratamiento y no puede ser atribuido a ninguno de ellos. Desafortunadamente, este término aparece con demasiada frecuencia, en ocasiones por malos diseños experimentales, y no siempre se actúa adecuadamente ante él. Ejemplo de un mal diseño que diera lugar a un factor de confusión sería enviar las muestras a secuenciar en dos días diferentes, de tal manera que el primer día se mandasen las NCM460 y otro día las HT29; el efecto observado no podría ser atribuido a que son líneas celulares diferentes, pues las diferencias observadas podrían deberse a que el análisis transcriptómico se ha realizado en dos días diferentes con cada línea celular. Por ello, una estrategia ampliamente usada, aunque menos de lo que se debería, es la aleatorización, pues protege contra los factores de *confusión*.
- *Diseño ciego*: el evaluador o analista de datos no conoce *a priori* qué tratamientos han sido dados a cada unidad. Esto es interesante para evitar errores indeseados por el sesgo a la hora de interpretar los resultados. Por ejemplo, en el análisis de microarrays, en todos aquellos pasos en los que se pueda evitar tener conocimiento sobre qué tratamiento corresponde a cada muestra es más adecuado, pues así se evitan sesgos a la hora de eliminar o no eliminar *outliers*, por ejemplo.
- *Muestras pareadas*: las muestras o unidades experimentales son pareadas cuando están correlacionadas entre sí por algún factor (aleatorio o no). Por ejemplo, pueden estar pareadas por pertenecer al mismo paciente, por haber sido extraídas del mismo cultivo, por haber sido cultivadas el mismo día...

- *Hipótesis*: es una afirmación que se plantea sobre un parámetro poblacional y que se acepta o no en función de las estimaciones sobre la muestra. A tal efecto, hay dos tipos de hipótesis: la *Hipótesis nula* (H_0), que es aquella que se desea contrastar a partir de la evidencia -por norma general, es la que el investigador busca rechazar o la contraria a la que se plantea como hipótesis-, y la *Hipótesis alternativa* (H_1), que es la que se acepta tras rechazar H_0 considerarse falsa a partir de la evidencia. Como apunte, hay que destacar que H_0 nunca queda probada del todo, sino que se desmiente o se rechaza a favor de H_1 en base a la evidencia proporcionada por las pruebas (los datos). Para poder rechazar o admitir una hipótesis es necesario hacer un *test o contraste de hipótesis*, que no es más que un procedimiento aceptado para desmentir o rechazar una hipótesis. Por ejemplo, cuando se compara si dos líneas celulares tienen igual SOCE, se comienza estimando los parámetros poblacionales $SOCE_{HT29}$ y $SOCE_{NCM460}$, cuyas estimaciones son \widehat{SOCE}_{HT29} y \widehat{SOCE}_{NCM460} , la H_0 es $SOCE_{HT29} = SOCE_{NCM460}$ y la alternativa $SOCE_{HT29} \neq SOCE_{NCM460}$. Sin embargo, en otras ocasiones se busca aceptar la hipótesis nula, como, por ejemplo, cuando se evalúa la normalidad y la homocedasticidad del error aleatorio a partir de los residuos del modelo, en el que la H_0 es $\epsilon_i \sim (0, \sigma)$.

En relación con la pregunta anteriormente planteada respecto a por qué experimentar en vez de llevar a cabo un estudio observacional, si en los estudios observacionales también hay tratamientos, variables respuesta, unidades de medida... esta se contesta a través de algunas de las ventajas que tienen los experimentos frente a los estudios observacionales: permiten establecer comparaciones directas entre tratamientos de interés, el adecuado diseño previo del experimento minimiza los errores en las comparaciones y conlleva una disminución del error en estas. Es más, estructurar la investigación en un adecuado diseño experimental permite tener bajo control el experimento lo cual permite obtener unas inferencias mucho más claras en relación a las comparaciones o a las relaciones causa-efecto del estudio.

Tras la breve introducción a los tipos de estudios y a cierta terminología que se empleará a lo largo de la tesis, se procede a presentar brevemente los diferentes modelos, o al menos algunos de sus fundamentos, que se emplearán en el análisis de los datos generados por los diferentes experimentos. Por un lado, se van a presentar los modelos lineales de efectos fijos que engloban desde la regresión simple o multivariante, ampliamente usada en todos los laboratorios, hasta otros más complejos como los modelos lineales generalizados en aquellos casos en los que los datos no cumplen los supuestos de los modelos (normalidad, homogeneidad de las varianzas...).

Posteriormente, se presentarán unos modelos más complejos, denominados modelos lineales mixtos o multinivel, usados tanto en experimentos longitudinales, en los que se miden diferentes variables sobre la misma unidad experimental a lo largo del tiempo, como en experimentos en los que los datos tienen estructura jerárquica, como pueden ser experimentos de imagen de Ca^{2+} en los que, por ejemplo, se estudie un tratamiento a dos niveles (control y DFMO), dentro de cada tratamiento hay diferentes cristales medidos en diferentes días y dentro de cada cristal hay un conjunto de células.

A.6.1. Modelos lineales de efectos fijos

Para una variable aleatoria de interés Y cuyo valor verdadero es η cada vez que se observa su valor fluctúa alrededor de dicho valor verdadero (A.22):

$$Y = \eta + \epsilon \quad (\text{A.22})$$

por lo que en cada observación hay una perturbación o error aleatorio respecto de dicho valor verdadero o esperado ϵ . En el caso de que el valor esperado de la variable aleatoria Y dependa de otras p variables independientes (covariables), entonces el valor verdadero o esperado puede expresarse como (ecuación A.23):

$$\eta = \beta_1 x_1 + \dots \beta_p x_p \quad (\text{A.23})$$

de tal manera que los valores β_j son parámetros desconocidos asociados a las variables independientes y x_j son los valores conocidos de las variables independientes (puede ser, por ejemplo, un tratamiento o una condición experimental).

Por lo tanto, si la variable aleatoria se observa n veces, siendo cada observación de esta variable aleatoria y_i , las cuales son independientes entre ellas (el valor de, por ejemplo, la primera observación, no determina el valor de alguna otra observación) y siguen la misma distribución de Y , entonces estas observaciones siguen el siguiente modelo lineal (ecuación A.24):

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots \beta_j x_{ji} + \epsilon_i \quad (\text{A.24})$$

La anterior expresión se refiere a una sola observación. No obstante, haciendo uso de la notación matricial el modelo teórico del modelo lineal puede expresarse como sigue (ecuación A.25):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (\text{A.25})$$

donde: $\mathbf{Y} = (y_1, \dots, y_n)'$ es el vector de observaciones, puesto que contiene los n valores observados y_i ; \mathbf{X} , denominada matriz de diseño -modelos de regresión- o del modelo -ANOVA-, contiene los valores conocidos de cada covariable para cada observación; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, y $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)'$ el vector de errores aleatorios, desviaciones aleatorias de y_i o perturbaciones.

En esta primera sección se van a estudiar una serie de modelos lineales cuyas variables predictoras son fijas, es decir, que el modelo considera que no son variables aleatorias; en concreto, estas serán las variables predictoras o explicativas, pues las variables respuesta sí son consideradas como aleatorias. Los modelos lineales de efectos fijos más habituales son los modelos de *regresión lineal múltiple* que, a su vez, puede estimarse mediante el método de *mínimos cuadrados ordinarios*

(MCO) o de *mínimos cuadrados ponderados (MCP)*, los cuales son un caso especial de los *mínimos cuadrados generalizados (MCG)*.

Otros modelos lineales de efectos fijos menos habituales pero también muy usados son los *modelos lineales generalizados (MLG)*, los cuales no hay que confundir con el método de MCG, puesto que mientras que MLG permiten que la distribución de los errores de la variable respuesta no se distribuya normalmente y que la varianza dependa del valor predicho, los MCG permiten la heterocedasticidad y la correlación entre las observaciones.

Sin entrar en mucho detalle, puesto que este trabajo no pretende ser un texto sobre modelos, se van a explicar brevemente los modelos anteriores (para profundizar en esta materia se aconseja leer Faraway (2004) y Faraway (2016) los cuales incluyen ejemplos con R , o Oehlert (2010) y Peña (2002b), entre otros).

A.6.1.1. Mínimos cuadrados ordinarios (MCO)

El primer método de estimación de los parámetros del modelo lineal es el de los MCO. Para ello, se parte de la ecuación A.25 se obtiene la expresión del modelo ajustado a los datos observados (ecuación A.26).

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad (\text{A.26})$$

a partir de la cual se puede estimar el error aleatorio a través de los residuos los cuales se obtiene de la diferencia entre el vector de valores observados \mathbf{Y} y ajustados $\hat{\mathbf{Y}}$ (ecuación A.27).

$$\hat{\boldsymbol{\epsilon}} = \mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \quad (\text{A.27})$$

Para este método la expresión anterior es de suma importancia, ya que, precisamente, su nombre se debe a que intenta minimizar la expresión anterior. Es decir, minimiza la suma de cuadrados de los residuos, $\text{SSR} = \sum_{i=1}^n e_i^2$ y, a partir de ello se obtiene el sistema de ecuaciones normales del hiperplano \mathbb{R}^p (ecuación A.28).

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y} \quad (\text{A.28})$$

No obstante, para poder resolver la ecuación anterior es imprescindible que $\mathbf{X}'\mathbf{X}$ sea invertible, por lo que ha de ser de rango máximo p (en el apartado “A.2” se mencionaba la importancia de eliminar variables redundantes para evitar matrices singulares con el fin de poder emplear ciertos métodos como MCO). Así, en el caso de que $\mathbf{X}'\mathbf{X}$ sea invertible (no singular, determinante diferente de cero), se llega a las siguientes expresiones que permiten estimar los parámetros del modelo mediante el método MCO (ecuación A.29):

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} ; \hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y} \quad (\text{A.29})$$

En la expresión anterior aparece la matriz \mathbf{H} , también conocida como matriz de proyección o *Hat*, muy importante a la hora de detectar posibles *outliers*.

Hay que tener en cuenta que la estimación del modelo mediante el método MCO se basa en unas hipótesis y supuestos, de tal manera que la calidad de las inferencias extraídas a partir de este dependerán del grado de cumplimiento de dichas suposiciones, puesto que la violación de alguna de ellas, en ocasiones, puede dar lugar a conclusiones erróneas. Por ello, no hay que olvidar realizar el correspondiente diagnóstico del modelo tras haberlo ajustado a los datos. A continuación, se enumeran los supuestos del modelo de los cuales muchos de ellos suelen ser aplicables a otro tipo de métodos y modelos.

A.6.1.1.1. Supuestos del modelo estimado por mínimos cuadrados ordinarios

En concreto, la principal forma de evaluar el grado de cumplimiento de los supuestos son las estimaciones de las perturbaciones -residuos-, puesto que las suposiciones, en su mayoría, están referidas a dichas perturbaciones (Hipótesis de Gauss-Markov). En concreto, los supuestos básicos de la estimación del modelo multivariante estimado por mínimos cuadrados son:

- *Hipótesis de Gauss-Markov*: estas hipótesis hacen referencia al comportamiento de los errores en el modelo lineal estimado por mínimos cuadrados ordinarios:
 - *Valor esperado nulo de los errores*: se espera que, en promedio, los errores sean 0; es decir, que $E[\epsilon_i] = 0 \quad \forall i$ y, en consecuencia, el valor esperado del modelo lineal es $E[\mathbf{Y}] = \mathbf{X}\beta$.
 - *Homocedasticidad*: se espera que el error se distribuya de forma homogénea y que no dependa de variables como por ejemplo del valor de la variable respuesta, de las predictoras, de alguna variable temporal considerada o no... En resumen, esto se expresa como $\text{Var}[\epsilon_i] = \sigma^2 \quad \forall i$.
 - *Incorrelación entre los errores de las diferentes observaciones*, puesto que en caso contrario estos no serían aleatorios ya que supondría que se podría predecir, por ejemplo, el valor de la observación registrada en tercer lugar a partir de la primera. Por lo tanto, la covarianza entre los diferentes errores ha de ser cero, $\text{Cov}[\epsilon_i, \epsilon_j] = 0 \quad \forall i \neq j$.

Lo anterior se puede resumir con notación matricial: $E[\epsilon] = \mathbf{0}$ y $\text{Var}[\epsilon] = \sigma^2 \mathbf{I}_n$, donde \mathbf{I}_n es la matriz identidad $n \times n$. Cabe destacar que el teorema de Gauss-Markov asegura que el mero cumplimiento de estas tres hipótesis (sin necesidad de que se cumpla la normalidad), proporciona unos estimadores de los coeficientes de regresión, $\hat{\beta}$, que son los *Mejores Estimadores Lineales e Insesgados de Mínima Varianza (BLUE)*.

- *Normalidad de las perturbaciones:* a parte de las hipótesis anteriores, también se hacen suposiciones sobre la distribución de los errores. En particular, si $\epsilon_i \sim (0, \sigma^2)$ y, como se ha visto antes, los ϵ_i son estocásticamente independientes, entonces el modelo lineal se denomina *modelo lineal normal*. Por lo tanto, $\mathbf{Y} \sim (\mathbf{Y}\beta, \sigma^2\mathbf{I}_n)$
- *Especificación:* la elección de las variables predictoras es un proceso complejo (se verá más adelante), pues es necesario que el modelo describa de forma adecuada la relación de dependencia existente entre la variable dependiente y las variables independientes.
- *Linealidad:* en el modelo lineal, como su propio nombre indica, debe existir una relación de dependencia lineal entre la variable respuesta y cada una de las variables predictoras.
- *Aditividad y multicolinealidad:* en estos modelos es de suma importancia que no exista información redundante en las variables independientes. Como se comentó en el apartado “A.2”, cuando en los modelos de regresión multivariante se da la situación de información redundante, en función de si la correlación entre las variables predictoras que causan esta redundancia es o no perfecta, se dice que existe *multicolinealidad perfecta o exacta*, respectivamente.

A.6.1.1.2. Contrastes de hipótesis

En ocasiones, no se suele prestar demasiada importancia al cumplimiento de estos supuestos. En particular, el cumplimiento de los supuestos de homocedasticidad e independencia, $\text{Var}(\epsilon) = \sigma^2\mathbf{I}$, es de suma importancia en la estimación de la varianza de los parámetros del modelo estimados por MCO. Es más, estos parámetros son los que se usan para realizar los contrastes de hipótesis, comparaciones dos a dos... con ánimo a contestar las preguntas que suscitaron la investigación y, de hecho, la varianza juega un papel de suma relevancia en estos contrastes. La estimación de esta se muestra a continuación (ecuación A.30):

$$\begin{aligned} \text{Var}(\hat{\beta}_{\text{MCO}}) &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \text{Var}(\epsilon) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \Sigma \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{I} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (\text{A.30})$$

No obstante, σ^2 es un parámetro poblacional de la distribución de los errores el cual es desconocido, por lo que ha de ser estimado. En particular, se estima a partir de SSR (suma de cuadrados residuales) como

$$\hat{\sigma}_{\text{MCO}}^2 = \frac{\text{SSR}}{n - p} \quad (\text{A.31})$$

por lo que la estimación de la varianza de los parámetros de regresión es (ecuación A.32):

$$\widehat{\text{Var}}(\hat{\beta}_{\text{MCO}}) = \frac{\text{SSR}}{n-p} (\mathbf{X}'\mathbf{X})^{-1} = \hat{\sigma}_{\text{MCO}}^2 (\mathbf{X}'\mathbf{X})^{-1} \quad (\text{A.32})$$

Hasta ahora se ha mostrado cómo estimar los parámetros del modelo, cada uno de los cuales muestra el efecto sobre la variable respuesta o, en caso de ser categórica, el efecto de cada uno de sus niveles o categorías sobre dicha variable, puesto que han de ser codificadas como *dummy variables* de tal manera que a cada categoría le corresponderá un parámetro (ver apartado “A.1” para la codificación *dummy*). No obstante, aun no se ha mostrado cómo saber o, mejor dicho, inferir o contrastar si alguno de ellos tiene efecto sobre la variable respuesta. Así, hay que diferenciar entre inferir si el modelo *per se* es significativo (contraste global de la regresión sobre el conjunto de los parámetros) o si alguno de los parámetros lo es (contraste individual de los parámetros).

En cuanto al contraste global de la regresión, la hipótesis nula que se desea contrastar es $H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0$; o, lo que es lo mismo, $H_0 = \mathbf{X}\beta = 0$, lo que equivale a un modelo sin variables explicativas que se conoce como modelo reducido $\mathbf{Y} = \beta_0 + \epsilon$, donde β_0 será el valor esperado de Y (media). Es decir, se plantea la hipótesis nula de si los datos están mejor descritos por el modelo reducido (H_0) o si, por el contrario, introducir alguna variable mejora el modelo o no. Antes de comenzar con el contraste, hay que tener presente la descomposición de la variabilidad total; es decir, la variabilidad de la variable respuesta (ecuación A.33).

$$\text{SST} = \sum_{i=1}^n (y_i - \bar{y})^2 = \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{\text{SSR}} + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{\text{SSE}} = \text{SSR} + \text{SSE} \quad (\text{A.33})$$

Como puede observarse en la ecuación A.33, sumando y restando en SST por el valor predicho \hat{y}_i se llega a que la variabilidad total puede descomponerse en la suma de dos factores: SSR, que es el error aleatorio, y SSE, que mide la variabilidad entre el valor predicho y la media general (también llamado centroide). Analizando un poco más la expresión anterior, es obvio que SST, puesto que los datos ya han sido observados, siempre es la misma y no cambia, por lo que cuanto mayor sea SSE (variabilidad explicada por el modelo) menor será SSR (variabilidad aleatoria no explicada por el modelo). Si cada una de estas sumas de cuadrados SSE y SSR se divide entre sus grados de libertad entonces se obtienen los *cuadrados medios*, que son estimaciones de las varianzas explicada y no explicada por el modelo, respectivamente (ecuación A.34):

$$\begin{aligned} \text{CME} &= \frac{\text{SSE}}{p-1} \\ \text{CMR} &= \frac{\text{SSR}}{n-p} \end{aligned} \quad (\text{A.34})$$

Estas sumas de cuadrados medios también son variables aleatorias, puesto que ambos dependen del error asociado a los datos; es decir, que si se repite el experimento con una muestra diferente la estimación de cada uno de ellos cambiará. Por lo tanto, se puede obtener el valor esperado de cada uno de ellos y puede demostrarse que tienen las siguientes expresiones (ecuación A.35):

$$E[CME] = \sigma^2 + \frac{n}{p-1} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sigma^2 + \frac{n}{p-1} SSEE [CMR] = \sigma^2 \quad (A.35)$$

analizando los valores esperados se puede concluir que si $H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0$, es decir, si el modelo no explicara nada, entonces $SST = SSR$ y la variabilidad explicada por el modelo $SSE = 0$ y $E[CME] = E[CMR]$. Por lo tanto, si la H_0 es cierta entonces los valores esperados tanto de CMR como de CME serían estimaciones centradas del error aleatorio σ^2 . Por lo tanto, para contrastar la hipótesis nula basta con comparar la varianza explicada (SSE/k) con la varianza no explicada ($SSR/(n-p)$), puesto que si la H_0 es cierta entonces ambas serán iguales y, en caso de que sea significativamente diferente, entonces se rechaza $H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0$ y se acepta $H_1 : \exists \beta_i \neq 0$, entonces algún parámetro es diferente de cero y $E[CME] > E[CMR]$. El contraste empleado para comparar varianzas es el estadístico F el cual sigue una distribución F de *Snedecor* con $k, n-p$ grados de libertad (ecuación A.36):

$$F = \frac{SSE/k}{SSR/n-p} \quad (A.36)$$

Por lo tanto, para un nivel de significación α (por ejemplo, 0.05), el modelo completo explicará significativamente la variable respuesta cuando $F > F_{k,n-p,\alpha}$ o cuando el p valor asociado al valor de F calculado sea inferior al nivel de significación α . El anterior contraste también puede ser visto desde el punto de vista de los modelos reducido y completo, pues la H_0 anterior es similar a decir que los datos están mejor descritos por el modelo reducido, mientras que la H_1 significaría que introducir alguna variable mejora el modelo en el sentido de que el SSR del modelo completo es significativamente más pequeño que el del modelo reducido.

No obstante, aunque evaluar la significación global de la regresión, cuando hay más de una variable (o más de dos categorías para una variable categórica), no basta con saber si el modelo explica o no la variable respuesta. En estos casos, lo que interesa es saber qué variables influyen, de forma individual, sobre el valor de la variable respuesta. Es decir, los contrastes individuales sobre cada uno de los parámetros. Para ello, por ejemplo, se contrasta la hipótesis nula $H_0 : \beta_j = 0$ o, de forma genérica, $H_0 : \beta_j = b$. Para ello, se calcula el estadístico t que, en el caso de que $\mathbf{Y} \sim (\mathbf{Y}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$, este seguirá una distribución t de *Student* con $n-p$ grados de libertad (ecuación A.37):

$$t = \frac{\hat{\beta}_j - b}{SE_{\beta_j}} \quad (A.37)$$

Por lo tanto, para un nivel de significación α (por ejemplo, 0.05), un parámetro tendrá un efecto significativo sobre la variable respuesta si $t > t_{n-p, \alpha/2}$. Es importante señalar que este contraste emplea SE_{β_j} , la raíz cuadrada de $\widehat{\text{Var}}(\beta_j)$, que es el j elemento del vector de estimaciones de las varianzas de los parámetros estimado en A.30, estimaciones que se hacen bajo la suposición $\text{Var}(\epsilon) = \sigma^2 I$. Por lo tanto, el cumplimiento de los supuestos de homocedasticidad e independencia son sumamente necesarios para realizar dicho test.

A parte de contrastar si cada uno de los parámetros individuales es o no significativo, también se puede contrastar si diferentes tratamientos tienen o no el mismo efecto sobre la variable respuesta, para lo que se emplean combinaciones lineales de las estimaciones de los tratamientos de interés. En particular, si se tiene una variable explicativa cualitativa formada por los k diferentes tratamientos de interés, de tal manera que se construyen k variables dicotómicas cada una de las cuales corresponde con un tratamiento, se estima un parámetro $\hat{\beta}_j$ para cada tratamiento. Así, para ver si dos tratamientos j y j' tienen el mismo efecto sobre la variable respuesta, se puede estimar si la diferencia entre ambos es cero. O sea, contrastar la hipótesis nula $H_0 : \beta_j - \beta_{j'} = 0$ frente a la alternativa $H_1 : \beta_j - \beta_{j'} \neq 0$.

A.6.1.1.3. Comparaciones múltiples

Cuando se realiza un único contraste de hipótesis se fija un nivel de significación α , que es la probabilidad de cometer un error de tipo I o de cometer un falso positivo (FP); es decir, de rechazar la H_0 cuando es verdadera. Por lo tanto, para un contraste dado se rechaza la H_0 cuando el p valor asociado al estadístico de contraste es inferior al nivel de significación, $p \leq \alpha$. La consecuencia de fijar este valor o esta cota conlleva que de cada 100 veces que se realizase ese mismo contraste de hipótesis con 100 muestras extraídas de la misma población, en el caso de que no fuera significativo, se esperaría que en $\alpha \cdot 100$ de ellos rechazase la H_0 . Por lo tanto, cada vez que se realiza un contraste, para dicho contraste el error de tipo I se controla fijando el nivel de significación.

No obstante, cuando se realizan múltiples contrastes de forma simultánea, como por ejemplo todas las comparaciones dos a dos entre los diferentes tratamientos o todas las comparaciones de cada uno de los tratamientos respecto del control, la probabilidad de cometer un error de tipo I en cada uno de los contrastes realizados sigue siendo α . Si se emplea el estadístico T para realizar los contrastes de hipótesis, siendo T_i el estadístico estimado para el contraste de hipótesis i , de un total de N contrastes simultáneos, existe un valor crítico T_α asociado a la distribución de dicho estadístico para el nivel de significación α tal que se rechaza la $H_{i,0}$ (hipótesis nula para el contraste i) cuando el T_i , en valor absoluto, sea muy grande, $|T_i| > T_\alpha$ ¹. En consecuencia, $P(|T_i| > T_\alpha | H_{i,0}) = \alpha$.

¹Para el contraste $H_0 : \beta_j = 0$ se calcula el estadístico de la ecuación A.37, donde la $H_1 : \beta_j \neq 0$; es decir, que la hipótesis alternativa se compone de dos eventos excluyentes que la satisfacen, $H_1 : \beta_j > 0$ o $\beta_j < 0$. En consecuencia, se realiza una prueba de dos colas, lo que significa que si el nivel de significación es α entonces $\alpha/2$ se corresponde a $H_{1,1} : \beta_j > 0$ y el otro $\alpha/2$ a $H_{1,2} : \beta_j < 0$. Dado que, en este caso, el estadístico t_{β_j} sigue una distribución t de Student, el valor crítico será $t_{\alpha/2}$ para $H_{1,2} : \beta_j < 0$ y $t_{1-\alpha/2}$ para $H_{1,1} : \beta_j > 0$. Además, puesto que esta distribución es simétrica, resulta que $|t_{1-\alpha/2, \text{gdl}}| = t_{1-\alpha/2, \text{gdl}}$ y, en consecuencia, H_0 se rechaza cuando el estadístico estimado $|t_{\beta_j}| > t_{1-\alpha/2, \text{gdl}}$.

Por lo tanto, a pesar de que para el contraste *i-ésimo* se controla el error de tipo I, al realizar N contrastes simultáneamente se esperaría que el $\alpha \cdot 100\%$ de dichos contrastes, en el caso de no ser significativos, rechazasen igualmente la H_0 . De hecho, el número de hipótesis nulas erróneamente rechazadas o los FP cometidos es una variable aleatoria que sigue una distribución binomial $p(\alpha, N)$. En consecuencia, la probabilidad de cometer “al menos” un falso positivo, considerando que cometer o no un falso positivo sigue una distribución binomial, es $P(\text{FP} \geq 1) = 1 - P(\text{FP} = 0) = 1 - \binom{N}{0} \alpha^0 \cdot (1 - \alpha)^N$.

Por ejemplo, si se tienen 4 tratamientos diferentes, se pueden realizar $\binom{4}{2} = 6$ comparaciones dos a dos, por lo que $P(\text{FP} \geq 1) = 0.2649081$; es decir, que de los 6 contrastes de hipótesis realizados, hay un 26.5% de probabilidad de inferir que una de las diferencias es diferente aun en el caso de que todas fueran iguales. Así, si con solo 6 contrastes aumenta la probabilidad en tal magnitud, hay que pensar cuánto aumenta la probabilidad de cometer al menos un falso positivo en experimentos transcriptómicos en los que se realizan miles de contrastes de hipótesis simultáneamente, pues la probabilidad de cometer un falso positivo es 1. Por ejemplo, si se contrastase la expresión diferencial de 1000 genes, para un nivel de significación de $\alpha = 0.05$ se esperaría encontrar 50 falsos positivos aun cuando ningún gen estuviera diferencialmente expresado. Por lo tanto, es evidente que hay que controlar de alguna manera este problema, conocido como problema de las comparaciones múltiples.

En consecuencia, cuando hay múltiples contrastes, en vez de controlar el error tipo I (nivel de significación) lo que se controlan son las *tasas de error de tipo I*. Además, para cada tasa de error de tipo I existen diferentes procedimientos que la controlan a un nivel de significación α de tal manera que una tasa de error de tipo I dada será controlada si es inferior o igual a dicho nivel de significación. Las tasas de error más interesantes son:

- *Tasa de error por comparación* (PCER, del inglés *per-comparison error rate*): es el cociente entre el valor esperado de errores tipo I entre el número de contrastes realizados, $\text{PCER} = E(\text{FP})/N$.
- *Tasa de error global* (FWER, del inglés *family-wise discovery rate*): esta tasa de error proporciona unos contrastes de hipótesis más conservativos; es decir, que le “cuesta” rechazar la H_0 . Se puede definir como la probabilidad de cometer uno o más errores, $\text{FWER} = P(\text{FP} \geq 1)$, por lo que mediante esta medida se controla que no se cometa ningún error considerando a todos los contrastes simultáneamente o, en otras palabras, se fija la probabilidad de cometer al menos un error al realizar todos los contrastes en lugar de fijar la probabilidad de cometer un FP en cada uno de los contrastes de forma individual. No obstante, esta tasa de error es muy peligrosa cuando el número de contrastes es elevado, pues a medida que aumenta N se hace cada vez más exigente y, en consecuencia, cada vez le cuesta más rechazar H_0 , por lo que puede aumentar de forma desorbitada la probabilidad de cometer falsos negativos (FN) y, en consecuencia, disminuye la potencia del test; es

decir, de no rechazar la H_0 cuando sí que es falsa. No obstante, la forma más habitual de controlar las tasa de error de tipo I, por razones históricas, es, precisamente, controlando FWER (bonferroni...).

- *Tasa de contrastes falsamente rechazados* (FDR ,del inglés *false Discovery Rate*): puesto que el control de la tasa de error a través de FWER puede ser inútil cuando N es muy elevado, se han definido otras tasas de error que dan lugar a contrastes de hipótesis más potentes. En particular, Benjamini and Hochberg (1995) propusieron controlar la tasa o porcentaje de H_0 falsamente rechazadas (FP) de entre todas las H_0 rechazadas (R), por lo que puede construirse la variable aleatoria $Q = FP/R$, donde $Q = 0$ si no se rechaza ninguna H_0 ($R = 0$). Por lo tanto, la tasa de falsamente rechazados, que es una variable aleatoria, se define como $FDR = E(Q) = E\left(\frac{FP}{R} | R > 0\right) P(R > 0)$. Para entender esto y cuál es su diferencia respecto de controlar simplemente el error de tipo I, si se controla $FDR = \alpha$, entonces el $\alpha \cdot 100\%$ de R (H_0 rechazadas) son ciertas y, en consecuencia, se rechazan erróneamente $\alpha R \cdot 100\%$ H_0 . Por el contrario, si se controla únicamente el error de tipo I, tal que se rechaza H_0 cuando $p_i \leq \alpha$, entonces se rechazan erróneamente $\alpha \cdot N 100\%$ H_0 .

Cabe destacar que, a diferencia del control de FWER, donde se controla el no cometer ningún error de tipo I, que es demasiado conservador, FDR permite o tolera algunos errores de tipo I en relación con el número total de H_0 que se rechazan, por lo que se emplea cuando el número de contrastes total es muy elevado. Por ello, el control de FDR se emplea en el análisis de datos de experimentos transcriptómicos donde se tienen miles de contrastes simultáneamente.

- *Tasa de falsamente rechazados modificada* (pFDR): partiendo de FDR, se define $pFDR = E\left(\frac{FP}{R} | R > 0\right)$, que, cuando N es muy elevado entonces $P(R > 0) \approx 1$ (la probabilidad de rechazar al menos una H_0 es casi seguro) y $FDR \approx pFDR$.

Cabe destacar que $PCER \leq FDR \leq FWER$, de tal manera que cualquier procedimiento que controle la tasa de error FWER a un nivel de significación α entonces también controla FDR y PCER. En consecuencia, FWER es más conservador, menos potencia del contraste, que FDR y este, a su vez, más conservador que FWER.

En cuanto a cómo controlar estas tasas de error a partir de los contrastes empleados, lo que se hace es ajustar el p_i valor proporcionado por el contraste *i-ésimo*, de tal manera que ahora no se rechaza la $H_{0,i}$ cuando $p_i \leq \alpha$, sino cuando $p_i \leq \alpha_i$. No obstante, con el fin de no tener que comprobar un valor diferente α_i para cada contraste, lo que se hace es corregir o ajustar el p_i original. Para ello, lo que se hace es corregir el estadístico de contraste de tal manera que ahora se rechaza la H_0 si $p_{i,ajustado} \leq \alpha$.

Como se ha dicho anteriormente, la tasa de error de tipo I más habitualmente controlada, sobretodo en las comparaciones dos a dos tras realizar el ANOVA de un modelo, es la FWER. De entre los diferentes métodos existentes, el más extendido es el método de Bonferroni, en el cual, si se realizan N contrastes de hipótesis, entonces se rechaza la H_0 cuando $p_i \leq \alpha/N$, por lo

que $Np_i \leq \alpha$ y, en consecuencia, el $p_{i,\text{ajustado}} = \min\{Np_i, 1\}$, pues toda probabilidad pertenece al intervalo $[0, 1]$. Existen otros métodos más complejos que controlan FWER, como el de Tukey.

En cuanto al control de FDR, en la presente tesis, para el análisis de datos transcriptómicos, se ha empleado el procedimiento propuesto por Benjamini and Hochberg (1995) para controlar FDR. Este consiste en ordenar los p_i valores originales de menor a mayor, donde $p_{1_i} \leq p_{r_i} \leq p_{N_i}$; a continuación obtener el $i^* = \max \{i | p_{r_i} \leq i \cdot \alpha/N\}$ y rechazar las $H_{i,0}$ para aquellos $i = 1, \dots, i^*$. Obviamente, si no existe i^* entonces $R=0$. En consecuencia, los $p_{r_i,\text{ajustados}}$ son (ecuación A.38):

$$p_{r_i,\text{ajustado}} = \text{FDR} = \min_{k=1,\dots,N} \left\{ \min \left\{ \frac{N}{k} p_{r_k}, 1 \right\} \right\} \quad (\text{A.38})$$

A.6.1.2. Mínimos cuadrados generalizados

Hasta el momento, se ha considerado que $\text{Var}(\epsilon_{\text{MCO}}) = \sigma^2 I$, pero cuando se violan los supuestos de homocedasticidad y/o de independencia las estimaciones MCO pierden sus buenas propiedades y las inferencias alcanzadas pueden ser totalmente contrarias a la realidad. En tal caso, es conveniente plantear otras estrategias para estimar los parámetros del modelo. Una opción habitual es recurrir a los métodos de estimación por mínimos cuadrados generalizados (MCG) en los que se relaja el supuesto de heterocedasticidad y homocedasticidad. Esto se debe a que ahora no se presupone la homogeneidad de las varianzas ni la no correlación entre los errores (ecuación A.39).

$$\text{Var}(\epsilon_{\text{MCG}}) = \sigma^2 \Sigma = \sigma^2 \Delta C \Delta' \neq \text{Var}(\epsilon_{\text{MCO}}) = \sigma^2 I \quad (\text{A.39})$$

y, por ende, $\epsilon_i \sim N(0, \sigma_i^2)$; es decir, que cada error sigue una normal con una varianza que no tiene por qué ser igual a la de otros errores. En la ecuación A.39, la matriz Σ se descompone como el producto matricial $\Sigma = \Delta C \Delta'$, donde C es la matriz de correlaciones (que en MCO era la matriz identidad que no tiene elementos de correlación) y Δ es la matriz que modela la heterocedasticidad (que en MCO era el escalar σ^2 común a todos los errores). No obstante, emplear MCG no es una tarea tan sencilla como podría parecer, pues hay un exceso de parámetros a estimar superior al número de observaciones, lo que da lugar a un sistema compatible indeterminado y, por ende, los parámetros no quedarían determinados sino que podrían tener infinitas soluciones. Para hacer frente a esta situación, la estrategia más común es imponer una serie de restricciones sobre las σ_i^2 , para lo que hay diferentes aproximaciones:

- Una de las estrategias más usadas es suponer que las varianzas σ_i^2 son proporcionales a σ^2 con unos pesos w_i de tal manera que $\epsilon_i \sim N(0, \sigma^2/w_i)$, de tal manera que $\Sigma = \Delta \Delta'$, siendo Δ es la matriz diagonal que contiene el inverso de la raíz cuadrada de los pesos y, en consecuencia, Σ es la matriz diagonal del inverso de los pesos, $1/w_i$. No obstante, hay que suponer el conocimiento *a priori* de dichos pesos. En particular, esta estimación se conoce

como *mínimos cuadrados ponderados (MCP)*, pues son un caso especial de los MCG, los cuales permiten obtener unos estimadores óptimos del modelo. Así, una vez se “conocen” los pesos, la estimación del modelo se lleva a cabo a través de la transformación de los datos mediante dichos pesos w_i (ecuación A.40):

$$\begin{aligned} Y &= X\beta + \epsilon \rightarrow \Delta^{-1}Y = \Delta^{-1}X\beta + \Delta^{-1}\epsilon \\ Y^* &= X^*\beta + \epsilon^* \end{aligned} \tag{A.40}$$

En consecuencia, ahora la varianza del modelo es $\text{Var}(\epsilon^*) = \sigma^2 I$, por lo que, evidentemente, se ha transformado el problema de la heterocedasticidad en un problema de MCO. No obstante, para aplicar este método es necesario conocer el patrón de heterocedastidad de los errores con el fin de encontrar la matriz de pesos más adecuada. Por ejemplo, si tras aplicar MCO se representan los residuos frente al valor predicho y se observa que la dispersión de estos aumenta con el valor de la variable respuesta predicha, entonces la matriz Σ podría ser la matriz diagonal cuyos elementos fueran el inverso de los valores predichos del modelo r_i^2 vs. \hat{y} , el inverso del valor de la variable respuesta predicha, e incluso el inverso de los valores de la variable que se piense que está produciendo dicho patrón de heterocedasticidad (para evaluarlo se representan los residuos frente a cada una de las variables hasta encontrar la culpable).

- Otra opción es emplear funciones para representar la matriz que modela la varianza, Δ , las cuales dependerán de una serie de parámetros: $\text{Var}(\epsilon_{\text{MCG}}) = \sigma^2 \lambda^2(\delta, \mu_i; v_i)$, siendo μ_i el valor esperado de la variable respuesta para la observación i ; v_i un vector conocido de variables dependientes capaces explicar la varianza para la observación i ; δ un vector cuyos elementos son los parámetros de varianza comunes a todas y cada una de las observaciones, y λ la función que modela la desviación estándar. En consecuencia, se tiene que el error para la observación i sigue una normal $\epsilon_i \sim N(0, \sigma^2 \lambda_i^2)$. Un caso especial de este método serían los *MCP* en los que $\lambda(w_i) = 1/w_i$.
- En cuanto a la correlación entre las perturbaciones, se puede representar la matriz que la modela, C_Z , como una función que dependerá de diferentes parámetros. Para ello, es imperativo conocer los datos y la estructura subyacente de correlación entre los errores. Por ejemplo, si se tiene un conjunto de datos que resultan estar agrupados cada uno en un grupo z en cada uno de los cuales hay n_z observaciones, se podría modelar la estructura de la correlación como $\text{Cor}(\epsilon_{iz}, \epsilon_{i^*z}) = h[d(t_{iz}, t_{i^*z}), q]$, siendo q el vector de correlación de los parámetros, $d(t_{iz}, t_{i^*z})$ el vector de las variables de posición t_{iz} y t_{i^*z} para ϵ_{iz} y ϵ_{i^*z} , y $h(\cdot, \cdot)$ es una función continua respecto a q que toma valores entre -1 y $+1$, siendo $h(0, q) = 1$.

Aunque lo anterior pueda parecer abrumador, sobretodo lo correspondiente al uso de funciones para modelar la heterocedasticidad y la correlación, la buena noticia es que existen funciones que facilitan considerablemente esta tarea. Por ejemplo, en R el paquete `nlme` permite implementar funciones correspondientes a las matrices de varianzas y correlación.

A.6.1.3. Modelos Lineales Generalizados (MLG)

Los métodos de estimación anteriores consideran que los errores siguen una distribución normal. No obstante, esto no siempre es así y, por ello, es necesario recurrir a otros métodos de estimación de los modelos lineales. Así, surgen los *modelos lineales generalizados (MLG)*, que son una generalización del modelo de regresión lineal en el que se no se exige la normalidad de la variable respuesta o, más bien, de los errores; incluso, estas variables respuesta pueden ser tanto continuas (distribución gamma) como discretas (distribución binomial). Además, también permite que la varianza no sea constante sino que esta esté acorde a la distribución considerada para la variable respuesta; por ejemplo, si la variable respuesta sigue una distribución binomial negativa, la varianza aumentará con el valor predicho de la variable respuesta.

A diferencia de los métodos MCO o MCG, en los cuales la relación entre la esperanza de la variable es lineal respecto de las variables independientes (ecuación A.41):

$$E [Y] = X\beta \quad (\text{A.41})$$

en los MLG esta relación tiene lugar entre una función de la esperanza de la variable respuesta, $l(E [Y])$, y las variables explicativas (ecuación A.42):

$$l(E [Y]) = X\beta = \theta \quad (\text{A.42})$$

siendo θ se conoce como el parámetro canónico de localización de la distribución considerada; por ejemplo, en el caso de que la distribución de los errores o de la variable respuesta sea normal, entonces $l(E [Y])=identidad$ y $\theta = \mu$. A esta función $l(E [Y])$ se le conoce como *función link* la cual dependerá de la distribución de la variable respuesta Y . Por ejemplo, para la *distribución normal* esta función será la identidad, $l(E [Y]) = E [Y]$, mientras que para la *distribución Gamma* esta la función link canónica es el recíproco, $l(E [Y]) = 1/E [Y]$, a pesar de que también puede hacer uso de otras funciones link tales como como la logarítmica o la identidad. La única restricción que se hace sobre esta función es que la distribución de Y ha de pertenecer a una distribución de la familia exponencial (Normal, Gamma, Exponencial, Bernouilli, Poisson, Binomial, ...); es decir, que la función de densidad (o función de masa en el caso de tratarse de una distribución de variable discreta) pueda expresarse como (ecuación A.43):

$$f(y|\theta_1, \dots, \theta_k) = a(\theta_1, \dots, \theta_k)b(y)\exp \left\{ \sum_{i=1}^k c_i(\theta_1, \dots, \theta_k)d_i(y) \right\} \quad (\text{A.43})$$

siendo θ_i son los k parámetros poblacionales (pueden ser uno -Poisson-, dos -Normal-...), $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ y $d(\cdot)$ funciones concretas de la distribución de Y . Por ejemplo, para una variable respuesta que sigue una distribución normal se tiene que (ecuación A.44):

$$\begin{aligned}
 f(y|\mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2 \right\} \\
 &= \underbrace{\frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{1}{2} \left(\frac{\mu^2}{\sigma^2} \right) \right\}}_{a(\mu, \sigma)} \underbrace{1}_{b(y)} \exp \left\{ \underbrace{-\frac{1}{2\sigma^2}}_{c_1(\mu, \sigma)} \underbrace{y^2}_{d_1(y)} + \underbrace{\frac{\mu^2}{\sigma^2}}_{c_2(\mu, \sigma)} \underbrace{y}_{d_2(y)} \right\}
 \end{aligned} \tag{A.44}$$

Una de las estrategias que se siguen en las estimaciones MCO o MCG cuando la variable respuesta no es normal es transformar dicha variable, por ejemplo, mediante su logaritmo. No obstante, uno podría pensar que esta solución sería suficiente y que sería lo mismo que emplear MLG, pero es aquí donde radica uno de los grandes malos entendidos, pues la función *link* no transforma los datos, sino la esperanza de la variable respuesta. Cabe destacar que las estimaciones que pertenecen a la familia exponencial son insesgados, uniforme y de mínima varianza.

A.6.1.4. Suposiciones del modelo de regresión lineal multivariante

Los modelos anteriores han de cumplir una serie de supuestos de tal manera que cuando alguno de ellos es violado entonces hay que estudiar las razones de dicha violación y actuar acorde a ella. No obstante, hay que tener muy presente que en el mundo real el total cumplimiento de las suposiciones sobre los errores (independencia, valor esperado cero, homocedasticidad, distribuidos normalmente con media 0 y varianza σ^2) no suele cumplirse, pues, por ejemplo, la distribución normal no es más que un modelo y en la realidad los datos no la siguen de forma exacta.

Afortunadamente, siempre que estas violaciones no sean demasiado entonces pueden obtenerse inferencias razonablemente buenas. A esta propiedad por la cual las inferencias obtenidas son razonablemente buenas en un rango de violación de las suposiciones se la conoce como *robustez de validación*. No obstante, esta propiedad también suscita otras preguntas como, por ejemplo, cuánto de grande ha de ser la violación para no confiar en el modelo. De hecho, autores como Sokal (1995) o FitzMaurice Garrett, Laird Nan, and Ware James (2004) aseguran que la violación de la normalidad no es un problema grave como consecuencia del teorema central del límite, e incluso innecesaria cuando el tamaño muestral es grande. En relación con la calidad de las inferencias, podría establecerse el siguiente orden en términos de violación de los supuestos:

$$\text{Independencia} > \text{Homocedasticidad} > \text{Normalidad} > \text{Multicolinealidad} \tag{A.45}$$

La mayoría de los supuestos se evalúan a partir de los residuos; es decir, de la diferencia entre el valor observado y el predicho, por lo que el vector de residuos se obtiene como (ecuación A.46):

$$\hat{\mathbf{e}} = \mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \quad (\text{A.46})$$

donde cada uno de los elementos del vector de residuos $\hat{\mathbf{e}}$ es (ecuación A.47):

$$r_i = y_i - \hat{y}_i \quad (\text{A.47})$$

No obstante, estos suelen modificarse ligeramente para hacerlos más interpretables y poder decidir, por ejemplo, si un residuo es o no demasiado grande. Una opción es escalarlos dividiéndolos entre la estimación de su error estándar con el fin de que tengan todos la misma varianza, los cuales se denominan *residuos estandarizados (re)* (ecuación A.48):

$$re_i = \frac{r_i}{\hat{\sigma} \sqrt{(1 - H_{ii})}} \quad (\text{A.48})$$

donde H_{ii} es el elemento i de la diagonal de la matriz *hat*, conocido como *leveraje* o apalancamiento. Cabe destacar que esta estandarización sólo es válida cuando se cumple la suposición de homocedasticidad. No obstante, estos residuos aun no proporcionan un criterio para poder decidir si la observación correspondiente es o no un posible *outliers* en función de cuán lejos está un valor predicho del observado. Para ello, a partir de la estimación de la varianza de cada residuo i (ecuación estimada a partir de un modelo en el cual dicha observación se haya eliminado -se identifica por el subíndice “(i)”- (ecuación A.49):

$$\widehat{\text{Var}}(y_i - \hat{y}_{(i)}) = \hat{\sigma}_{(i)}^2 \left(1 + \mathbf{x}'_i (\mathbf{X}'_{(i)} \mathbf{X}_{(i)})^{-1} \mathbf{x}_i \right) \quad (\text{A.49})$$

por lo que empleando este factor de escalado se obtienen los conocidos como *residuos estudentizados (rt)*(ecuación A.50):

$$rt_i = \frac{y_i - \hat{y}_{(i)}}{\hat{\sigma}_{(i)} \left(1 + \mathbf{x}'_i (\mathbf{X}'_{(i)} \mathbf{X}_{(i)})^{-1} \mathbf{x}_i \right)^{1/2}} \quad (\text{A.50})$$

obviamente, puede parecer que calcular un modelo para cada una de las observaciones en el cual dicha observación no se incluya para poder obtener estas $\widehat{\text{Var}}(y_i - \hat{y}_{(i)})$ podría suponer un coste computacional muy grande cuando se usan bases de datos muy grandes. No obstante, la expresión anterior se puede expresar como una corrección de los residuos escalados *re* (ecuación A.51):

$$rt_i = re_i \left(\frac{n-p-1}{n-p-re_i^2} \right)^{1/2} \quad (\text{A.51})$$

Así, a partir de los residuos estudentizados, rt_i , es posible decidir si un dato sigue o no el mismo patrón de comportamiento que el resto, puesto que si el modelo es adecuado entonces los rt_i deberán seguir una distribución t de *Student* con $n - p$ grados de libertad.

No obstante, los residuos anteriores son los que se emplean comúnmente en los modelos *gaussianos* (MCO, MCG...). Sin embargo, en modelos más complejos en los que, por ejemplo, la varianza no es constante, como en los MLG, la cual depende de la distribución considerada para la variable respuesta, es necesario emplear otro tipo de residuos, como el *residuo de Pearson* (rp):

$$rp_i = \frac{y_i - \hat{y}_{(i)}}{\text{Var}(\hat{y}_i)} \quad (\text{A.52})$$

el cual es un tipo de residuo estandarizado cuya suma de sus cuadrados sigue una distribución $\sum_{i=1}^n (rp_i^2) \sim \chi^2$, más recomendado para otros modelos como MLG o modelos lineales mixtos.

A continuación, en los siguientes subapartados se exponen los principales supuestos que han de cumplir los modelos ordenados de menor a mayor importancia, se explica cómo identificar la violación de estos, qué la puede causar, qué consecuencias tiene y cómo se trata.

A.6.1.4.1. Normalidad

Para evaluar el supuesto de normalidad de los errores se emplean los residuos. Para ello, se han desarrollado diferentes métodos tanto gráficos como basados en contrastes de hipótesis:

- *Gráficos*: entre los más usados destacan los histogramas y los *Q-Q plot* de los residuos, a poder ser estudentizados para fijar un valor máximo a partir del cual considerar que una observación es un *outliers*. Cuando el número de observaciones es bajo, es preferible emplear este tipo de gráficos. En particular, *Q-Q plot*, es la representación de los cuantiles de la distribución de los residuos (ordenados de menor a mayor) enfrentados a los cuantiles teóricos de la variable aleatoria $\epsilon_i \sim N(0, \sigma^2)$. Así, se cumple el supuesto de normalidad cuando los valores representados se sitúan en la diagonal principal, y dependiendo de la forma en la que se desvíen de dicha diagonal, la violación de este supuesto será mayor o menor, más o menos peligroso... En cuanto al histograma de los residuos, este debería ser similar a la función de densidad de la normal, por lo que se suele representar sobre este la estimación de la función de densidad de los residuos y la distribución teórica normal y se comparan.
- *Contrastes de hipótesis*: la evaluación de la normalidad mediante test o contrastes de hipótesis. Entre los más comunes destaca el test de *Shapiro Wilks* o el de *Jarque Bera*. En general, se contrastan las siguientes hipótesis:
 - H_0 : ϵ_i se distribuyen según una normal.
 - H_1 : ϵ_i no se distribuyen según una normal.

La violación del supuesto de normalidad puede tener varias causas. Una causa muy común es un tamaño muestral pequeño, por lo que observar fluctuaciones importantes en la normalidad no indicaría necesariamente la falta de normalidad. Otra causa muy común es cuando hay presencia de heterocedasticidad o cuando hay un error de especificación que da lugar a un modelo que no es adecuado para explicar la variable respuesta y , por ende, que los residuos muestren una distribución diferente a la normal. Otro de los grandes protagonistas en esta violación es la presencia de observaciones atípicas o influyentes.

En cuanto a sus consecuencias, por un lado, puede afectar a los estimadores MC , puesto que dejan de ser los estimadores *máximo-verosímiles*. Es más, este efecto se agrava ante la presencia de unos pocos *outliers* y, en consecuencia, la eficiencia de los estimadores se reduce o, en otras palabras, aumenta su variabilidad, si bien es cierto que siguen siendo centrados o insesgados. Esto da lugar a un aumento del error de tipo I en los contrastes de hipótesis de los estimadores lo que se traduce en la disminución en la potencia del test. No obstante, algunos autores como Sokal (1995) o FitzMaurice Garrett, Laird Nan, and Ware James (2004) aseguran que la violación del supuesto de normalidad, como consecuencia del *teorema central del límite*, no es un problema grave y, de hecho, para tamaños muestrales grandes es innecesario. De hecho los contrastes t (contrastos individuales de los parámetros, comparaciones dos a dos...) y F (comparación entre modelos, significación global del modelo) proporcionan unos resultados aproximados pero válidos; es decir, habrá que tener cautela con aquellos valores del test que estén cercanos al nivel de significación. Lo que sí es cierto es que

Por lo tanto, aunque esta violación sea la menos perjudicial, es necesario que haya procedimientos para que cuando esta sea demasiado evidente y poder afrontar este problema. Las estrategias más habituales son *transformar la variable respuesta* -logaritmos, raíz cuadrada, inverso... (transformaciones de BoxCox), *eliminar posibles outliers o puntos influyentes*, o recurrir a otros métodos como los MCG, MLG, e incluso otros más sofisticados, pero más costosos en términos computacionales, como los basados en remuestreo (*bootstrap*):

- *Transformaciones de Box-Cox*: La falta de normalidad puede disminuirse mediante cambios de escala de la variable respuesta y . Una estrategia muy interesante que se usa no solo para la falta de normalidad, sino también para la falta de homocedasticidad y/o de linealidad, son la *familia de transformaciones de potencia de Box-Cox* (Box and Cox (1964)):

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda y^{\lambda-1}} & \text{Si } \lambda \neq 0 \\ \dot{y} \log(y) & \text{Si } \lambda = 0 \end{cases} \quad (\text{A.53})$$

donde \dot{y} es la media geométrica. Así, a partir de la función anterior se transforma la variable respuesta para diferentes valores de λ , de tal manera que para cada valor se realiza un ANOVA del modelo resultante y se escoge aquel valor de λ que proporcione un modelo con una menor suma de cuadrados del error aleatorio ($SSR(\lambda)$); es decir, donde $SSR(\lambda)$ alcanza su mínimo. No obstante, la función anterior requiere de valores positivos de la variable respuesta. Por ello,

Hawkins and Weisberg (2017) propusieron una modificación para poder considerar valores no positivos de la variable respuesta en la que se emplea una constante γ tal que la nueva variable respuesta a transformar sea $z > 0$:

$$z = 0.5 \cdot (y + \sqrt{(y^2 + \gamma^2)}) \quad (\text{A.54})$$

donde gamma puede ser estimado o seleccionado por el analista de tal manera que z sea positivo. En consecuencia, se tiene que (ecuación A.55):

$$z^{(\lambda)} = \begin{cases} \frac{z^\lambda - 1}{\lambda z^{\lambda-1}} & \text{Si } \lambda \neq 0 \\ \dot{z} \log(z) & \text{Si } \lambda = 0 \end{cases} \quad (\text{A.55})$$

De forma general, cuando se observa asimetría a la derecha esta puede ser disminuída mediante la raíz cuadrada, el logaritmo y otra potencia inferior a 1 de y , mientras que si la asimetría es a la izquierda entonces esta puede ser “reparada” por su cuadrado, el cubo u otras superiores a 1.

Obviamente, después de estimar el modelo, para obtener las estimaciones en la escala original habrá que deshacer las transformaciones anteriores. Es importante tener en cuenta que para proporcionar los intervalos de confianza ($\mu \pm t_{\text{gdl}, \alpha/2} \cdot \text{SEM}$) o los valores $\mu \pm \text{SEM}$, estas operaciones han de realizarse previamente a deshacer la transformación, tanto para μ como para los límites superiores e inferiores anteriores y, una vez estimados, deshacer la transformación de cada uno de ellos. Así, si se define como $T(\cdot)$ como la función transformación, a $T^{-1}(\cdot)$ como la función inversa (deshacer la transformación), a $\mu^{(\lambda)}$ al valor estimado con el modelo en las unidades de la variable transformada y a $\text{SEM}^{(\lambda)}$ al valor de la desviación estimada también en las unidades de la variable transformada, se tiene que, por lo general, $T(\mu^{(\lambda)} \pm t_{\text{gdl}, \alpha/2} \cdot \text{SEM}^{(\lambda)}) \neq T(\mu^{(\lambda)}) \pm T(t_{\text{gdl}, \alpha/2} \cdot \text{SEM}^{(\lambda)}) \neq \mu \pm t_{\text{gdl}, \alpha/2} \cdot \text{SEM}$.

- *Bootstrap*: En ocasiones, cuando a pesar de emplear diferentes estrategias, las violaciones de los supuestos del modelo, tales como normalidad u homocedasticidad, son demasiado evidentes, es necesario emplear otro tipo de enfoque más robustos. Entre las diferentes estrategias, existen las basadas en estimadores robustos u otras más sofisticadas conocidas como *bootstrapping*. En particular, las técnicas de *bootstrapping* paramétrico se basa en que la muestra es una muestra representativa de la población. Por ello, se puede considerar a esta muestra, siempre que sea lo suficientemente grande, como la población y extraer muestras de igual tamaño de esta misma, lo que se conoce como remuestreo. En el caso de los modelos lineales (de efectos fijos o de efectos mixtos, que se verán más adelante), se parte de la muestra original y se estima el modelo lineal deseado, a cuyas estimaciones se llamarán t^0 . A continuación, bajo el supuesto de que la muestra es una muestra representativa de la población, se han llevado a cabo remuestreos de un tamaño muestral igual al original; es decir, que sobre la muestra original de N elementos se toman N elementos con reemplazo,

por lo que un elemento puede ser escogido más de una vez y otros no ser escogidos. Así, para cada nueva muestra o *remuestra* se ajusta un modelo lineal, de tal manera que este procedimiento se repite un número elevado de veces, por ejemplo $B = 1000$ -de hecho, en la presente tesis se han realizado 1000 remuestreos para cada estimación *bootstrapping*.

Puesto que se estiman $B = 1000$ modelos, cada uno de ellos con una *remuestra*, cada parámetro del modelo se estima $B = 1000$ veces. Por lo tanto, en vez de suponer una distribución para cada parámetro, por ejemplo la normal, lo que se hace con este método es generar la hipótesis nula para cada parámetro (coeficientes del modelo, diferencias dos a dos...). En consecuencia, puesto que se tienen los parámetros originales (β_j^0 para $j = 1, \dots, k$), se puede ver qué cuantil ocupan en dicha distribución nula y así inferir si son o no significativos. Para ello, por ejemplo se toman los $\hat{\beta}_0^{(i)}$ donde el superíndice (i) indica el remuestreo $i \in \{1, 2, \dots, 1000\}$ y $\hat{\beta}_0^{(i)}$ la estimación del parámetro β_0 en el remuestreo (i).

A partir de estas estimaciones se obtienen los estadísticos correspondientes a cada parámetro de cada remuestreo, como puede ser el estadístico *t* de *Student*, $t_0^{(i)}$, y se construye la distribución de la hipótesis nula no del parámetro en cuestión, sino del estadístico correspondiente al contraste de hipótesis de interés para dicho parámetro, como por ejemplo $\beta_0 = 0$ s. A continuación, se calcula el estadístico correspondiente a $\hat{\beta}_0$ para el modelo original; es decir, para $\hat{\beta}_0^{(0)}$ (el clásico, sin remuestreo), que se denominará t_0^0 . Finalmente, se evalúa cómo de lejos está t_0^0 de la distribución nula construida a partir de los $B = 1000$ $t_0^{(i)}$ (figura A.1). En cuanto al *p* valor asociado a cada uno de estos contrastes, existen diferentes estrategias; pero en la presente tesis se ha usado la siguiente (ecuación A.56):

$$p - \text{valor} = 2 \cdot \min \{ \# \{ t^i \leq t^0 \}, \# \{ t^i \geq t^0 \} \} \quad (\text{A.56})$$

que, posteriormente, será corregido mediante algún método de comparaciones múltiples como Bonferroni o Tukey. C.5 se muestra cómo se han obtenido los *p* valores de los diferentes contrastes de hipótesis de interés y en las tablas C.14 y C.13 los resultados numéricos correspondientes a dichos contrastes.

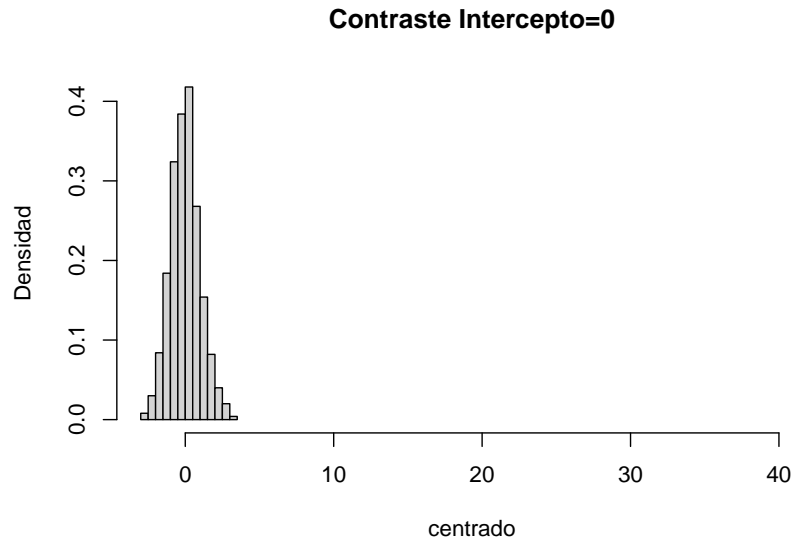


Figura A.1: Ejemplo de la hipótesis nula generada mediante *Boostrapping* paramétrico partir de los $t_j^{(i)}$ para el estadístico t_j asociado al parámetro β_0 . Se muestran las distribuciones nulas, con ayuda de un histograma, de cada uno de los estadísticos correspondientes a los contrastes de hipótesis de interés, y el estadístico original se representa en cada uno de ellos mediante una línea vertical roja para evidenciar cómo de cerca o lejos está de dicha distribución.

A.6.1.4.2. Homocedasticidad

El modelo lineal estimado mediante MCO establece que la varianza de las perturbaciones es constante, o sea, que $\epsilon = \sigma^2 I_n$, lo que se conoce como el supuesto de homocedasticidad, mientras que en caso de no cumplirse se dice que hay heterocedasticidad. Al igual que para el supuesto de normalidad, para evaluar el supuesto de homocedasticidad se analizan los residuos del modelo. Además, en este caso también existen tanto métodos gráficos como basados en contrastes de hipótesis:

- Contrastes de hipótesis para evaluar la homocedasticidad hay muchos; entre los más habituales están el test de *Breusch-Pagan*, *White*, *Goldfeld-Quandt* o *Barlett*.
- Un gráfico sumamente útil es la representación de los residuos frente a los valores ajustados o predichos de la variable respuesta o frente a las variables explicativas (figura A.2). Así, en caso de homocedasticidad se esperaría ver una dispersión de los residuos constante que no variase en función de ninguna de dichas variables. Por el contrario, en caso de violación del supuesto de homocedasticidad, se observaría una clara variación de estos residuos a lo largo de alguna de las variables o del valor predicho de la respuesta. Otra opción es representar la raíz cuadrada del valor absoluto de los residuos estudentizados frente al valor predicho o a las variables explicativas.

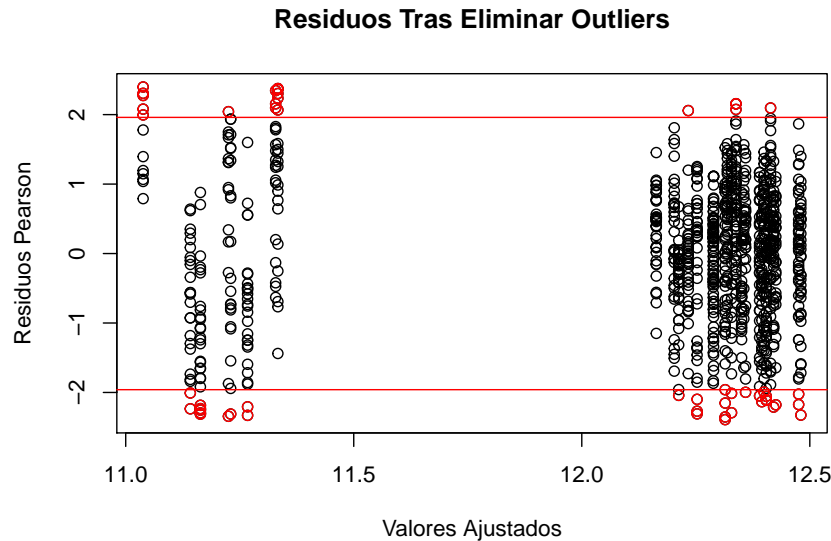


Figura A.2: Residuos vs. valores predichos por el modelo. En el gráfico se representan los residuos de un modelo lineal frente a los valores predichos por este. Las líneas rojas marcan un punto de corte común a partir del cual se podría considerar a una observación como atípica.

La presencia de heterocedasticidad puede deberse a muy diferentes causas. A continuación se enumeran algunas de las más importantes:

- *Función de variables explicativas:* puede suceder que la varianza de los residuos sea función de alguna de las variables explicativas, situación que suele ser habitual en modelos econométricos, ecológicos o ambientales. Por ejemplo, si se desea medir el gasto en productos no básicos (variable respuesta) en función de la renta de una familia (variable independiente o explicativa), se observará una menor varianza entre las familias con rentas bajas, puesto que se preocupan más por productos básicos y tienen menos dinero sobrante para gastar en productos no básicos. Por el contrario, esta varianza será mayor entre aquellas familias con rentas más altas, pues al tener más dinero sobrante tras adquirir los productos básicos pasan a entrar en juego otras variables tales como ser más o menos ahorrador, tener o no *hobbies* más o menos caros, tener o no previsto la adquisición de otros bienes -empresas, casas, coches...-, etc.
- *Uso de datos agrupados:* cuando el modelo tiene variables predictoras categóricas (factores) como sucede en la inmensa mayoría de los experimentos realizados en un laboratorio la dispersión de la variable respuesta puede ser muy diferente para cada una de las diferentes categorías. Por ejemplo, al comparar un tejido tumoral vs. uno normal, puede darse el caso que, debido a las características propias del tumor, las muestras pertenecientes a dicha categoría muestren una mayor variabilidad que las que puede haber en el tejido normal.

- *Errores de especificación:* cuando se realiza un estudio observacional puede suceder tanto que haya demasiadas variables como que no todas las que de verdad influyen en la variable respuesta hayan sido registradas. No obstante, en los estudios experimentales también puede pasar que, a pesar de controlar la población en estudio, haya alguna variable no controlada que no se haya considerado. En cualquier caso, se estaría omitiendo alguna variable del modelo que podría ser importante en el proceso generador de los datos. En consecuencia, las perturbaciones estimadas del modelo incompleto (sin las variables que, erróneamente, no se han considerado) contendrán no sólo la aleatoriedad, sino también el efecto determinista de la variable omitida, por lo que la esperanza de las perturbaciones no es cero, lo que resulta en el incumplimiento de otro de los supuesto de Gauss-Marcov y, por ende, los estimadores de los coeficientes de regresión serán sesgados. Para más inri, si la varianza de la variable omitida no es constante también se producirá heterocedasticidad en el modelo incompleto.
- *Linealidad:* hay ocasiones en los que la variable respuesta depende de una variable explicativa pero no de forma lineal, sino que, por ejemplo, dependa del cuadrado de esta. Es decir, que la relación entre la variable respuesta y la/s explicativa/s no sea lineal.

En cuanto a las consecuencias de la violación de la homocedasticidad, si esta es pequeña entonces no supone un problema serio en la estimación del modelo (Sokal (1995)). No obstante, si la violación es grande conlleva que las estimaciones ya no sean BLUE de mínima varianza. En consecuencia, los estimadores MCO dejan de ser eficientes, de tal manera que dicha eficiencia se hará menor a medida que aumente la magnitud de la heterocedasticidad. De hecho, como se verá a continuación en la estimación *sandwich* (ecuación A.57) la estimación de las varianzas de los estimadores de los parámetros por MCO no es adecuada en presencia de heterocedasticidad y, por ende, los test individuales t y F para cada uno de ellos dejan de ser válidos y, por ende, se obtendrían inferencias erróneas. No obstante, el test de significación global de la regresión y los R^2 siguen siendo válidos. Por lo tanto, los estimadores MCO todavía son insesgados y consistentes en presencia de heterocedasticidad, pero no óptimos.

Es obvio que las observaciones con una varianza baja proporcionan una información más fiable en la estimación del modelo que aquellas que presentan una varianza mayor. Por ello, una de las estrategias para tratar la heterocedasticidad es dar más peso a las observaciones más fiables. Para ello, se pueden emplear *MCG* o la *estimación sandwich*. En concreto, la *estimación sandwich* se basa en cómo se estima la varianza de los errores en los modelos estimados por MCO, $\text{Var}(\epsilon) = \sigma^2 I_n$, pues la expresión anterior, evidentemente, deja de ser válida ya que no hay un σ^2 común a todos los errores. Por lo tanto, la estimación de la varianza de los parámetros estimados $\text{Var}(\hat{\beta}) \neq \sigma^2 (X'X)^{-1}$, sino que pasa a estimarse como (ecuación A.57):

$$\text{Var}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} X' \Sigma X (X'X)^{-1} \neq \sigma^2 (X'X)^{-1} \quad (\text{A.57})$$

A partir de la expresión anterior diferentes autores han propuesto diferentes estrategias para estimar $\hat{\sigma}^2 \Sigma$. Algunas de estas estrategias son, por ejemplo, la propuesta por H. White and others (1980), que consiste en $\hat{\sigma}^2 \Sigma = \text{diag}(\hat{\epsilon}_i^2)$, o por Long and Ervin (2000), en la que $\hat{\sigma}^2 \Sigma = \text{diag}(\hat{\epsilon}_i^2 / (1 - h_i)^2)$. No obstante, hay más estrategias en este sentido (consultar Zeileis (2004) para más información).

En estos y otro tipo de modelos, otra estrategia para resolver la heterocedasticidad de los datos es transformar la variable respuesta mediante la familia de transformaciones potencia de Box-Cox explicada en A.53 (Box and Cox (1964)).

A.6.1.4.3. Independencia

Según las hipótesis de Gauss-Markov, los errores han de ser independientes entre sí y no mostrar ningún patrón no aleatorio; es decir, que las observaciones lo sean, pues no hay que olvidar que cada error estimado corresponde a una observación. Métodos gráficos que permitan evaluar la aleatoriedad o independencia entre los errores son los mismos que para evaluar la homocedasticidad. No obstante, también hay métodos basados en contrastes de hipótesis para tal fin como el *Test de Durbin Watson*. La violación de este supuesto suele darse, por ejemplo en datos agrupados (células de un mismo cristal en los experimentos de imagen de Ca^{2+} , muestras en un mismo gel en experimentos de *Western-blot*...), en series temporales donde los valores de la variable respuesta de un tiempo t se ven influenciados por valores anteriores cercanos a dicho tiempo t (valores de fluorescencia en experimentos de imagen de Ca^{2+} ...).

Es más, hay ocasiones en las que puede darse la violación del supuesto de independencia incluso cuando no debería de haber ningún patrón de correlación entre las observaciones, como sucede cuando el modelo, de forma sistemática, sobreestima o infraestima los valores predichos, lo que resulta en residuos “seguidos” con el mismo signo. Además, también puede observarse un patrón no aleatorio de los residuos en situaciones tales como falta de linealidad o cuando se ha omitido erróneamente alguna variable explicativa...

Para tratar de solventar esta violación del modelo hay que analizar cuál es el culpable de dicha violación. Entre las diferentes estrategias empleadas para tratar este problema está el uso de MCG en los que se consideran diferentes estructuras de correlación para la matriz de varianzas-covarianzas. No obstante, en aquellos casos en los que la correlación sea debida a medidas repetidas sobre los mismos individuos pueden emplearse *modelos lineales mixtos*. No obstante, no siempre es necesario recurrir a otro tipo de métodos de estimación; por ejemplo, si el problema es la no linealidad de la respuesta frente a alguna variable predictora entonces se realiza alguna transformación de las variables problemáticas.

A.6.1.4.4. Aditividad y multicolinealidad

La *multicolinealidad perfecta* puede dar lugar a problemas a la hora de obtener los modelos. En la estimación por mínimos cuadrados, para obtener las estimaciones de los coeficientes de regresión se requiere que $\mathbf{X}'\mathbf{X}$ sea invertible, lo que no se cumple en presencia de multicolinealidad perfecta al ser $\mathbf{X}'\mathbf{X}$ singular (si bien es cierto que existen inversas generalizadas). Además, cuando el número de variables regresoras es elevado, se corre el riesgo de que la matriz $\mathbf{X}'\mathbf{X}$ sea casi singular ($\det(\mathbf{X}'\mathbf{X}) \approx 0$), pues suele darse el caso de que algunas de las variables predictoras puedan ser expresadas, aproximadamente, como combinación lineal de otras y ser redundantes, por lo que seguramente estén introduciendo ruido pero no información adicional. A este problema se le conoce como *multicolinealidad no perfecta o de grado*. El problema surge porque al $\det(\mathbf{X}'\mathbf{X}) \approx 0$, las varianzas y covarianzas de los parámetros estimados se inflan, ya que $\widehat{\text{Var}}(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$. Como consecuencia de esto, la precisión en la estimación de los coeficientes de regresión disminuye. Por ello, se tiene una baja eficiencia de las estimaciones de los coeficientes, por lo que el intervalo de confianza de éstos se hace mayor y acaba por incluir a cero. A pesar de ello, los estimadores minimocuadráticos siguen siendo óptimos, insesgados y lineales. Por ello, es fácilmente detectable cuando los test de t individuales para los contrastes de hipótesis de que $H_0 : \beta_j = 0$ no son significativos, a pesar de que el contraste de significación global sí lo sea al no verse afectado en presencia de *multicolinealidad*, pues R^2 no está afectado por dicho problema. Además, cabe destacar que a mayor número de variables, R^2 tiende a crecer y la bondad de ajuste será parecida ante la presencia de variables que introducen ruido pero no información adicional al modelo, por lo que una medida alternativa al R^2 es el R^2_{ajustado} (ecuación A.58) que penaliza por el número de variables explicativas en el modelo. No obstante, un sutil cambio en las matrices \mathbf{X} e \mathbf{Y} pueden provocar grandes cambios en las estimaciones de los parámetros.

Para diagnosticarla puede emplearse métodos basados en la correlación muestral entre variables regresoras, como el cálculo de inflación de la varianza, o *métodos basados en la matriz $\mathbf{X}'\mathbf{X}$* . Para evitar los problemas relacionados con la multicolinealidad, pueden emplearse diferentes enfoques, como *eliminar variables explicativas* bajo la premisa de que si dos variables explicativas están muy correlacionadas entre sí es posible excluir del modelo alguna de ellas sin pérdida de información. Sin embargo, en caso de que el modelo más complejo, sin eliminar la variable explicativa j fuera correcto, el modelo reducido, sin la variable j , produciría unas estimaciones sesgadas. Otras medidas pueden ser *aumentar el tamaño muestral*, *emplear información extramuestral* o *transformar las variables* de tal forma que se sustituyen las variables redundantes por la suma, promedio, etc ... de las mismas, o empleando ratios respecto de aquella variable explicativa que mayor multicolinealidad presente. Sin embargo, en caso de que las suposiciones de normalidad y homocedasticidad se cumplan en el modelo sin transformar (sin el ratio), es posible que no se cumplan en el modelo que emplé los ratios.

A.6.1.5. Selección de Variables

A la hora de plantearse un estudio, ya sea observacional o experimental, uno puede pensar en diferentes variables como candidatas a ser variables explicativas para explicar la variable respuesta de interés. No obstante, no todas ellas tienen por qué aportar información relevante o adicional al modelo a la hora de explicar dicha variable respuesta. Es más, como ya se comentó anteriormente en el apartado “A.6.1.4.4 Aditividad y multicolinealidad”, el uso de muchas variables regresoras en el modelo puede generar multicolinealidad no perfecta. Por lo tanto, es imprescindible realizar un proceso de selección de variables con el fin de obtener estimaciones más precisas, pues cuanto mayor sea el número de elementos innecesarios en el modelo menor será la eficacia del modelo y mayor el riesgo de obtener conclusiones erróneas a partir de él. En este sentido, entra en juego el conocido *principio de parsimonia* o la *navaja de Ockham*, pues se busca aquel modelo lo más simple que explique lo mejor posible la variable dependiente. Así, entre dos modelos muy parecidos donde ambos proporcionen unos ajustes razonablemente buenos y similares se seleccionará el más sencillo a pesar de que el ajuste sea ligeramente peor. Además, a la hora de interpretar un modelo siempre será más sencillo e intuitivo un modelo más sencillo con pocas o ninguna interacción que un modelo con una complejidad elevada.

Para llevar a cabo la selección de variables pueden plantearse diferentes enfoques: selección basada en contraste de hipótesis o selección basada en criterios. A continuación, se explican brevemente cada uno de ellos:

- *Selección de variables basada en contrastes de hipótesis*: estos métodos emplean contrastes de hipótesis, como test de significación (t o F), para medir la importancia de las variables explicativas en el modelo. No obstante, estos métodos han de ser tomados con cautela, pues entre los problemas que pueden plantearse está el de las comparaciones múltiples debido a que los estimadores de los coeficientes de regresión se están estimando simultáneamente.
- *Selección de variables basada en criterios*: estos métodos se basan en buscar aquel modelo, de entre los 2^p posibles modelos distintos, siendo p el número de variables regresoras candidatas, más adecuado según algún criterio. Los criterios más empleados son:
 - *El coeficiente de determinación ajustado*: Se calculan todos los coeficientes de determinación ajustados R_a^2 :

$$R_a^2 = 1 - \frac{RSS/(n - k)}{TSS/(n - 1)} \quad (A.58)$$

para todos los modelos posibles que pueden formarse a partir de k variables regresoras, siendo k un subconjunto de las p potenciales variables regresoras. Se usa este coeficiente de determinación ajustado porque el coeficiente de determinación *no ajustado* (el R^2) siempre será mayor cuantas más variables regresoras contenga el modelo, lo cual no sería un buen criterio. Por lo tanto, el R_a^2 , puesto que penaliza por el número de variables

que contiene el modelo, sólo aumentará en el caso de que la variable regresora añadida tuviera algún valor predictivo adicional.

- *El estadístico Cp de Mallows*: aquí se plantea un estadístico tal que, para $k+1$ parámetros (subconjunto de variables regresoras más intercepto), el mejor modelo será aquel cuyo C_p sea menor:

$$C_k = \frac{SSR_k}{\hat{\sigma}^2} + 2(k+1) - n \quad (\text{A.59})$$

donde $\hat{\sigma}^2$ es la estimación de la varianza del modelo completo y SSR_k la suma de cuadrados residuales del modelo con k parámetros. Si el modelo es adecuado, puede demostrarse que $E(SSR_k) = (n-k)\sigma^2$ y $E(C_k) \approx k$, siendo malos modelos aquellos cuyo $C_p > k$

- *Información de Akaike (AIC)*: basado en la información de Kullback-Leibler y la teoría de verosimilitud, el criterio de Akaike tiene el siguiente aspecto (ecuación A.60):

$$AIC = -2\log(\hat{\theta}) + 2(k+1) = n \ln(SSR/n) + 2(k+1) \quad (\text{A.60})$$

donde la primera parte de la igualdad, $\log(\hat{\theta})$, es el menos dos veces la función de log-máximoverosimilitud, y donde la segunda igualdad es el caso especial para la regresión por mínimos cuadrados. Hay que destacar que *AIC* no es un test de contraste de hipótesis para realizar inferencias sobre el modelo, sino que es un criterio paramétrico que se emplea para comparar diferentes modelos con el fin de seleccionar, de entre todos los evaluados, el más parsimonioso (aquel modelo más simple que explique lo mejor posible los datos observados), que será aquel que proporcione un menor valor de *AIC*. Por un lado, este criterio de información emplea el valor de la bondad de ajuste el cual será menor cuanto mejor se ajuste el modelo a los datos. Por otro lado, penaliza la cantidad de parámetros (variables...) empleados en el modelo, pues suma $2(k+1)$. Por lo tanto, un buen modelo será aquel que ajuste bien los datos, es decir, que proporcione un valor pequeño de bondad de ajuste, y, además, que contenga una cantidad lo suficientemente reducida de parámetros. Hay otros tipos de métodos de selección que emplean criterios en los que se requiere que los modelos sean anidados, pero *AIC* no tienen tal limitación.

- *Información bayesiana*: otro criterio muy empleado es la información bayesiana (*BIC*, del inglés *bayesian information criterion*), desarrollado por Schwarz (1978), el cual es similar a *AIC* salvo en el penalti, pues en vez de usar $2(k+1)$ emplea $2(k+1)\log(n)$, donde n es el número de observaciones.

Tras decidir el o los criterios empleados para llevar a cabo el proceso de selección de variables, el siguiente paso es aplicar dichos criterios y encontrar el modelo más adecuado y parsimonioso. No obstante, si se tienen p variables explicativas parecería lógico plantearse generar todos los posibles modelos, evaluarlos y compararlos entre sí. No obstante, cuando el número de variables

explicativas candidatas o potenciales es reducido entonces este planteamiento es posible. Sin embargo, cuando el número de posibles variables explicativas es elevado tiene lugar la *explosión exponencial* consecuencia del elevado número de posibles modelos a evaluar, pues aumenta de forma exponencial con cada variable adicional. Es decir, que el número de posibles nuevos modelos que se generarían al aumentar el número de variables en 1 (por ejemplo, pasar de 2 a 3 variables explicativas conllevaría pasar de evaluar de 8 a 16, mientras que pasar de 10 a 11 conllevaría pasar de evaluar 2048 modelos a 4096). En este último caso, existen diferentes algoritmos que evitan esta explosión:

- *Procedimiento secuencial*: hay algoritmos que se plantean la selección del modelo mediante un enfoque secuencial; es decir, que van incluyendo o eliminando variables una a una del modelo de forma secuencial. Estos algoritmos se conocen como *forward*, *backward* o *stepwise*, y se empleará algún criterio de información, como los anteriormente presentados, para decidir si dicha variable introducida o eliminada mejora o empeora el modelo. Los algoritmos secuenciales son un tipo de algoritmo *eurístico*, pues al no probar todos los posibles modelos, sino que se quita o se pone una variable en función de un modelo, no aseguran que se encuentre el modelo *óptimo*, aunque sí, en caso de no ser el óptimo, estará muy cercano a él. A continuación, se presentan los tres algoritmos mencionados:
 - *Forward*: a partir de un modelo sencillo (sin variables explicativas, por ejemplo), se van añadiendo de forma secuencial nuevas variables explicativas, una a una, al modelo considerando algún criterio. Así, en cada etapa, de todas las variables explicativas disponibles se añade aquella que genere una mayor mejora del modelo. Este proceso se repite hasta que el introducir una nueva variable no de lugar a una mejora del modelo.
 - *Backward*: en este método se parte del modelo completo, el cual contiene las p variables e interacciones posibles. Entonces, en cada paso o etapa el número de variables se reducirá en uno, pues se eliminará aquella variable que menos contribuya al modelo según algún criterio o test de significación escogido. Por lo tanto, se selecciona aquel modelo que, tras eliminar una a una las variables que no contribuyen, se llegue a un modelo en el que todas las variables predictoras que permanezcan sean significativas (según el test F o el test de t) o que según algún criterio el modelo empeore al quitar cualquiera de ellas. Además, cabe destacar que a la hora de eliminar las variables hay que tener en cuenta las posibles estructuras jerárquicas; es decir, que una interacción de diferentes variables estará “por encima” en la jerarquía que las variables individuales que participan en ella y, por ende, no se podrá eliminar ninguna variable si la interacción en cuestión está presente en el modelo.
 - *Stepwise*: este método es una mezcla de los dos métodos anteriores, pues en cada etapa de inclusión de una nueva variable al modelo se evalúa si en el modelo actual sería o no adecuado eliminar alguna de las variables presentes. El proceso sigue hasta que no se produzca ningún modelo que mejore el anterior según alguno de los criterios o reglas consideradas.

- *Best subset o exhaustivos*: este otro tipo de algoritmos generan todos los posibles modelos a partir de las p variables y las posibles interacciones deseadas. Así, es posible escoger el óptimo o el mejor según alguno de los criterios que se hayan elegido. Aunque pueda parecer el más adecuado, como se ha dicho anteriormente, si el número de posibles variables explicativas es muy elevado se produce la explosión exponencial.

A.6.1.6. Detección y tratamiento de Outliers

Una vez se ha seleccionado el modelo, el siguiente paso es realizar un diagnóstico del mismo con el fin de comprobar si las suposiciones del modelo se cumplen, las cuales ya se explicaron en el apartado “A.6.1.4 Suposiciones del modelo de regresión lineal multivariante”. Además, la selección de variables anteriormente explicada es muy sensible a valores atípicos o influyentes. Por lo tanto, es necesario realizar ‘reviamente a la selección un adecuado análisis de los residuos y eliminar estos de los *outliers* detectados.

Hay que diferenciar entre dos tipos de valores “aberrantes” en el modelo: valores atípicos y valores influyentes. Mientras que los primeros pueden desviarse del modelo y no tienen por qué distorsionar las estimaciones que este genera, los segundos pueden influir en las estimaciones y generar modelos erróneos. A continuación, se explica como tratar cada uno de ellos.

- *Residuos estudentizados*: a partir de los residuos estudentizados o de Pearson es posible considerar una observación como *outliers* si el residuo correspondiente es mayor a dos en valor absoluto. No obstante, es importante considerar las comparaciones múltiples y obtener un valor crítico más adecuado a cada modelo, para lo que pueden emplearse diferentes correcciones, como por ejemplo Bonferroni.
- *Distancia de cook*: esta distancia mide la influencia de cada una de las observaciones. A partir de ella, se considerará a una observación como influyente si su distancia de Cook es superior a la del resto de una forma clara, si tiene un valor superior a 1 o si es superior a $Cook_{\text{critico}} = 4/(n - k - 1)$, donde k es el número de coeficientes estimados (uno por cada variable explicativa más uno del intercepto).
- *Valor hat*: también es posible detectar puntos influyente a partir de los valores de la matriz sombrero o matriz de influencia, $\hat{H} = X(X'X)^{-1}X'$. Esta tiene una serie de características: dimensión $n \times n$, rango k , traza k y los elementos de la diagonal principal con valores entre 0 y 1. Por lo tanto, si algún elemento de la diagonal principal h_{ii} es 1 entonces el resto serían iguales a 0 y, por ende, el hiperplano ajustado pasaría por dicha observación independientemente del valor del resto de observaciones. Por lo tanto, un punto influyente suele considerarse a aquel cuyo $h_{ii} > 2(k + 1)/n$. Estos elementos h_{ii} se les conoce como valores *hat*. Cabe destacar que en regresiones multivariantes con más de dos predictores los valores *hat* son determinantes para detectar las observaciones de influencia, pues el diagnóstico por medio de gráficos puede resultar complejo y tedioso.

Finalmente, tras haber detectado los *outliers* y puntos influyentes, estos se eliminan de la base de datos y se estima nuevamente el modelo completo para, posteriormente, llevar a cabo la selección de variables. No obstante, será necesario reevaluar los supuestos de este nuevo modelo y repetir el proceso de detección de observaciones atípicas y/o influyentes.

A.6.2. Modelos Lineales Mixtos

En ocasiones, es habitual encontrarse con datos que están relacionados entre sí en forma de clúster, como por ejemplo cuando se dispone de varias observaciones correspondientes a un mismo individuo. Es decir, que sobre cada una de las unidades experimentales se realizan diferentes medidas, a cada una de las cuales se les podría considerar como subindividuos de un mismo individuo. Por ejemplo, en el caso de querer evaluar diferentes métodos de enseñanza diferentes y para ello se escogen, al azar, 9 colegios. A su vez, cada método de enseñanza se asigna a 3 clases diferentes de cada colegio, por lo que en todos los colegios se aplican los tres métodos de enseñanza pero en cada clase sólo uno. Por lo tanto, los alumnos de una misma clase emplean un único método de enseñanza.

En consecuencia, se forma una estructura de correlación entre las diferentes medidas: las notas de los alumnos se agrupan dentro de una única clase y las clases se agrupan dentro de cada colegio, por lo que los alumnos de una misma clase están correlacionados entre sí y las clases de un mismo colegio también entre sí. Este mismo ejemplo se puede trasladar, por ejemplo, a datos obtenidos de experimentos de imagen de Ca^{2+} , en los que las células se agrupan dentro de cada cubreobjetos en el que son sembradas y cada cubreobjetos se agrupa dentro de cada unidad experimental en caso de que para cada individuo haya varios cubreobjetos. Para los datos de *Western Blot* también se tiene una estructura jerárquica similar, pues las muestras analizadas dentro de un mismo gel estarán relacionadas entre sí.

A continuación, se muestran algunos planteamientos que no tienen en cuenta la correlación entre las medidas sobre un mismo individuo y, por ello, no suelen ser las más recomendables:

- *Resumen de los datos mediante algún estadístico*: por ejemplo, para analizar datos longitudinales en los que para cada punto temporal se calcula la media y la varianza, o en datos provenientes de técnicas como *Patch Clamp* en los que para cada “parche” se tiene una curva corriente vs. voltaje y para cada voltaje se hace la media y la desviación de todos los parches de una misma condición experimental. Posteriormente, se analizan los estadísticos obtenidos, por ejemplo, mediante ANOVA o test de t. El problema de este planteamiento es que se reduce el análisis de datos longitudinales o de medidas repetidas a un mero análisis de observaciones independientes, uno para cada punto temporal o para cada voltaje.
- *Análisis de los puntos finales*: otro enfoque que puede utilizarse en los ejemplos anteriores es el de analizar ya no cada uno de los puntos temporales o de voltaje, sino únicamente el último y comparar entre condiciones. No obstante, el problema es el mismo, la pérdida de la

información que contiene la estructura de correlación entre las medidas pertenecientes a un mismo individuo.

- *Análisis del área bajo la curva (AUC) o incrementos de la señal:* cuando se tienen datos longitudinales, por ejemplo, se pueden obtener curvas y, por ende, áreas bajo cada una de ellas. Por lo tanto, es posible usar como variable respuesta dicho área. El problema de este planteamiento es que diferentes curvas pueden proporcionar áreas iguales. También, en el caso de analizar los incrementos de la señal respecto de la basal, dicho incremento no tiene por qué producirse a un mismo tiempo para cada individuo. Por lo tanto, hay que tener claro cual es el fin del análisis para sopesar si basta con analizar el AUC o el incremento o si, por el contrario, hay que modelizar la señal a lo largo del tiempo. Por ejemplo, en los experimentos de imagen de Ca^{2+} de la presente tesis el interés no es modelizar la señal sino evaluar el incremento y el AUC, por lo que en dicho caso es adecuado este planeamiento. No obstante, a pesar de que se reduce el análisis a usar como variables respuesta el AUC y el incremento, en los experimentos de imagen llevados a cabo sí se tiene en cuenta la correlación existente entre las células medidas en un mismo cubreobjetos y en un día en concreto, por lo que en este caso se emplearán modelos lineales mixtos (explicados a continuación) para controlar este factor aleatorio de la agrupación de las células de un mismo cubreobjetos y de los cubreobjetos medidos en un mismo día que dan estructura de correlación entre dichas células y cubreobjetos.
- *Regresión lineal simple:* una estrategia también de resumen de las observaciones dentro de cada individuo es realizar una regresión lineal simple, tal que para cada uno de los sujetos se estima una recta de regresión mediante mínimos cuadrados y, posteriormente, se comparan los promedios de las pendientes e interceptos entre los diferentes grupos de interés.

Sin embargo, como se ha dicho anteriormente, se ha dicho que este enfoque anterior no suele ser el más acertado (salvo ciertas ocasiones como el planteado para las áreas e incrementos en función del interés del estudio). Esto se debe a que las estrategias anteriores conllevan la pérdida de información resultante de la estructura de correlación de las medidas repetidas; es decir, de las observaciones realizadas sobre un mismo individuo (cubreobjetos, paciente...). Además, en muchas ocasiones suele haber datos faltantes en los diferentes estudios, ya sea porque son observacionales o porque alguna condición experimental no ha sido posible realizarla en el laboratorio para algunos individuos (se ha perdido la muestra, se ha roto el cubreobjetos...).

En consecuencia, surgen los conocidos como *modelos lineales mixtos* los cuales permiten considerar estas estructuras de agrupación entre las diferentes observaciones. Es decir, tener en cuenta en el análisis la estructura de los datos agrupados, los cuales, por definirlo más formalmente, son una colección de observaciones que están relacionados entre sí de forma estocástica, espacial, temporal... Es importante también comprender cómo se generan estas agrupaciones o en qué contextos, que puede dividirse en dos procesos: *procesos aleatorios jerárquicos*, como el caso de los experimentos de imagen, y *medidas repetidas o datos longitudinales*. Es más, permite estimar el comportamiento individual respecto del comportamiento poblacional.

Estos modelos consideran tanto factores fijos como aleatorios y, por ello, recibe el apellido de “mixtos.” No obstante, no siempre es sencillo decidir si un factor es fijo o aleatorio y, en ocasiones, también depende del interés del análisis. Por un lado, un factor aleatorio es aquel cuyos niveles han sido extraídos al azar de una población de posibles niveles, como por ejemplo el factor *Paciente* del que se dispone tanto de muestras normales como tumorales, pues los niveles de dicho factor son varios pacientes que han sido extraídos al azar de una población que contiene a todos los pacientes con cáncer. Otro ejemplo serían los geles o membranas en un experimento de *Western Blot*, en el que las observaciones se agrupan dentro de diferentes geles y, además, dichos geles han sido extraídos de forma aleatoria de una población de los infinitos geles que podrían haberse realizado para ese mismo experimento. También hay que pensar en un factor aleatorio como aquel factor del cual no se desea obtener una estimación del efecto de cada uno de sus niveles, como el caso de los geles en el *Western Blot*, pues sería extraño que el fin del estudio fuera evaluar el efecto de un gel sobre la expresión de una proteína. Por otro lado, un factor aleatorio es aquel cuyos niveles son los únicos posibles y que han sido determinados por el investigador, como por ejemplo el factor *Tipo de tejido*, en el que sólo hay dos posibles niveles: *Tumoral* y *Normal*. Otra forma de entender la diferencia entre ambos tipos de factores es pensar en qué sucedería si el mismo experimento o estudio se realizara en otras condiciones diferentes, tiempo, ciudad... de tal manera que si los niveles de un factor fueran los mismos entonces el factor sería fijo (*Tipo de Tejido, Tratamiento con DFMO...*) y, por el contrario, si fueran diferentes, entonces sería aleatorio (*Gel, Paciente, Cubreobjetos...*).

Por lo tanto, un enfoque más acertado y práctico para analizar datos agrupados sería emplear *modelos lineales mixtos* en los que no solo se estiman los factores fijos (los de interés para el estudio y que siempre serán los mismos en cada realización del experimento) sino que también se controla el efecto de los factores aleatorios que dan lugar a las agrupaciones entre las observaciones. Además, estos modelos también permiten estimar no solo la variabilidad que introducen los efectos aleatorios de cada factor aleatorio, sino también el efecto *per se* sobre la variable respuesta. Para comprender mejor estos modelos, es interesante plantearlos, en una primera aproximación, como un modelado en dos etapas que les hace más sencillos de entender y, además, para facilitar su comprensión, se considerará como factor aleatorio el *Sujeto* y como factores fijos tanto aquellos que varían dentro de cada sujeto como otros que no varían dentro de cada sujeto pero que también sean de interés para el estudio:

- *Primera etapa:* en esta primera etapa se describe la variabilidad intra-sujeto, es decir, el comportamiento de cada uno de los individuos en función de variables las conocidas que sí varían dentro de cada sujeto (un ejemplo sería el tipo de tejido dentro de cada paciente). Para ello, se ajusta un modelo de regresión a cada uno de los individuos. Así, si $Y_{i,j}$ es la variable respuesta para el *i-ésimo* sujeto medido en el nivel *j* de la variable conocida para $i = 1, 2, \dots, m$ y $j = 1, 2, \dots, n_i$. Así, el vector de respuestas es Y_i para cada individuo *i*. En consecuencia, el modelo que explica la variabilidad dentro de los sujetos es (ecuación A.61):

$$Y_i = Z_i\beta_i + \epsilon_i \quad (\text{A.61})$$

siendo Z_i la matriz de orden $n_i \times q$, siendo q de covariables conocidas que varían dentro de cada sujeto, β_i el vector q -dimensional con los coeficientes de regresión sujeto-específicos y $\epsilon_i \sim N(0, R_i)$ tal que R_i es la matriz de varianzas-covarianzas.

- *Segunda etapa:* en esta segunda etapa se modeliza la variabilidad existente entre los diferentes sujetos; es decir, cómo de diferente es el comportamiento entre los diferentes sujetos. Por lo tanto, se explica la variabilidad en los coeficientes de regresión sujeto-específicos β_i a través de su modelización de estos en función de las variables explicativas de interés o efectos fijos las cuales (ecuación A.62):

$$\beta_i = K_i\beta + b_i \quad (\text{A.62})$$

siendo K_i la matriz de orden $q \times p$ de p covariables conocidas o matriz de diseño (efectos fijos), β el vector p -dimensional de parámetros de regresión desconocidos a estimar y $b_i \sim N(0, D)$ tal que D es la matriz de varianzas-covarianzas de orden $q \times q$.

Sin embargo, esta aproximación en dos etapas sigue siendo otra estrategia basada en estadísticos de resumen, pues el vector de respuestas Y_i se está resumiendo mediante β_i el cual se modela en la segunda etapa como si de la variable respuesta se tratase. Por lo tanto, para evitar que se trate de otro método más basado en estadísticos de resumen lo que se hace es combinar ambas etapas en un único modelo (ecuación A.63):

$$\begin{aligned} Y_i &= Z_i K_i \beta + Z_i b_i + \epsilon_i \\ &= X_i \beta + Z_i b_i + \epsilon_i \end{aligned} \quad (\text{A.63})$$

Por lo tanto, se obtiene un modelo en el que se tienen los efectos fijos, β , y los aleatorios b_i y ϵ_i (este último es el error experimental o aleatorio), que se corresponden con la desviación de β debido a cada individuo y al error experimental, respectivamente. Cabe destacar que la matriz de diseño X_i contiene tanto las covariables conocidas en la primera etapa que varían dentro de cada sujeto como las empleadas en la segunda etapa. Además, los componentes de la varianza son los elementos de las matrices D y R_i . En consecuencia, la matriz de varianzas-covarianzas del vector de respuestas Y_i es:

$$\text{Cov}(Y_i) = Z_i D Z_i' + R_i \quad (\text{A.64})$$

La formulación más habitual de los modelos lineales mixtos es la de Laird and Ware (1982) (ecuación A.65):

$$Y_i = X_i \beta + Z_i b_i + \epsilon_i \quad (\text{A.65})$$

donde $\epsilon_i \sim N(0, R_i)$, $b_i \sim N(0, D)$, $\text{Cov}(\epsilon_i, b_i) = 0$, $Y_i | b_i \sim N(X_i \beta + Z_i b_i, R_i)$ y $Y_i \sim N(X_i \beta, Z_i D Z_i' + R_i)$.

y que

$$\text{Cov}(Y_i) = Z_i D Z_i' + R_i = V_i \quad (\text{A.66})$$

donde V_i es una matriz normalmente definida por bloques (niveles o individuos) que puede presentar distintas estructuras (como se definió en los modelos de *MCG*): Sin estructura, Simetría Compuesta, Autorregresiva AR(1), Versiones heterogéneas, etc

En definitiva, en los modelos mixtos, a parte de estimar los coeficientes del modelo correspondientes a los efectos fijos (β), también se estiman los efectos aleatorios para cada individuo o nivel de agrupación, b_i , se consideran las estructuras de correlación entre las observaciones dentro de un mismo nivel de agrupación, $Z_i D Z_i'$ y pueden ser o no homocedásticas.

Para un mismo factor aleatorio o factor de agrupación pueden existir diferentes tipos de efectos aleatorios. Por ejemplo, en el caso de medidas repetidas a lo largo del tiempo sobre un mismo paciente, el factor aleatorio *Paciente* puede tener asociado tanto interceptos aleatorios (cómo varía su valor basal respecto del valor basal poblacional) y pendiente aleatoria (la diferente tendencia de la variable respuesta a lo largo del tiempo para un paciente respecto de la tendencia poblacional).

A.6.2.1. Inferencias

Está claro que la complejidad de los modelos lineales mixtos es mayor que la de los modelos lineales ordinarios. En particular, los modelos lineales mixtos estiman dos tipos de parámetros: fijos y aleatorios. Si bien es cierto que el interés suele centrarse en los parámetros fijos, una mala especificación de la parte aleatoria puede perjudicar a las estimaciones de la parte fija. Además, existen diferentes métodos para estimar los parámetros de estos modelos: *maxima verosimilitud (ML)* y la de *máxima verosimilitud restringida (REML)*. A continuación, se explica brevemente las inferencias sobre cada tipo de parámetro:

- *Parámetros Fijos*: autores como Bates et al. (2015), autor del paquete de R *lme4* ampliamente usado para la obtención de los modelos lineales mixtos, están de acuerdo con que la obtención de los grados de libertad de los parámetros fijos no es fiable en el contexto de este tipo de modelos, pues, a diferencia de los modelos lineales ordinarios en los que los grados de libertad son $n - p$, esto no es tan sencillo en los modelos lineales mixtos. Por lo tanto, en estos últimos se emplean métodos aproximados para estimar los grados de libertad. Por lo tanto, la distribución de los estadísticos *t* o *F* será poco fiable y, por ende, también lo será el contraste de significación de los parámetros fijos basados en dicho estadístico. Por ello, la estrategia seguida por muchos es generar diferentes modelos de diferente complejidad en la parte fija (con diferentes factores fijos) y, a continuación, compararlos mediante diferentes criterios como *AIC* o *BIC*. No obstante, estos modelos que se comparan han de tener la misma parte aleatoria y, además, el mismo número de observaciones. Afortunadamente, no es necesario que sean modelos anidados. Es importante tener en cuenta que, como sugieren

algunos autores tales como R. (2013), ningún criterio de información es mejor que otro y, por ende, es aconsejable el uso de diferentes criterios de información y escoger aquel modelo que más adecuado sea según la mayoría de los criterios de información considerados.

- *Parámetros Aleatorios*: del mismo modo que para la parte fija, para evaluar la significación de los parámetros aleatorios se construyen diferentes modelos de diferente complejidad manteniendo la parte fija sin modificar y variando la parte aleatoria para, posteriormente, comparar los diferentes modelos mediante diferentes criterios como *AIC* o *cAIC*, versión mejorada de *AIC* para los modelos mixtos. En particular, el contraste de interés en la parte aleatoria es que la varianza del factor aleatorio sea cero (hipótesis nula) frente a que sea mayor de cero (hipótesis alternativa). Cabe destacar que no es lo mismo que la hipótesis nula de los parámetros de la covarianza esté en el límite del espacio de los posibles valores de dicha covarianza (problema del límite), cero en este caso, o no:

 - En el caso de que no lo esté, como puede pasar por ejemplo al comparar un modelo con varianza residual heterogénea frente a otro con varianza homogénea, o cuando se contrasta que la covarianza entre dos efectos aleatorios sea cero (pues en este caso el cero no está en el límite), el test de razón de verosimilitudes se distribuye asintóticamente como una χ^2_{gdl} , donde los grados de libertad (gdl) son la diferencia entre el número de parámetros de covarianza del modelo completo y el modelo anidado.
 - Por el contrario, cuando la hipótesis nula de los parámetros de covarianza está en el límite del espacio parametral, como es el caso de contrastar si un efecto aleatorio ha de permanecer o no en el modelo, se contrasta que su varianza sea igual a cero. Este valor, una varianza nula, está en el límite de todos los posibles valores que puede tomar este parámetro, pues ha de ser estrictamente igual o mayor que cero. En estas situaciones, la forma exacta de la distribución nula es difícil de obtener, pues estaría dada por una mixtura de varias distribuciones χ^2 , donde los resultados son obtenidos asumiendo que los errores son independientes y homocedásticos. Se ha visto que los test estadísticos tienen una distribución asintótica que son mixturas de χ^2 , como por ejemplo la distribución nula $0.5\chi_0^2 + 0.5\chi_1^2$ si el modelo sólo tiene un efecto aleatorio se desea contrastar la varianza de este o $0.5\chi_1^2 + 0.5\chi_2^2$ cuando también se contrasta la varianza de un único efecto aleatorio en un modelo con dos efectos aleatorios. En consecuencia, en el caso de emplear el test de verosimilitud para evaluar la significación de la parte aleatoria aparece el *problema del límite* ya que se contrasta que la varianza sera diferente de cero y, por ello, los p valores obtenidos en el test que sean cercanos al nivel de significación han de ser tomados con cautela. Esto se debe a que el test de verosimilitud parte de la base de que no hay un límite en el espacio del parámetro estimado, por lo que hacer contrastes en el límite de la distribución del parámetro puede ser peligroso. Así, se proponen técnicas *bootstrapping* o las estimaciones *REML* frente a *ML*, pues corrigen en cierto modo el problema del límite o de la estimación de la varianza en el límite de su distribución.

A.6.2.2. Selección del Modelo

Tras entender la dificultad de las inferencias en este tipo de modelos, es hora de plantear cómo seleccionar el mejor modelo que, al igual que en los modelos ordinarios, también se busca encontrar el más simple que explique lo mejor posible los datos observados. Existen diferentes estrategias para ello. En particular, en la presente tesis se ha seguido la siguiente:

1. Se estima el modelo *sobre-óptimo*, aquel que contiene todas las posibles variables explicativas y sus interacciones. No obstante, no siempre es posible generar dicho modelo debido a que no haya grados de libertad suficientes debido a un número bajo de observaciones. En tal caso, se genera el modelo más complejo posible considerando sólo aquellas posibles variables explicativas que, según el conocimiento del investigador, sean las más plausibles a la hora de explicar la variable respuesta.
2. Se busca la estructura aleatoria más adecuada a partir del modelo *sobre-óptimo*, para lo que se generan diferentes modelos con la misma parte fija sobre-óptima y diferentes estructuras aleatorias. Para ello, se emplea el método de estimación *REML* con el fin de abordar el problema de la estimación de la varianza en el límite de su distribución (se contrasta que sea igual a cero). A continuación, se comparan los diferentes modelos generados mediante los valores *cAIC*. Cabe destacar que se escoge el método de estimación *REML* frente a *ML* debido a que el primero reduce el sesgo inherente en las estimaciones *ML* de los parámetros de la covarianza.

En ocasiones puede suceder que se obtenga una varianza estimada para un efecto aleatorio igual a 0 o correlación ± 1 entre los efectos aleatorios de un mismo factor aleatorio, lo que puede deberse a un reducido número de efectos aleatorios (niveles del factor aleatorio) o a estructuras aleatorias demasiado complejas. En consecuencia, se obtiene una matriz de varianzas-covarianzas de los efectos aleatorios que no es de rango máximo y, en definitiva, no habrá suficientes datos para estimar de forma fiable todos los parámetros. Así, algunos autores como Barr et al. (2013) sugieren comenzar por la estructura aleatoria más compleja e ir eliminando efectos y factores aleatorios cuando los algoritmos empleados no converjan o cuando haya algún problema de singularidad de la matriz de varianzas-covarianzas de la parte aleatoria como consecuencia de varianzas 0 y/o correlaciones entre efectos aleatorios ± 1 . Otros como Matuschek et al. (2017), por el contrario, sugieren emplear modelos simplificados a priori siempre que sea posible. En cualquier caso, los factores y efectos aleatorios asociados que den alguno de estos problemas se eliminarán del modelo.

3. Tras seleccionar la estructura aleatoria óptima, se evalúa la estructura fija óptima. Para ello, de forma similar a como se procede para seleccionar la parte aleatoria, se generan diferentes modelos en los que se varía la parte fija y se mantiene la parte aleatoria sin modificar entre ellos (la parte aleatoria seleccionada en el paso 2), y se emplea el método de estimación *ML*. A continuación, se comparan todos los modelos mediante *AIC*.

4. Por último, el modelo seleccionado en el paso 3 se reestima empleando el método *REML*.

Por lo tanto, para inferir si un efecto, tanto fijo como aleatorio, o incluso considerar diferentes estructuras para las matrices Σ , D que tengan en cuenta la heterocedasticidad, correlación entre los errores o entre los efectos aleatorios, se recurre a la comparación de modelos mediante el uso de criterios tales como la *AIC* $AIC = -2\log(\text{likelihood}) + 2k$, siendo k el número de parámetros estimados. Algunos autores argumentan que el criterio de *AIC* marginal es sesgado y conviene emplear en su lugar el criterio de *AIC* condicional. No obstante, es costoso computacionalmente, y para llevar a cabo una selección de diferentes modelos podría resultar inviable en cuestión de tiempo.

A.7. Modelos Predictivos: Machine Learning

Mediante diferentes métodos o algoritmos de *machine learning* o *aprendizaje automático* es posible generar modelos predictivos y clasificar, a partir de diferentes variables, muestras desconocidas en diferentes clases; por ejemplo, una muestra desconocida en tumoral o sana. Aunque existen diferentes algoritmos de *machine learning*, en la presente tesis, en concreto, se ha empleado el algoritmo de árboles de clasificación para, de entre todas las componentes principales consideradas como variables explicativas en la clasificación entre muestras sanas y tumorales, identificar cuál o cuales son las más importantes en dicha clasificación y así, posteriormente, correlacionar dichas componentes más importantes con las variables originales; es decir, con la expresión de los genes estudiados (ver apartado “A.5. Coeficiente de correlación de Pearson”). Un libro interesante y didáctico para la implementación de este tipo de algoritmos con R es Lantz (2013).

A.7.1. Algoritmos de Árboles de Clasificación y Bosques Aleatorios (Random Forest)

Los *árboles de clasificación (CT)* son un método de clasificación que trata de buscar relaciones entre las variables explicativas y una variable respuesta cualitativas. No obstante, también es posible emplear este tipo de algoritmos para variables respuesta continuas, de tal manera que este tipo de algoritmos se conocen como *árboles de regresión*. Estos algoritmos se basan en la idea de “divide y vencerás,” pues el procedimiento, a *grosso modo*, consiste en ir dividiendo el conjunto de observaciones en grupos más pequeños de tal manera que la heterogeneidad para cada uno de los grupos que se van formando sea mínima. Para tal fin, se comienza por el nodo raíz, que es el conjunto de datos original, y este se va dividiendo en función de la variable explicativa que se considere (según el algoritmo) que es la más explicativa (más importante). Así, se forman las ramas del árbol, cada una con un subconjunto de datos del conjunto original. Posteriormente, sobre cada una de las ramas formadas se aplica nuevamente este procedimiento, para lo que se escogen las

siguientes variables más predictivas, como si de un árbol en crecimiento se tratase. El procedimiento finaliza cuando, según un criterio de parada, el algoritmo decide que la homogeneidad de cada nuevo grupo es suficiente.

Es interesante que estos algoritmos no solo permiten clasificar las observaciones del conjunto de datos original según las variables explicativas de interés, sino que también permiten encontrar un orden de importancia en las variables explicativas consideradas. Por lo tanto, éstos algoritmos permiten encontrar un modelo e interpretarlo, lo que supone una característica importante que los diferencia de otros algoritmos de *machine learning* conocidos como *algoritmos de caja negra* (redes neuronales artificiales, por ejemplo).

Como criterio para elegir la variable más predictiva (más importantes) existen las conocidas como *medidas de impureza del nodo*. De entre todas las medidas que existen, cuando la variable es dicotómica la más empleada es el *índice de Gini* (ecuación A.67):

$$I(\Omega) = 2p_{\Omega}(1 - 2p_{\Omega}) \quad (\text{A.67})$$

donde Ω es el *nodo padre* (*raíz al principio*), $2p_{\Omega} = P(Y = 1|\Omega)$ la probabilidad condicional de que $Y = 1$ si pertenece al nodo Ω . También existe la *ganancia de información* que emplea el concepto de *entropía de Shannon* (ecuación A.68), siendo ésta última 0 cuando el grupo es totalmente homogéneo y 1 (máxima) cuando hay el mismo número de observaciones de cada categoría:

$$S = \sum_{i=1}^{\text{categorías}} -p_i \log_2(p_i) \quad (\text{A.68})$$

siendo p_i la proporción de observaciones de la categoría i en la muestra. Por lo tanto, si se divide el conjunto según diferentes variables explicativas, se obtiene un valor de S para cada rama, se suman proporcionalmente las entropías de cada rama, S_{total} y se resta de S_{original} ; como resultado se obtiene el valor de *ganancia de información*, de tal manera que aquella variable explicativa con mayor *ganancia de información* es la seleccionada para ese nodo. Finalmente, los árboles pueden ser podados, con el fin de reducir el número de ramas y evitar problemas como posibles sobreajustes.

Entre las diferentes ventajas de este tipo de algoritmos, destaca que pueden emplearse en multitud de problemas, que funciona bien con datos faltantes, que no importa el tipo de variable explicativa, que el tamaño reducido del conjunto de datos no supone un problema y que el modelo generado es de fácil interpretación, pues, recordemos, se obtiene un orden en las variables explicativas según su importancia. Como desventajas, cabe destacar que existe un sesgo hacia variables con muchas categorías, problemas de sobreajuste, poco robusta a cambios en el conjunto de entrenamiento y, aunque fáciles de interpretar con pocas variables, a medida que el número aumenta también lo hace la complejidad en su interpretación.

Una generalización de los *CT* son los *bosques aleatorios (random forest, RF)*, que se basan en generar un conjunto de *CT* y sólo un subconjunto aleatorio de las variables explicativas en cada *CT* (cuyo tamaño puede ser variado por el analista). Este algoritmo facilita el uso de conjuntos grandes de datos y afronta el problema de *curse of dimensionality* o *problema de la dimensionalidad*², como evitar el efecto negativo de la multicolinealidad y mejorar la precisión de los resultados. En otras palabras, el algoritmo selecciona un subconjunto de las observaciones mediante muestreo con reemplazo, se generan diferentes *CT*, cada uno con un subconjunto de variables explicativas que se usan para predecir sobre un nuevo set de datos. Con la información obtenida de los diferentes *CT*, se considera que una observación pertenece a la categoría *i* si la mayoría de los *CT* así lo predicen. Este tipo de algoritmo tiene las ventajas de los *CT* salvo que el resultado final no es un *CT* interpretable, sino una medida de la importancia de la variable explicativa sobre la variable respuesta.

²Fenómenos que tienen lugar cuando se analizan datos en espacios de elevada dimensionalidad que no tienen lugar cuando se analizan datos de menor dimensionalidad.

Apéndice B

Análisis de datos transcriptómicos

Los datos que se presentan en este apéndice se corresponden con los análisis transcriptómicos presentados en la presente tesis. En primer lugar, para poder seguir de una forma más clara los análisis realizados y comprender en mayor medida los resultados, se procede a presentar un caso práctico en el que se explica, muy brevemente, los pasos seguidos y las razones, así como la interpretación resumida de las figuras, gráficos, tablas y resultados obtenidos. En segundo lugar, para el resto de análisis transcriptómicos realizados, habrá un subapartado en el apéndice en el que se mostrará el material suplementario que, con el fin de no entorpecer la lectura y asimilación de los resultados, se ha evitado mostrar en el apartado de resultados pero que, sin embargo, es importante a la hora de validar los resultados presentados.

B.1. Análisis de datos transcriptómicos basados en RNA-seq

En este apartado se presentan los datos complementarios (análisis exploratorio y control de calidad, normalizaciones, transformaciones...) a los análisis de los experimentos basados en *RNA-seq* mostrados en el apartado “7.1.2 Análisis transcriptómico de las líneas celulares de cáncer de colon HT29 y de colon normal NCM460” .

B.1.1. Control de calidad y preprocesado de los datos transcriptómicos de los experimentos basados en RNA-seq para la comparación entre las líneas celulares HT29 y NCM460

Aquí se presenta el control de calidad y detección de *outliers* mediante métodos generales (figura B.1), a partir de los cuales no se detectó ninguna muestra atípica; el filtrado mediante el método *K sobre A*, en el que un gen permanece en la matriz de expresión si al menos *K* muestras ≥ 4 en este caso por ser el número de réplicas en cada condición- presentan una expresión superior a un valor mínimo

-en este caso el primer cuartil de todos los valores de expresión-; la normalización de los datos, pues, para poder llevar a cabo cualquier análisis sobre las muestras, éstas han de ser comparables entre ellas, y la transformación de los datos discretos. No obstante, hay que distinguir entre dos tipos de transformación diferentes que pueden hacerse; por un lado, para realizar análisis exploratorios de datos de RNA-seq se puede emplear la normalización *rlog* (Michael I. Love, Huber, and Anders (2014b)), implementado por el paquete *DESeq2* (Michael I. Love, Huber, and Anders (2014a)), que realiza una transformación logarítmica en base 2 de los datos de expresión y minimiza las diferencias entre las muestras para aquellos genes que tengan un número bajo de lecturas (ruido de Poisson) y lo normaliza frente al tamaño de la librería (figura B.2). Por otro lado, a la hora de realizar análisis de expresión diferencial y de significación biológica, pueden usarse diferentes métodos, tales como *DESeq2* (Michael I. Love, Huber, and Anders (2014a)) o *edgeR* (Robinson, McCarthy, and Smyth (2010), McCarthy, Chen, and Smyth (2012)); sin embargo, en este caso ha usado el método *limma* (Ritchie et al. (2015)), para el cual pues es necesario, en primer lugar, realizar una normalización de los datos mediante el método TMM (Robinson and Oshlack (2010)) para obtener los factores de escala que permiten convertir los datos de conteo originales en aquellos que tienen unos tamaños de librería efectivos. En segundo lugar, una transformación de estos datos normalizados empleando el método *voom* (Law et al. (2014)) -diferente al *rlog* (Michael I. Love, Huber, and Anders (2014b)) empleado para el análisis exploratorio- el cual lleva a cabo una transformación logarítmica de los datos de conteo en base 2 (logCMP) y una estimación de la relación entre la media y la varianza que se usará posteriormente para calcular los pesos o ponderaciones a nivel observacional para poder aplicar sobre estos datos el modelo lineal *limma* (B.2).

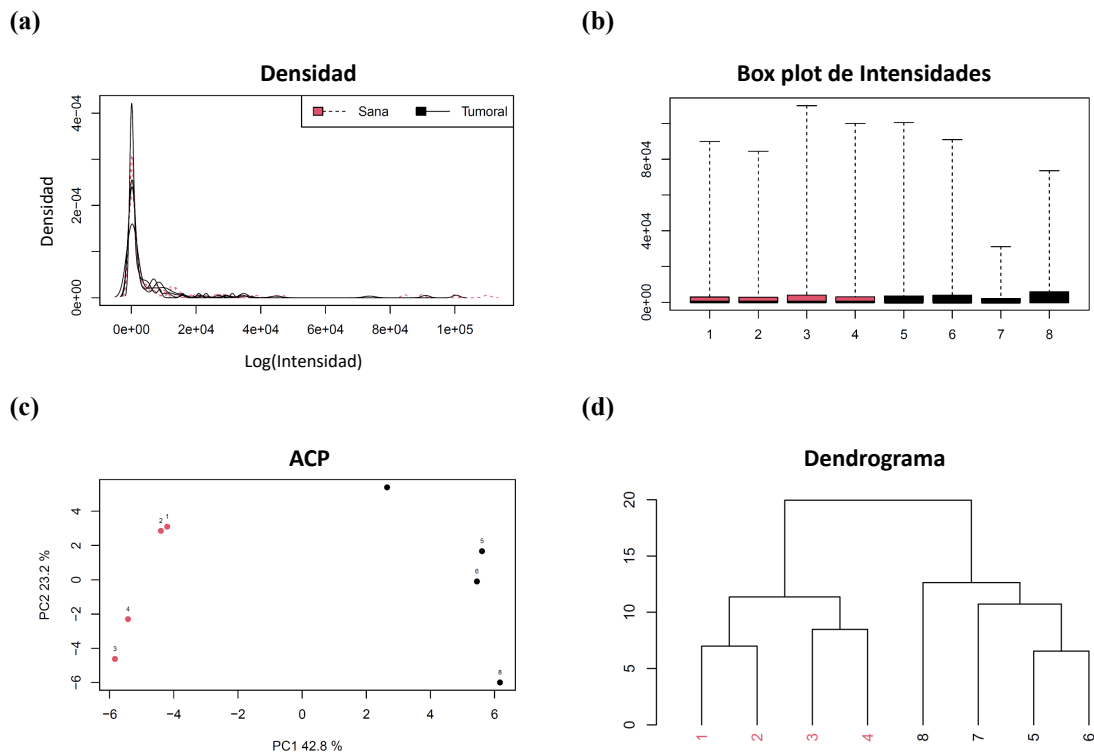


Figura B.1: Control de calidad y detección de *outliers* con métodos generales. b) Gráfico de densidad; b) Gráfico de densidad; (c) Análisis de componentes principales coloreada en función de los factores de interés *Línea* y *Tratamiento*; Análisis Clúster Jerárquico, se ha usado el método de Ward y como medida de disimilaridad se ha usado la distancia euclídea. Las muestras se han coloreado en función del factor *Tumor* (negro para muestras tumorales y rojo para normales).

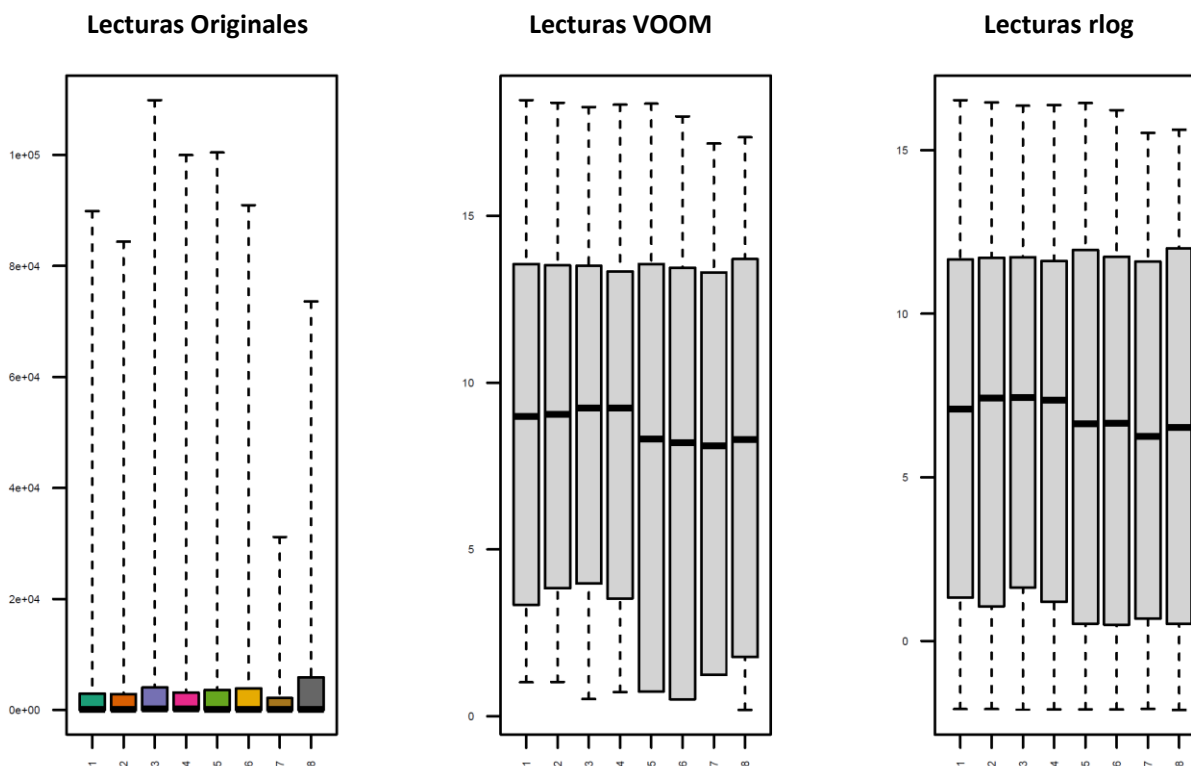


Figura B.2: Normalización y transformación de los datos de conteo de RNA-seq. Izquierda) lecturas originales; Centro) normalización y transformación rlog; Derecha) normalización TMM y transformación voom.

B.1.2. Análisis de enriquecimiento de genes de la expresión diferencial entre las líneas celulares HT29 y NCM460 de los experimentos basados en RNA-seq

En la tabla B.1 se muestra un resumen numérico del análisis de enriquecimiento realizado a partir del análisis de expresión diferencial entre las células HT29y NCM460. Para ello, se han cruzado estos datos con 8 colecciones de conjuntos de genes que han sido empeladas a lo largo de la presente tesis, de tal manera que en dicha tabla se muestra no sólo cuántos conjuntos de genes rechazan alguna de las dos hipótesis nulas evaluadas (autocontenida y competitiva), sino también cuántos de aquellos que rechazan alguna de dichas hipótesis nulas pertenecen a las bases de datos relacionadas con el cáncer (MCC y MOG) y/o contienen en su nombre alguna de las siguientes palabras clave de interés para la la temática de la tesis: *calcium*, *mitochondria*, *reticulum*, *cancer*, *wnt* y *polyamines*. Posteriormente, en las tablas B.2-B.10 se muestra en detalle el análisis de enriquecimiento de conjuntos de genes para aquellos conjuntos con un FDR < 0.1 divididos en diferentes tablas en función tanto de la base de datos a la que pertenecen como del método empleado para su análisis.

Tabla B.1: Análisis de enriquecimiento de genes para la expresión diferencial entre las células HT29 y NCM460 llevada a cabo con Illumina (RNA-seq).

| Base de Datos | Conjuntos | | Conjuntos Significativos (FDR <0.05) | | | | | | | |
|---------------|--------------|---------------------|--------------------------------------|-----------|---------------------|--------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | Métodos | Totales | Calcio ¹ | Mitocondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO BP | 12342 | 146 | CAMERA ORA | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO MF | 4245 | 66 | CAMERA ORA | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| MH | 50 | 0 | CAMERA ORA | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| MCDB | 6290 | 12 | CAMERA ORA | 4 | 0 | 0 | 0 | 1 | 0 | 0 |
| MCC | 858 | 0 | CAMERA ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MOG | 189 | 0 | CAMERA ORA | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MIM | 5219 | 0 | CAMERA ORA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 29537 | 225 | CAMERA ORA | 13 | 0 | 0 | 0 | 1 | 1 | 0 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Omncogenic; MIM: Msigdb Immunology definidas por nuestro grupo.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

Tabla B.2: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA).

| Conjuntos | nºGenes | Sentido | P-valor | FDR |
|--|---------|---------|---------|-----------|
| GO:0014807 regulation of somitogenesis | 6 | -1 | 2e-07 | 0.0009764 |

Tabla B.3: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOMF (CAMERA).

| Conjuntos | nºGenes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| GO:0008046 axon guidance receptor activity | 6 | -1 | 1.30e-06 | 0.0018881 |
| GO:0003836 beta-galactoside (CMP) alpha-2,3-sialyltransferase activity | 3 | 1 | 5.09e-05 | 0.0358590 |
| GO:1903136 cuprous ion binding | 5 | 1 | 8.86e-05 | 0.0416125 |

Tabla B.4: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA).

| Conjuntos | nºGenes | Sentido | P-valor | FDR |
|---|---------|---------|----------|-----------|
| M17954 STARK HYPOCAMPUS 22Q11 DELETION DN | 19 | 1 | 1.40e-06 | 0.0048470 |
| M18387 KORKOLA SEMINOMA UP | 42 | -1 | 1.60e-06 | 0.0048470 |
| M8556 OSADA ASCL1 TARGETS UP | 34 | -1 | 2.01e-05 | 0.0396037 |
| M8293 NIKOLSKY BREAST CANCER 8P12 P11 AMPLICON | 39 | -1 | 2.92e-05 | 0.0431503 |
| M3010 HECKER IFNB1 TARGETS | 64 | 1 | 5.71e-05 | 0.0675036 |
| M39364 WP DEVELOPMENT AND HETEROGENEITY OF THE ILC FAMILY | 17 | -1 | 1.16e-04 | 0.0792074 |
| M10276 LEE NEURAL CREST STEM CELL DN | 86 | -1 | 1.24e-04 | 0.0792074 |
| M5091 EINAV INTERFERON SIGNATURE IN CANCER | 27 | 1 | 1.27e-04 | 0.0792074 |
| M3218 MOSERLE IFNA RESPONSE | 31 | 1 | 1.48e-04 | 0.0792074 |
| M8943 TING SILENCED BY DICER | 24 | 1 | 1.60e-04 | 0.0792074 |
| M18724 XU AKT1 TARGETS 6HR | 24 | 1 | 1.62e-04 | 0.0792074 |
| M39869 WP SOMITOGENESIS IN THE CONTEXT OF SPONDYLOCOSTAL DYSOSTOSIS | 7 | -1 | 1.83e-04 | 0.0792074 |
| M16227 REACTOME CHOLESTEROL BIOSYNTHESIS | 24 | 1 | 1.87e-04 | 0.0792074 |
| M18837 BIOCARTA EGFR SMRTE PATHWAY | 10 | 1 | 1.93e-04 | 0.0792074 |
| M18660 ROETH TERT TARGETS UP | 14 | 1 | 2.01e-04 | 0.0792074 |
| M4551 SANA RESPONSE TO IFNG UP | 62 | 1 | 2.58e-04 | 0.0953149 |
| M12730 KORKOLA EMBRYONAL CARCINOMA UP | 36 | -1 | 2.84e-04 | 0.0987485 |

Tabla B.5: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA).

| Conjuntos | n ^o Genes | n ^o Genes Significativos | P-valor | FDR |
|---|----------------------|-------------------------------------|----------|-----------|
| M13661 CREIGHTON ENDOCRINE THERAPY RESISTANCE 5 | 458 | 395 | 3.00e-07 | 0.0018324 |
| M1741 ZHENG BOUND BY FOXP3 | 407 | 349 | 5.10e-06 | 0.0116250 |
| M2157 DUTERTRE ESTRADIOL RESPONSE 24HR DN | 475 | 404 | 5.90e-06 | 0.0116250 |
| M1804 WANG SMARCE1 TARGETS UP | 227 | 200 | 1.33e-05 | 0.0196541 |
| M4306 ONDER CDH1 TARGETS 2 DN | 399 | 339 | 3.99e-05 | 0.0334240 |
| M1684 DURCHDEWALD SKIN CARCINOGENESIS DN | 235 | 205 | 4.65e-05 | 0.0334240 |
| M7514 BERENJENO TRANSFORMED BY RHOA DN | 360 | 307 | 4.94e-05 | 0.0334240 |
| M2413 PEDERSEN METASTASIS BY ERBB2 ISOFORM 7 | 365 | 311 | 5.10e-05 | 0.0334240 |
| M38972 JINESH BLEBBISHIELD TO IMMUNE CELL FUSION PSHMS UP | 354 | 302 | 5.32e-05 | 0.0334240 |
| M929 SABATES COLORECTAL ADENOMA UP | 114 | 104 | 5.94e-05 | 0.0334240 |
| M9192 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL DN | 400 | 339 | 6.22e-05 | 0.0334240 |
| M7094 MARKEY RB1 ACUTE LOF DN | 173 | 153 | 8.71e-05 | 0.0429040 |
| M3335 HELLER HDAC TARGETS UP | 284 | 244 | 1.01e-04 | 0.0459239 |
| M5490 WELCSH BRCA1 TARGETS UP | 183 | 161 | 1.10e-04 | 0.0464436 |
| M3015 HUANG DASATINIB SENSITIVITY UP | 76 | 71 | 1.28e-04 | 0.0491352 |
| M7396 SWEET LUNG CANCER KRAS DN | 287 | 246 | 1.33e-04 | 0.0491352 |
| M4665 BAELDE DIABETIC NEPHROPATHY DN | 401 | 338 | 1.60e-04 | 0.0525768 |
| M19541 REN ALVEOLAR RHABDOMYOSARCOMA DN | 380 | 321 | 1.64e-04 | 0.0525768 |
| M798 FLORIO NEOCORTEX BASAL RADIAL GLIA UP | 197 | 172 | 1.69e-04 | 0.0525768 |
| M15368 ROSS AML WITH PML RARA FUSION | 59 | 56 | 2.02e-04 | 0.0583390 |
| M2466 DELACROIX RAR BOUND ES | 403 | 339 | 2.17e-04 | 0.0583390 |
| M39508 WP TYROBP CAUSAL NETWORK | 42 | 41 | 2.26e-04 | 0.0583390 |
| M18149 RUTELLA RESPONSE TO HGF UP | 356 | 301 | 2.27e-04 | 0.0583390 |
| M10702 JAEGER METASTASIS DN | 199 | 173 | 2.71e-04 | 0.0664396 |
| M38971 JINESH BLEBBISHIELD TRANSFORMED STEM CELL SPHERES UP | 159 | 140 | 2.81e-04 | 0.0664396 |
| M12602 WOO LIVER CANCER RECURRENCE UP | 91 | 83 | 3.45e-04 | 0.0782760 |
| M8451 SENESE HDAC3 TARGETS UP | 438 | 366 | 3.71e-04 | 0.0782760 |
| M1919 COLINA TARGETS OF 4EBP1 AND 4EBP2 | 309 | 262 | 3.88e-04 | 0.0782760 |
| M14801 HOOI ST7 TARGETS DN | 103 | 93 | 3.90e-04 | 0.0782760 |
| M1461 PETROVA ENDOTHELIUM LYMPHATIC VS BLOOD UP | 109 | 98 | 4.03e-04 | 0.0782760 |

Tabla B.5: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA) (continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M289 OISHI CHOLANGIOMA STEM CELL LIKE DN | 234 | 201 | 4.14e-04 | 0.0782760 |
| M17572 SMID BREAST CANCER LUMINAL B DN | 365 | 307 | 4.31e-04 | 0.0782760 |
| M2351 TORCHIA TARGETS OF EWSR1 FLI1 FUSION DN | 271 | 231 | 4.37e-04 | 0.0782760 |
| M17299 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL UP | 430 | 359 | 4.82e-04 | 0.0837971 |
| M1077 REACTOME PLATELET ACTIVATION SIGNALING AND AGGREGATION | 199 | 172 | 5.45e-04 | 0.0897833 |
| M11896 LINDGREN BLADDER CANCER CLUSTER 2A DN | 131 | 116 | 5.52e-04 | 0.0897833 |
| M16407 IVANOVA HEMATOPOIESIS STEM CELL LONG TERM | 171 | 149 | 5.62e-04 | 0.0897833 |
| M14127 HELLER HDAC TARGETS SILENCED BY METHYLATION UP | 376 | 315 | 6.49e-04 | 0.0989714 |
| M259 MANALO HYPOXIA UP | 181 | 157 | 6.53e-04 | 0.0989714 |

Tabla B.6: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MH (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|---|---------|---------|----------|--------|
| M5895 HALLMARK WNT BETA CATENIN SIGNALING | 41 | -1 | 0.000282 | 0.0141 |

Tabla B.7: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MH (ORA).

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--------------------------------|---------|------------------------|---------|-------|
| M5893 HALLMARK MITOTIC SPINDLE | 196 | 175 | 0.00122 | 0.061 |

Tabla B.8: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| M41023 ERWIN COHEN BLOOD VACCINE TC 83 AGE 23 48YO VACCINATED VS CONTROL 7DY UP | 64 | 1 | 1.41e-05 | 0.0314943 |
| M41109 HOWARD NEUTROPHIL INACT MONOV INFLUENZA A INDONESIA 05 2005 H5N1 AGE 18 49YO 3DY UP | 100 | 1 | 1.54e-05 | 0.0314943 |
| M41047 QUEREC PBMC YF 17D VACCINE AGE 18 45YO 3DY UP | 40 | 1 | 2.44e-05 | 0.0314943 |
| M41048 QUEREC PBMC YF 17D VACCINE AGE 18 45YO 7DY UP | 40 | 1 | 2.44e-05 | 0.0314943 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA).

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|---|---------|------------------------|----------|-----------|
| M7635 GSE21360 TERTIARY VS QUATERNARY MEMORY CD8 TCELL UP | 157 | 145 | 2.00e-07 | 0.0010687 |
| M7626 GSE21360 NAIVE VS PRIMARY MEMORY CD8 TCELL UP | 139 | 129 | 5.00e-07 | 0.0011849 |
| M3236 GSE13306 RA VS UNTREATED MEM CD4 TCELL UP | 167 | 151 | 3.80e-06 | 0.0065398 |
| M6781 GSE3203 INFLUENZA INF VS IFNB TREATED LN BCELL DN | 172 | 154 | 1.28e-05 | 0.0152825 |
| M5431 GSE3982 MAST CELL VS MAC DN | 165 | 148 | 1.48e-05 | 0.0152825 |
| M6514 GSE5542 UNTREATED VS IFNA TREATED EPITHELIAL CELLS 6H UP | 157 | 141 | 2.00e-05 | 0.0172100 |
| M4553 GSE24142 EARLY THYMIC PROGENITOR VS DN3 THYMOCYTE UP | 162 | 145 | 2.39e-05 | 0.0176280 |
| M4606 GSE24634 TREG VS TCONV POST DAY7 IL4 CONVERSION DN | 160 | 143 | 3.28e-05 | 0.0211683 |
| M6200 GSE2405 HEAT KILLED LYSATE VS LIVE A PHAGOCYTOPHILUM STIM NEUTROPHIL 24H UP | 171 | 152 | 3.81e-05 | 0.0215813 |
| M9574 GSE42021 TREG VS TCONV PLN DN | 188 | 166 | 4.18e-05 | 0.0215813 |
| M41152 FLETCHER PBMC BCG 10W INFANT BCG STIMULATED VS UNSTIMULATED 10W DN | 216 | 189 | 4.60e-05 | 0.0215907 |
| M8208 GSE27670 CTRL VS BLIMP1 TRANSDUCED GC BCELL DN | 175 | 155 | 5.08e-05 | 0.0218567 |
| M8820 GSE37605 FOXP3 FUSION GFP VS IRES GFP TREG NOD DN | 131 | 118 | 6.86e-05 | 0.0261700 |
| M4437 GSE22886 IGA VS IGM MEMORY BCELL UP | 161 | 143 | 7.10e-05 | 0.0261700 |
| M3918 GSE17721 LPS VS PAM3CSK4 6H BMDC UP | 172 | 152 | 7.93e-05 | 0.0261700 |
| M9124 GSE40273 XBP1 KO VS WT TREG DN | 166 | 147 | 8.11e-05 | 0.0261700 |
| M4951 GSE29618 PDC VS MDC UP | 165 | 146 | 9.42e-05 | 0.0269835 |
| M6522 GSE5542 UNTREATED VS IFNG TREATED EPITHELIAL CELLS 24H UP | 165 | 146 | 9.42e-05 | 0.0269835 |
| M5726 GSE7764 IL15 NK CELL 24H VS SPLENOCYTE DN | 182 | 160 | 9.93e-05 | 0.0269835 |
| M5767 GSE8515 IL1 VS IL6 4H STIM MAC DN | 133 | 119 | 1.34e-04 | 0.0345921 |
| M5774 GSE9006 HEALTHY VS TYPE 1 DIABETES PBMC AT DX UP | 179 | 157 | 1.52e-04 | 0.0350247 |
| M9751 GSE43863 TFH VS LY6C LOW CXCR5NEG EFFECTOR CD4 TCELL DN | 150 | 133 | 1.54e-04 | 0.0350247 |
| M7453 GSE21670 STAT3 KO VS WT CD4 TCELL TGFB IL6 TREATED DN | 161 | 142 | 1.70e-04 | 0.0350247 |
| M8220 GSE22443 NAIVE VS ACT AND IL2 TREATED CD8 TCELL DN | 155 | 137 | 1.74e-04 | 0.0350247 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M4574 GSE24142 DN2 VS DN3 THYMOCYTE FETAL UP | 166 | 146 | 1.89e-04 | 0.0350247 |
| M7525 GSE21546 SAPIA KO VS SAPIA KO AND ELK1 KO ANTI CD3 STIM DP THYMOCYTES UP | 183 | 160 | 1.91e-04 | 0.0350247 |
| M6212 GSE2405 0H VS 3H A PHAGOCYTOPHILUM STIM NEUTROPHIL UP | 160 | 141 | 1.96e-04 | 0.0350247 |
| M4946 GSE29618 MONOCYTE VS PDC DN | 171 | 150 | 2.09e-04 | 0.0350247 |
| M8186 GSE25677 MPL VS R848 STIM BCELL UP | 136 | 121 | 2.14e-04 | 0.0350247 |
| M4970 GSE29618 PDC VS MDC DAY7 FLU VACCINE UP | 165 | 145 | 2.18e-04 | 0.0350247 |
| M7658 GSE19401 NAIVE VS IMMUNIZED MOUSE PLN FOLLICULAR DC DN | 182 | 159 | 2.19e-04 | 0.0350247 |
| M8717 GSE30971 CTRL VS LPS STIM MACROPHAGE WBP7 KO 4H DN | 159 | 140 | 2.26e-04 | 0.0350247 |
| M5185 GSE360 DC VS MAC B MALAYI HIGH DOSE UP | 176 | 154 | 2.30e-04 | 0.0350247 |
| M5722 GSE7764 NKCELL VS SPLENOCYTE DN | 170 | 149 | 2.41e-04 | 0.0350247 |
| M8222 GSE22443 NAIVE VS ACT AND IL2 TREATED CD8 TCELL UP | 141 | 125 | 2.46e-04 | 0.0350247 |
| M6197 GSE2405 0H VS 12H A PHAGOCYTOPHILUM STIM NEUTROPHIL UP | 164 | 144 | 2.51e-04 | 0.0350247 |
| M7274 GSE19923 E2A KO VS HEB AND E2A KO DP THYMOCYTE UP | 164 | 144 | 2.51e-04 | 0.0350247 |
| M9262 GSE40068 CXCR5POS BCL6POS TFH VS CXCR5NEG BCL6NEG CD4 TCELL DN | 175 | 153 | 2.64e-04 | 0.0358693 |
| M7978 GSE25123 WT VS PPARG KO MACROPHAGE ROSIGLITAZONE STIM UP | 169 | 148 | 2.77e-04 | 0.0366705 |
| M6887 GSE9601 NFKB INHIBITOR VS PI3K INHIBITOR TREATED HCMV INF MONOCYTE DN | 151 | 133 | 3.13e-04 | 0.0381183 |
| M5434 GSE3982 MAST CELL VS BCELL UP | 168 | 147 | 3.18e-04 | 0.0381183 |
| M4407 GSE22886 NAIVE CD8 TCELL VS MEMORY TCELL DN | 179 | 156 | 3.28e-04 | 0.0381183 |
| M4426 GSE22886 TCELL VS BCELL NAIVE DN | 139 | 123 | 3.32e-04 | 0.0381183 |
| M7983 GSE25088 IL4 VS IL4 AND ROSIGLITAZONE STIM MACROPHAGE DAY10 DN | 162 | 142 | 3.33e-04 | 0.0381183 |
| M5665 GSE6566 STRONG VS WEAK DC STIMULATED CD4 TCELL UP | 121 | 108 | 3.46e-04 | 0.0381183 |
| M8506 GSE23568 ID3 TRANSDUCED VS ID3 KO CD8 TCELL DN | 173 | 151 | 3.47e-04 | 0.0381183 |
| M8722 GSE30971 WBP7 HET VS KO MACROPHAGE 2H LPS STIM UP | 173 | 151 | 3.47e-04 | 0.0381183 |
| M7490 GSE21774 CD62L POS CD56 BRIGHT VS CD62L NEG CD56 DIM NK CELL UP | 144 | 127 | 3.74e-04 | 0.0395128 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | nºGenes | nºGenes Significativos | P-valor | FDR |
|---|---------|------------------------|----------|-----------|
| M4578 GSE24142 ADULT VS FETAL EARLY THYMIC PROGENITOR DN | 178 | 155 | 3.75e-04 | 0.0395128 |
| M4545 GSE24142 EARLY THYMIC PROGENITOR VS DN2 THYMOCYTE UP | 161 | 141 | 3.83e-04 | 0.0395486 |
| M6716 GSE6092 B BURGDORFERI VS B BURGDORFERI AND IFNG STIM ENDOTHELIAL CELL UP | 166 | 145 | 4.18e-04 | 0.0419733 |
| M3351 GSE14000 4H VS 16H LPS DC TRANSLATED RNA UP | 177 | 154 | 4.29e-04 | 0.0419733 |
| M9584 GSE42021 CD24HI VS CD24INT TREG THYMUS UP | 188 | 163 | 4.33e-04 | 0.0419733 |
| M8318 GSE28726 NAIVE VS ACTIVATED NKTCELL UP | 160 | 140 | 4.39e-04 | 0.0419733 |
| M6818 GSE7509 UNSTIM VS FCGR1IB STIM DC DN | 131 | 116 | 4.59e-04 | 0.0430876 |
| M8748 GSE36888 STAT5 AB KNOCKIN VS WT TCELL IL2 TREATED 17H UP | 176 | 153 | 4.89e-04 | 0.0442931 |
| M9588 GSE42021 CD24HI VS CD24LOW TREG THYMUS DN | 176 | 153 | 4.89e-04 | 0.0442931 |
| M3690 GSE17721 CTRL VS LPS 4H BMDC DN | 159 | 139 | 5.04e-04 | 0.0444018 |
| M8418 GSE29949 DC BRAIN VS MONOCYTE BONE MARROW UP | 181 | 157 | 5.25e-04 | 0.0444018 |
| M5553 GSE3982 BCELL VS TH1 DN | 164 | 143 | 5.48e-04 | 0.0444018 |
| M7234 GSE19198 1H VS 24H IL21 TREATED TCELL UP | 164 | 143 | 5.48e-04 | 0.0444018 |
| M6846 GSE7348 UNSTIM VS LPS STIM MACROPHAGE DN | 147 | 129 | 5.54e-04 | 0.0444018 |
| M6201 GSE2405 HEAT KILLED LYSATE VS LIVE A PHAGOCYTOPHILUM STIM NEUTROPHIL 24H DN | 175 | 152 | 5.58e-04 | 0.0444018 |
| M3307 GSE13485 DAY7 VS DAY21 YF17D VACCINE PBMC UP | 186 | 161 | 5.59e-04 | 0.0444018 |
| M6128 GSE2770 IL12 AND TGFB VS IL4 TREATED ACT CD4 TCELL 2H DN | 186 | 161 | 5.59e-04 | 0.0444018 |
| M9909 GSE42088 UNINF VS LEISHMANIA INF DC 8H UP | 163 | 142 | 6.27e-04 | 0.0488966 |
| M8610 GSE32986 GMCSF VS GMCSF AND CURDLAN LOWDOSE STIM DC UP | 174 | 151 | 6.36e-04 | 0.0488966 |
| M405 GSE12003 4D VS 8D CULTURE MIR223 KO BM PROGENITOR UP | 99 | 89 | 6.44e-04 | 0.0488966 |
| M7684 GSE20152 SPHK1 KO VS WT HTNFA OVERXPRESS ANKLE UP | 168 | 146 | 6.76e-04 | 0.0493757 |
| M4944 GSE29618 BCELL VS MDC DN | 179 | 155 | 6.79e-04 | 0.0493757 |
| M6183 GSE2128 C57BL6 VS NOD THYMOCYTE UP | 179 | 155 | 6.79e-04 | 0.0493757 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M9482 GSE41867 MEMORY VS EXHAUSTED CD8 TCELL DAY30 LCMV DN | 184 | 159 | 7.21e-04 | 0.0512056 |
| M9203 GSE40666 NAIVE VS EFFECTOR CD8 TCELL UP | 173 | 150 | 7.24e-04 | 0.0512056 |
| M6867 GSE4590 PRE BCELL VS VPRES POS LARGE PRE BCELL UP | 122 | 108 | 7.37e-04 | 0.0514207 |
| M3038 GOLDRATH NAIVE VS MEMORY CD8 TCELL UP | 189 | 163 | 7.62e-04 | 0.0519697 |
| M8185 GSE25677 R848 VS MPL AND R848 STIM BCELL UP | 104 | 93 | 7.65e-04 | 0.0519697 |
| M9423 GSE33292 DN3 THYMOCYTE VS TCELL LYMPHOMA FROM TCF1 KO DN | 161 | 140 | 8.18e-04 | 0.0536952 |
| M5474 GSE3982 DC VS EFF MEMORY CD4 TCELL UP | 172 | 149 | 8.23e-04 | 0.0536952 |
| M5206 GSE360 L DONOVANI VS M TUBERCULOSIS DC DN | 155 | 135 | 8.67e-04 | 0.0536952 |
| M7320 GSE19888 ADENOSINE A3R INH PRETREAT AND ACT BY A3R VS TCELL MEMBRANES ACT MAST CELL UP | 155 | 135 | 8.67e-04 | 0.0536952 |
| M3875 GSE17721 PAM3CSK4 VS CPG 24H BMDC DN | 177 | 153 | 8.75e-04 | 0.0536952 |
| M4962 GSE29618 BCELL VS MDC DAY7 FLU VACCINE DN | 177 | 153 | 8.75e-04 | 0.0536952 |
| M3599 GSE15930 NAIVE VS 48H IN VITRO STIM IL12 CD8 TCELL UP | 166 | 144 | 8.77e-04 | 0.0536952 |
| M5700 GSE7460 FOXP3 MUT VS WT ACT TCONV UP | 166 | 144 | 8.77e-04 | 0.0536952 |
| M8294 GSE23114 PERITONEAL CAVITY B1A BCELL VS SPLEEN BCELL IN SLE2C1 MOUSE UP | 160 | 139 | 9.34e-04 | 0.0536952 |
| M4650 GSE25087 TREG VS TCONV ADULT UP | 171 | 148 | 9.36e-04 | 0.0536952 |
| M6415 GSE8921 UNSTIM VS TLR1 2 STIM MONOCYTE 6H UP | 171 | 148 | 9.36e-04 | 0.0536952 |
| M6488 GSE4748 CTRL VS CYANOBACTERIUM LPSLIKE STIM DC 1H DN | 171 | 148 | 9.36e-04 | 0.0536952 |
| M7228 GSE19198 6H VS 24H IL21 TREATED TCELL DN | 171 | 148 | 9.36e-04 | 0.0536952 |
| M7229 GSE19198 CTRL VS IL21 TREATED TCELL 6H DN | 171 | 148 | 9.36e-04 | 0.0536952 |
| M9263 GSE40068 CXCR5NEG BCL6NEG CD4 TCELL VS CXCR5POS BCL6NEG TFH DN | 176 | 152 | 9.92e-04 | 0.0558447 |
| M6703 GSE3920 IFNB VS IFNG TREATED ENDOTHELIAL CELL DN | 165 | 143 | 9.99e-04 | 0.0558447 |
| M8398 GSE29949 MICROGLIA BRAIN VS CD8 POS DC SPLEEN DN | 137 | 120 | 1.02e-03 | 0.0558447 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|---|---------|------------------------|----------|-----------|
| M4250 GSE17974 2.5H VS 72H IL4 AND ANTI IL12 ACT CD4 TCELL DN | 181 | 156 | 1.05e-03 | 0.0558447 |
| M4316 GSE20366 EX VIVO VS DEC205 CONVERSION UP | 181 | 156 | 1.05e-03 | 0.0558447 |
| M2921 GSE13522 CTRL VS T CRUZI BRAZIL STRAIN INF SKIN DN | 148 | 129 | 1.05e-03 | 0.0558447 |
| M3635 GSE15930 STIM VS STIM AND IFNAB 48H CD8 T CELL DN | 170 | 147 | 1.06e-03 | 0.0558447 |
| M6511 GSE4984 LPS VS VEHICLE CTRL TREATED DC DN | 170 | 147 | 1.06e-03 | 0.0558447 |
| M4942 GSE29618 BCELL VS PDC DN | 175 | 151 | 1.12e-03 | 0.0577643 |
| M6850 GSE7348 LPS VS TOLERIZED AND LPS STIM MACROPHAGE DN | 153 | 133 | 1.13e-03 | 0.0577643 |
| M8414 GSE29949 CD8 POS DC SPLEEN VS MONOCYTE BONE MARROW UP | 153 | 133 | 1.13e-03 | 0.0577643 |
| M6484 GSE4748 CTRL VS LPS AND CYANOBACTERIUM LPSLIKE STIM DC 3H UP | 180 | 155 | 1.19e-03 | 0.0590722 |
| M9256 GSE20727 DNFB ALLERGEN VS ROS INH AND DNFB ALLERGEN TREATED DC DN | 180 | 155 | 1.19e-03 | 0.0590722 |
| M3133 GSE11864 CSF1 VS CSF1 IFNG IN MAC UP | 169 | 146 | 1.21e-03 | 0.0590722 |
| M3510 GSE14769 UNSTIM VS 360MIN LPS BMDM DN | 169 | 146 | 1.21e-03 | 0.0590722 |
| M8567 GSE26030 UNSTIM VS RESTIM TH1 DAY5 POST POLARIZATION UP | 185 | 159 | 1.24e-03 | 0.0590722 |
| M9571 GSE42021 TCONV PLN VS TREG PRECURSORS THYMUS UP | 185 | 159 | 1.24e-03 | 0.0590722 |
| M3110 GSE11057 NAIVE CD4 VS PBMC CD4 TCELL DN | 174 | 150 | 1.27e-03 | 0.0590722 |
| M4050 GSE17721 LPS VS GARDIQUIMOD 6H BMDC UP | 174 | 150 | 1.27e-03 | 0.0590722 |
| M4415 GSE22886 NAIVE TCELL VS NKCELL DN | 174 | 150 | 1.27e-03 | 0.0590722 |
| M8561 GSE26030 TH1 VS TH17 DAY5 POST POLARIZATION DN | 174 | 150 | 1.27e-03 | 0.0590722 |
| M7947 GSE25088 IL4 VS IL4 AND ROSIGLITAZONE STIM STAT6 KO MACROPHAGE DAY10 DN | 163 | 141 | 1.29e-03 | 0.0593973 |
| M8060 GSE24726 WT VS E2 2 KO PDC UP | 190 | 163 | 1.30e-03 | 0.0593973 |
| M350 GSE10147 IL3 AND HIVP17 VS IL3 AND CPG STIM PDC DN | 135 | 118 | 1.34e-03 | 0.0606879 |
| M3423 GSE14350 IL2RB KO VS WT TEFF UP | 178 | 153 | 1.51e-03 | 0.0664419 |
| M9925 GSE41176 UNSTIM VS ANTI IGM STIM BCELL 24H DN | 194 | 166 | 1.52e-03 | 0.0664419 |
| M41002 PATEL SKIN OF BODY ZOSTAVAX AGE 70 93YO VZV CHALLENGE 6HR UP | 236 | 200 | 1.53e-03 | 0.0664419 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | n ^o Genes | n ^o Genes Significativos | P-valor | FDR |
|---|----------------------|-------------------------------------|----------|-----------|
| M6872 GSE4590 SMALL VS LARGE PRE BCELL UP | 134 | 117 | 1.54e-03 | 0.0664419 |
| M6573 GSE4535 BM DERIVED DC VS FOLLICULAR DC UP | 167 | 144 | 1.55e-03 | 0.0664419 |
| M40912 FLETCHER PBMC BCG 10W INFANT PPD STIMULATED VS UNSTIMULATED 10W DN | 156 | 135 | 1.57e-03 | 0.0664419 |
| M7463 GSE21670 UNTREATED VS TGFB TREATED STAT3 KO CD4 TCELL DN | 156 | 135 | 1.57e-03 | 0.0664419 |
| M8093 GSE24972 MARGINAL ZONE BCELL VS FOLLICULAR BCELL IRF8 KO DN | 156 | 135 | 1.57e-03 | 0.0664419 |
| M4555 GSE24142 DN2 VS DN3 THYMOCYTE UP | 161 | 139 | 1.66e-03 | 0.0696795 |
| M4374 GSE20715 0H VS 24H OZONE TLR4 KO LUNG UP | 166 | 143 | 1.75e-03 | 0.0701537 |
| M5049 GSE30962 ACUTE VS CHRONIC LCMV SECONDARY INF CD8 TCELL UP | 166 | 143 | 1.75e-03 | 0.0701537 |
| M6193 GSE2405 0H VS 24H A PHAGOCYTOPHILUM STIM NEUTROPHIL DN | 166 | 143 | 1.75e-03 | 0.0701537 |
| M6450 GSE3039 B2 VS B1 BCELL UP | 166 | 143 | 1.75e-03 | 0.0701537 |
| M8128 GSE25147 UNSTIM VS HELIOBACTER PYLORI LPS STIM MKN45 CELL DN | 133 | 116 | 1.76e-03 | 0.0701537 |
| M6112 GSE2770 TGFB AND IL4 VS IL4 TREATED ACT CD4 TCELL 48H UP | 182 | 156 | 1.78e-03 | 0.0701537 |
| M6247 GSE22601 IMMATURE CD4 SINGLE POSITIVE VS DOUBLE POSITIVE THYMOCYTE DN | 182 | 156 | 1.78e-03 | 0.0701537 |
| M9860 GSE46606 IRF4 KO VS WT UNSTIM BCELL UP | 182 | 156 | 1.78e-03 | 0.0701537 |
| M7745 GSE23308 CTRL VS CORTICOSTERONE TREATED MACROPHAGE UP | 171 | 147 | 1.84e-03 | 0.0704045 |
| M8849 GSE37301 MULTIPOTENT PROGENITOR VS GRAN MONO PROGENITOR DN | 171 | 147 | 1.84e-03 | 0.0704045 |
| M6626 GSE5589 LPS AND IL10 VS LPS AND IL6 STIM IL6 KO MACROPHAGE 45MIN DN | 187 | 160 | 1.85e-03 | 0.0704045 |
| M8487 GSE29164 DAY3 VS DAY7 UNTREATED MELANOMA DN | 187 | 160 | 1.85e-03 | 0.0704045 |
| M8500 GSE29164 CD8 TCELL VS CD8 TCELL AND IL12 TREATED MELANOMA DAY7 UP | 192 | 164 | 1.92e-03 | 0.0704045 |
| M3636 GSE15930 STIM VS STIM AND TRICHOSTATINA 48H CD8 T CELL UP | 176 | 151 | 1.92e-03 | 0.0704045 |
| M6513 GSE5542 UNTREATED VS IFNG TREATED EPITHELIAL CELLS 6H DN | 176 | 151 | 1.92e-03 | 0.0704045 |
| M6574 GSE4535 BM DERIVED DC VS FOLLICULAR DC DN | 176 | 151 | 1.92e-03 | 0.0704045 |
| M7382 GSE20484 MCSG VS CXCL4 MONOCYTE DERIVED MACROPHAGE DN | 176 | 151 | 1.92e-03 | 0.0704045 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|---|---------|------------------------|----------|-----------|
| M4073 GSE17721 0.5H VS 4H LPS BMDC DN | 165 | 142 | 1.98e-03 | 0.0704045 |
| M5092 GSE32423 IL7 VS IL7 IL4 MEMORY CD8 TCELL DN | 165 | 142 | 1.98e-03 | 0.0704045 |
| M6718 GSE6092 IFNG VS IFNG AND B BURGDORFERI INF ENDOTHELIAL CELL UP | 165 | 142 | 1.98e-03 | 0.0704045 |
| M9578 GSE42021 TREG PLN VS CD24LO TREG THYMUS UP | 165 | 142 | 1.98e-03 | 0.0704045 |
| M8690 GSE34156 NOD2 LIGAND VS TLR1 TLR2 LIGAND 6H TREATED MONOCYTE UP | 181 | 155 | 2.00e-03 | 0.0704045 |
| M7322 GSE19888 ADENOSINE A3R INH VS ACT WITH INHIBITOR PRETREATMENT IN MAST CELL UP | 143 | 124 | 2.04e-03 | 0.0704045 |
| M8095 GSE24972 MARGINAL ZONE BCELL VS FOLLICULAR BCELL DN | 143 | 124 | 2.04e-03 | 0.0704045 |
| M6119 GSE2770 IL4 ACT VS ACT CD4 TCELL 6H DN | 186 | 159 | 2.08e-03 | 0.0704045 |
| M3039 GOLDRATH NAIVE VS MEMORY CD8 TCELL DN | 170 | 146 | 2.08e-03 | 0.0704045 |
| M5832 GSE9650 NAIVE VS MEMORY CD8 TCELL DN | 170 | 146 | 2.08e-03 | 0.0704045 |
| M6965 GSE9239 CTRL VS TNF INHIBITOR TREATED DC UP | 170 | 146 | 2.08e-03 | 0.0704045 |
| M7232 GSE19198 1H VS 6H IL21 TREATED TCELL DN | 170 | 146 | 2.08e-03 | 0.0704045 |
| M4568 GSE24142 EARLY THYMIC PROGENITOR VS DN2 THYMOCYTE FETAL UP | 159 | 137 | 2.13e-03 | 0.0704045 |
| M8189 GSE26488 CTRL VS PEPTIDE INJECTION OT2 THYMOCYTE DN | 159 | 137 | 2.13e-03 | 0.0704045 |
| M8709 GSE30971 CTRL VS LPS STIM MACROPHAGE WBP7 HET 2H DN | 175 | 150 | 2.17e-03 | 0.0704045 |
| M6855 GSE7768 OVA WITH LPS VS OVA WITH MPL IMMUNIZED MOUSE WHOLE SPLEEN 6H DN | 148 | 128 | 2.17e-03 | 0.0704045 |
| M8243 GSE27241 CTRL VS DIGOXIN TREATED CD4 TCELL IN TH17 POLARIZING CONDITIONS UP | 137 | 119 | 2.17e-03 | 0.0704045 |
| M9092 GSE40274 CTRL VS XBP1 TRANSDUCED ACTIVATED CD4 TCELL UP | 137 | 119 | 2.17e-03 | 0.0704045 |
| M3220 GSE13306 TREG VS TCONV LAMINA PROPRIA UP | 164 | 141 | 2.24e-03 | 0.0704045 |
| M4966 GSE29618 MONOCYTE VS MDC DAY7 FLU VACCINE UP | 164 | 141 | 2.24e-03 | 0.0704045 |
| M5455 GSE3982 MAST CELL VS TH2 UP | 164 | 141 | 2.24e-03 | 0.0704045 |
| M5703 GSE7460 FOXP3 MUT VS WT ACT WITH TGFB TCONV UP | 164 | 141 | 2.24e-03 | 0.0704045 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | nºGenes | nºGenes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M7650 GSE22140 HEALTHY VS ARTHRITIC MOUSE CD4 TCELL UP | 164 | 141 | 2.24e-03 | 0.0704045 |
| M7958 GSE25123 WT VS PPARG KO MACROPHAGE IL4 STIM DN | 164 | 141 | 2.24e-03 | 0.0704045 |
| M8707 GSE30971 CTRL VS LPS STIM MACROPHAGE WBP7 HET 4H DN | 180 | 154 | 2.25e-03 | 0.0704045 |
| M5748 GSE7852 LN VS FAT TCONV DN | 169 | 145 | 2.34e-03 | 0.0727796 |
| M41159 NAKAYA B CELL FLUARIX FLUVIRIN AGE 18 50YO 7DY UP | 338 | 281 | 2.40e-03 | 0.0731931 |
| M9953 GSE43260 BTLA POS VS NEG INTRATUMORAL CD8 TCELL UP | 190 | 162 | 2.41e-03 | 0.0731931 |
| M365 GSE7568 IL4 VS IL4 AND TGFB TREATED MACROPHAGE 24H UP | 158 | 136 | 2.41e-03 | 0.0731931 |
| M5295 GSE36476 CTRL VS TSST ACT 16H MEMORY CD4 TCELL OLD UP | 158 | 136 | 2.41e-03 | 0.0731931 |
| M4417 GSE22886 NAIVE VS MEMORY TCELL DN | 174 | 149 | 2.44e-03 | 0.0736709 |
| M9102 GSE40274 CTRL VS FOXP3 AND EOS TRANSDUCED ACTIVATED CD4 TCELL UP | 136 | 118 | 2.48e-03 | 0.0744433 |
| M4575 GSE24142 DN2 VS DN3 THYMOCYTE FETAL DN | 163 | 140 | 2.53e-03 | 0.0746422 |
| M7469 GSE21670 STAT3 KO VS WT CD4 TCELL IL6 TREATED DN | 163 | 140 | 2.53e-03 | 0.0746422 |
| M8087 GSE23502 BM VS COLON TUMOR MYELOID DERIVED SUPPRESSOR CELL UP | 163 | 140 | 2.53e-03 | 0.0746422 |
| M8686 GSE34156 NOD2 LIGAND VS TLR1 TLR2 LIGAND 6H TREATED MONOCYTE DN | 152 | 131 | 2.60e-03 | 0.0761470 |
| M9599 GSE42021 TCONV PLN VS CD24HI TCONV THYMUS DN | 184 | 157 | 2.62e-03 | 0.0761470 |
| M4007 GSE17721 LPS VS CPG 16H BMDC UP | 168 | 144 | 2.64e-03 | 0.0761470 |
| M5476 GSE3982 DC VS CENT MEMORY CD4 TCELL UP | 168 | 144 | 2.64e-03 | 0.0761470 |
| M5660 GSE6269 E COLI VS STAPH AUREUS INF PBMC DN | 157 | 135 | 2.73e-03 | 0.0777287 |
| M7962 GSE25123 IL4 VS IL4 AND ROSIGLITAZONE STIM MACROPHAGE DAY10 UP | 157 | 135 | 2.73e-03 | 0.0777287 |
| M9725 GSE43863 NAIVE VS TFH CD4 EFF TCELL D6 LCMV UP | 173 | 148 | 2.74e-03 | 0.0777287 |
| M2945 GSE14415 ACT TCONV VS ACT NATURAL TREG UP | 146 | 126 | 2.79e-03 | 0.0782868 |
| M7010 GSE15330 LYMPHOID MULTIPOTENT VS MEGAKARYOCYTE ERYTHROID PROGENITOR IKAROS KO DN | 146 | 126 | 2.79e-03 | 0.0782868 |
| M4659 GSE26495 NAIVE VS PDILOW CD8 TCELL DN | 162 | 139 | 2.85e-03 | 0.0791105 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|---|---------|------------------------|----------|-----------|
| M5463 GSE3982 DC VS BCELL UP | 162 | 139 | 2.85e-03 | 0.0791105 |
| M4753 GSE27786 LSK VS NKCELL DN | 183 | 156 | 2.93e-03 | 0.0803133 |
| M8869 GSE37301 COMMON LYMPHOID PROGENITOR VS CD4 TCELL DN | 183 | 156 | 2.93e-03 | 0.0803133 |
| M8989 GSE37534 GW1929 VS ROSIGLITAZONE TREATED CD4 TCELL PPARG1 FOXP3 TRANSDUCED DN | 151 | 130 | 2.94e-03 | 0.0803133 |
| M4408 GSE22886 UNSTIM VS STIM MEMORY TCELL UP | 167 | 143 | 2.97e-03 | 0.0807058 |
| M8898 GSE37301 CD4 TCELL VS GRANULOCYTE MONOCYTE PROGENITOR DN | 140 | 121 | 3.00e-03 | 0.0810942 |
| M4343 GSE20366 CD103 KLRG1 DP VS DN TREG DN | 172 | 147 | 3.08e-03 | 0.0811329 |
| M4387 GSE2197 IMMUNOSUPPRESSIVE DNA VS UNTREATED IN DC UP | 172 | 147 | 3.08e-03 | 0.0811329 |
| M6636 GSE5589 LPS AND IL10 VS LPS AND IL6 STIM IL10 KO MACROPHAGE 45MIN DN | 172 | 147 | 3.08e-03 | 0.0811329 |
| M6207 GSE2405 S AUREUS VS A PHAGOCYTOPHILUM NEUTROPHIL UP | 156 | 134 | 3.08e-03 | 0.0811329 |
| M9414 GSE39556 UNTREATED VS 3H POLYIC INJ MOUSE CD8A DC UP | 156 | 134 | 3.08e-03 | 0.0811329 |
| M5339 GSE37416 CTRL VS 48H F TULARENSIS LVS NEUTROPHIL UP | 145 | 125 | 3.16e-03 | 0.0824539 |
| M7967 GSE25123 ROSIGLITAZONE VS IL4 AND ROSIGLITAZONE STIM MACROPHAGE DAY10 UP | 177 | 151 | 3.19e-03 | 0.0824539 |
| M8914 GSE32255 UNSTIM VS 4H LPS STIM DC UP | 161 | 138 | 3.21e-03 | 0.0824539 |
| M4489 GSE22886 NAIVE CD8 TCELL VS NEUTROPHIL DN | 134 | 116 | 3.21e-03 | 0.0824539 |
| M7130 GSE9960 GRAM NEG VS GRAM NEG AND POS SEPSIS PBMC UP | 134 | 116 | 3.21e-03 | 0.0824539 |
| M7423 GSE16450 CTRL VS IFNA 6H STIM MATURE NEURON CELL LINE DN | 182 | 155 | 3.28e-03 | 0.0834219 |
| M9903 GSE42088 UNINF VS LEISHMANIA INF DC 4H UP | 182 | 155 | 3.28e-03 | 0.0834219 |
| M6877 GSE4590 LARGE PRE BCELL VS VPRES POS LARGE PRE BCELL UP | 150 | 129 | 3.32e-03 | 0.0836154 |
| M9373 GSE11961 FOLLICULAR BCELL VS MARGINAL ZONE BCELL DN | 150 | 129 | 3.32e-03 | 0.0836154 |
| M3136 GSE11864 CSF1 VS CSF1 PAM3CYS IN MAC UP | 187 | 159 | 3.38e-03 | 0.0842641 |
| M9469 GSE41867 NAIVE VS DAY8 LCMV EFFECTOR CD8 TCELL DN | 187 | 159 | 3.38e-03 | 0.0842641 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | n ^o Genes | n ^o Genes Significativos | P-valor | FDR |
|--|----------------------|-------------------------------------|----------|-----------|
| M5644 GSE6269 FLU VS STAPH AUREUS INF PBMC DN | 139 | 120 | 3.40e-03 | 0.0842641 |
| M7686 GSE20152 SPHK1 KO VS WT HTNFA OVERXPRESS ANKLE DN | 192 | 163 | 3.46e-03 | 0.0842641 |
| M5494 GSE3982 MAC VS BCELL UP | 171 | 146 | 3.46e-03 | 0.0842641 |
| M5602 GSE39820 CTRL VS IL1B IL6 IL23A CD4 TCELL UP | 171 | 146 | 3.46e-03 | 0.0842641 |
| M5720 GSE7764 NKCELL VS SPLENOCYTE UP | 171 | 146 | 3.46e-03 | 0.0842641 |
| M5541 GSE3982 BCELL VS EFF MEMORY CD4 TCELL UP | 176 | 150 | 3.57e-03 | 0.0849696 |
| M9161 GSE40274 XBP1 VS FOXP3 AND XBP1 TRANSDUCED ACTIVATED CD4 TCELL UP | 176 | 150 | 3.57e-03 | 0.0849696 |
| M6502 GSE4984 GALECTIN1 VS LPS STIM DC UP | 144 | 124 | 3.58e-03 | 0.0849696 |
| M7511 GSE21379 WT VS SAP KO CD4 TCELL DN | 160 | 137 | 3.62e-03 | 0.0849696 |
| M8187 GSE25677 MPL VS R848 STIM BCELL DN | 160 | 137 | 3.62e-03 | 0.0849696 |
| M8212 GSE27670 CTRL VS BLIMP1 TRANSDUCED GC BCELL UP | 133 | 115 | 3.65e-03 | 0.0849696 |
| M5710 GSE7460 WT VS FOXP3 HET ACT TCONV DN | 181 | 154 | 3.67e-03 | 0.0849696 |
| M5750 GSE7852 THYMUS VS FAT TCONV DN | 181 | 154 | 3.67e-03 | 0.0849696 |
| M7251 GSE18281 SUBCAPSULAR CORTICAL REGION VS WHOLE CORTEX THYMUS UP | 181 | 154 | 3.67e-03 | 0.0849696 |
| M7436 GSE21670 TGFB VS TGFB AND IL6 TREATED CD4 TCELL DN | 181 | 154 | 3.67e-03 | 0.0849696 |
| M8997 GSE37533 PPARG1 FOXP3 VS PPARG2 FOXP3 TRANSDUCED CD4 TCELL DN | 181 | 154 | 3.67e-03 | 0.0849696 |
| M6260 GSE8835 HEALTHY VS CLL CD4 TCELL UP | 165 | 141 | 3.75e-03 | 0.0864342 |
| M8042 GSE17301 IFNA2 VS IFNA2 AND ACD3 ACD28 STIM CD8 TCELL UP | 191 | 162 | 3.86e-03 | 0.0881822 |
| M8539 GSE33425 CD161 INT VS NEG CD8 TCELL UP | 191 | 162 | 3.86e-03 | 0.0881822 |
| M7492 GSE21774 CD56 BRIGHT VS DIM CD62L POSITIVE NK CELL DN | 170 | 145 | 3.88e-03 | 0.0882486 |
| M8246 GSE27896 HDAC6 KO VS WT TREG DN | 143 | 123 | 4.05e-03 | 0.0909671 |
| M309 GSE10240 IL22 VS IL22 AND IL17 STIM PRIMARY BRONCHIAL EPITHELIAL CELLS DN | 159 | 136 | 4.07e-03 | 0.0909671 |
| M6899 GSE9509 LPS VS LPS AND IL10 STIM IL10 KO MACROPHAGE 20MIN UP | 159 | 136 | 4.07e-03 | 0.0909671 |
| M7506 GSE21380 TFH VS GERMINAL CENTER TFH CD4 TCELL UP | 159 | 136 | 4.07e-03 | 0.0909671 |
| M8851 GSE37301 MULTIPOTENT PROGENITOR VS RAG2 KO NK CELL DN | 180 | 153 | 4.11e-03 | 0.0914652 |

Tabla B.9: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M4869 GSE27786 NKTCELL VS NEUTROPHIL DN | 164 | 140 | 4.21e-03 | 0.0932885 |
| M6595 GSE5099 CLASSICAL M1 VS ALTERNATIVE M2 MACROPHAGE UP | 169 | 144 | 4.35e-03 | 0.0955704 |
| M9269 GSE39916 B CELL SPLEEN VS PLASMA CELL BONE MARROW UP | 169 | 144 | 4.35e-03 | 0.0955704 |
| M6347 GSE3720 LPS VS PMA STIM VD2 GAMMADelta TCELL DN | 104 | 91 | 4.37e-03 | 0.0956030 |
| M6091 GSE2770 UNTREATED VS IL12 TREATED ACT CD4 TCELL 2H DN | 195 | 165 | 4.39e-03 | 0.0956353 |
| M5585 GSE3982 CENT MEMORY CD4 TCELL VS NKCELL DN | 153 | 131 | 4.41e-03 | 0.0956674 |
| M6111 GSE2770 TGFB AND IL4 VS TGFB AND IL12 TREATED ACT CD4 TCELL 48H DN | 174 | 148 | 4.47e-03 | 0.0961609 |
| M9168 GSE40274 HELIOS VS FOXP3 AND HELIOS TRANSDUCED ACTIVATED CD4 TCELL DN | 174 | 148 | 4.47e-03 | 0.0961609 |
| M4952 GSE29618 PDC VS MDC DN | 158 | 135 | 4.57e-03 | 0.0971236 |
| M9370 GSE11961 MEMORY BCELL DAY7 VS GERMINAL CENTER BCELL DAY7 DN | 158 | 135 | 4.57e-03 | 0.0971236 |
| M8794 GSE37605 TREG VS TCONV C57BL6 FOXP3 FUSION GFP DN | 142 | 122 | 4.58e-03 | 0.0971236 |
| M9130 GSE40273 EOS KO VS WT TREG DN | 179 | 152 | 4.59e-03 | 0.0971236 |
| M8632 GSE32986 CURDLAN LOWDOSE VS CURDLAN HIGHDOSE STIM DC UP | 184 | 156 | 4.69e-03 | 0.0978689 |
| M8651 GSE26343 WT VS NFAT5 KO MACROPHAGE LPS STIM DN | 184 | 156 | 4.69e-03 | 0.0978689 |
| M2976 GSE14415 NATURAL TREG VS TCONV UP | 131 | 113 | 4.71e-03 | 0.0978689 |
| M6578 GSE5455 HEALTHY VS TUMOR BEARING MOUSE SPLEEN MONOCYTE 24H INCUBATION DN | 163 | 139 | 4.72e-03 | 0.0978689 |
| M8457 GSE27291 0H VS 7D STIM GAMMADelta TCELL DN | 163 | 139 | 4.72e-03 | 0.0978689 |
| M4512 GSE22886 TH1 VS TH2 48H ACT DN | 189 | 160 | 4.79e-03 | 0.0985290 |
| M5432 GSE3982 MAST CELL VS NEUTROPHIL UP | 189 | 160 | 4.79e-03 | 0.0985290 |
| M4366 GSE20715 0H VS 24H OZONE LUNG UP | 168 | 143 | 4.87e-03 | 0.0993826 |
| M6892 GSE9316 IL6 KO VS IFNG KO INVIVO EXPANDED CD4 TCELL DN | 168 | 143 | 4.87e-03 | 0.0993826 |

Tabla B.10: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MOG (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|-------------------------|---------|---------|----------|-----------|
| M2680 BCAT.100 UP.V1 UP | 41 | -1 | 0.000529 | 0.0722925 |
| M2808 CAHOY NEURONAL | 54 | -1 | 0.000765 | 0.0722925 |

B.2. Caso práctico: análisis de microarrays de las muestras pareadas de tejido humano de cáncer de colon

En este apartado se va a presentar muy brevemente un caso práctico, a modo de ejemplo, de cómo analizar los datos transcriptómicos de microarrays de Affymetrix (aunque tiene bastantes partes en común con el análisis de los datos transcriptómicos de *RNA-seq* a partir de la matriz de expresión, a excepción de lo expuesto en el apartado “B.1”) y de la interpretación de los resultados. Todos los experimentos basados en microarrays de la presente tesis han sido obtenidos a partir de *Affymetrix GeneChips*, tanto en los experimentos con muestras procesadas por nuestro grupo de investigación -líneas celulares y muestras de tejido- como con los experimentos extraídos de repositorios públicos (NCBI GEO), por lo que únicamente se va a presentar un ejemplo de análisis de este tipo de microarrays.

En primer lugar, tras el escaneo del chip, el *software* del equipo genera diferentes archivos: *.EXP*, que contiene información sobre el experimento; *.DAT*, que es la imagen original digitalizada; *.CEL*, generado a través del escaneo y segmentación de la imagen, el cual contiene las cuantificaciones realizadas por el *software* del equipo de las señales de fluorescencia de cada celda (fluorescencia de cada sonda), y suelen denominarse *features*; *.CDF*, el cual proporciona la correspondencia entre las sondas, las *features*, los conjuntos de sondas y sus genes diana; *.CHP*, proporciona el cálculo de los niveles de expresión de cada gen realizado por el *software* del equipo.

Tras obtener los archivos que proporciona el equipo, el análisis comienza a partir de los archivos de extensión *.CEL*, uno por muestra, para lo que se emplea el *software R* que, además, es el que se ha usado a lo largo de todos los análisis de la presente tesis. Tras cargar y leer estos archivos, se someten a un control de calidad y se preprocesan antes de realizar cualquier análisis. A continuación, se comprueba la calidad de cada uno de los microarrays, para lo que se emplean desde métodos generales no específicos para microarrays (ACP, detección de *outliers* multivariante...) hasta métodos diseñados específicamente para microarrays de Affymetrix (modelos a nivel de sonda y el análisis de RLE y NUSE, gráficos de degradación...). Posteriormente, se realizaron los análisis de expresión diferencial y de significación biológica a partir de la lista de genes diferencialmente expresados en cada uno de los contrastes de hipótesis realizados.

Como librerías esenciales, se han empleado *affy* (Gautier et al. (2004)), *affyPLM* (B. M. Bolstad (2004), B. M. Bolstad et al. (2005), Brettschneider et al. (2007)), *affycoretools* (MacDonald (2020)), *oligo* (B. S. Carvalho and Irizarry (2010)), *oligoClasses* (B. Carvalho and

Scharpf (2020)), AnnotationDbi (Pagès et al. (2020)), rrcov (Todorov and Filzmoser (2009)), genefilter (Gentleman et al. (2021)), limma (Ritchie et al. (2015)), C50 (Kuhn and Quinlan (2021)), EnrichmentBrowser (Geistlinger, Csaba, and Zimmer (2016)).

B.2.0.1. Análisis exploratorio y control de calidad

Para evaluar la calidad de las muestras y detectar posibles *outliers*, se han empleado diferentes métodos. En primer lugar, se realizó un análisis exploratorio de los datos para visualizar la distribución de la señal original (intensidades de las sondas sin normalizar) de las diferentes muestras mediante gráficos de cajas y bigotes (figuras B.3a) y la estimación de la densidad Kernel (figuras B.3b) de cada una de las muestras. Así, se esperaría ver una distribución de las intensidades similar a lo largo de todas las muestras. No obstante, que no se aprecie tal similitud no tiene por qué ser indicativo de que alguna de ellas esté en mal estado, pues hay que tener en cuenta que, en este punto, los datos todavía no han sido normalizados y, por ende, no tienen por qué ser comparables entre ellos en este punto del análisis. Lo que sí es cierto es que se puede observar algún patrón atribuible a algún efecto indeseado, lo que se conoce como *batch effect* y que, de ser así, deberá de ser corregido más adelante por ejemplo, usando este efecto como un factor de bloqueo en el modelo usando en el futuro análisis de expresión diferencial. Un ejemplo de *batch effect* podría ser que se observara diferencias entre las muestras 1 a 5 y 6 a 10, de tal manera que las primeras se hubieran hecho en un día diferente a las de l grupo 6 a 10.

En este caso, se observan distribuciones diferentes en cuanto a forma, pero no en posición, lo que puede indicar la necesidad de una normalización de los datos para hacerlas comparables entre sí en posteriores análisis (figuras B.3a, b). Cabe destacar que no se observa ningún *batch effect*; es decir, que no se observa ningún factor conocido ni desconocido que esté ejerciendo un efecto en la forma de la distribución de la señal original.

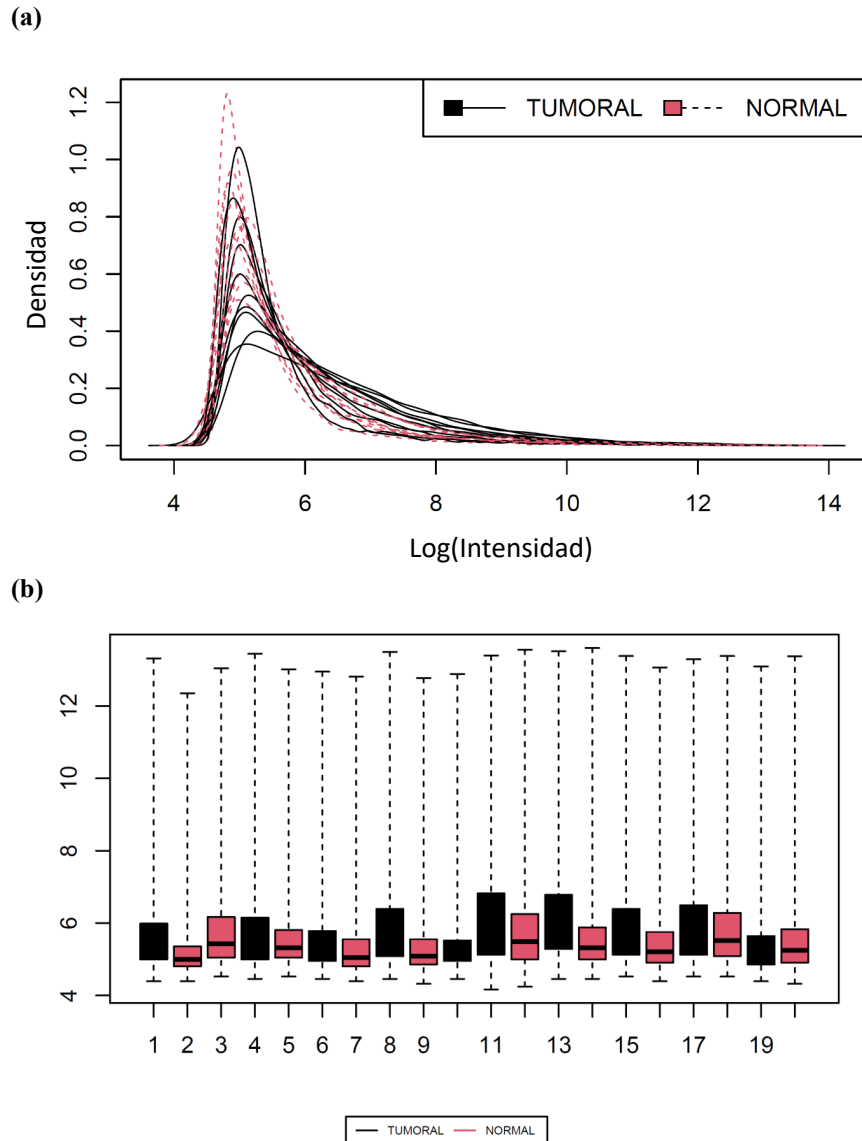


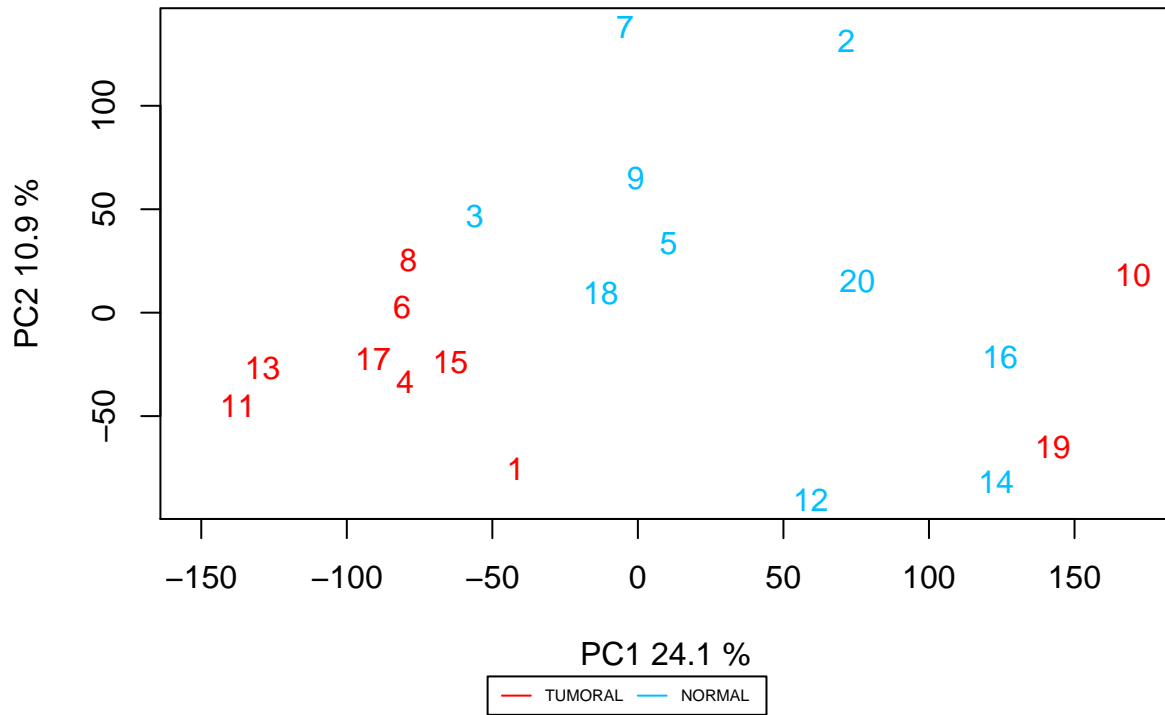
Figura B.3: Análisis exploratorio de los datos *outliers*. a) Gráfico de densidad

En segundo lugar, se emplean los datos normalizados (normalización por cuantiles) y transformados mediante la función logarítmica en base 2 para hacer los diferentes microarrays comparables entre sí. Además, estos no serán centrados ni estandarizados con el fin de tratar de forma diferente a las muestras atípicas del resto y a los genes poco informativos (ruido) de los genes importantes. Así, se realiza una búsqueda de patrones -agrupaciones- y detección de *outliers* mediante dos métodos de aprendizaje no supervisado: ACP combinado con la detección multivariante de *outliers* mediante la distancia robusta de Mahalanobis, y análisis de clúster jerárquico. El primer método combina la reducción de la dimensión de la matriz de expresión mediante ACP y la selección de un número reducido de componentes principales, sin tener en cuenta en cuenta las clases en las que se pueden agrupar las diferentes muestras, con el cálculo de las distancias de Mahalanobis robustas considerando cada una de las clases de forma independiente

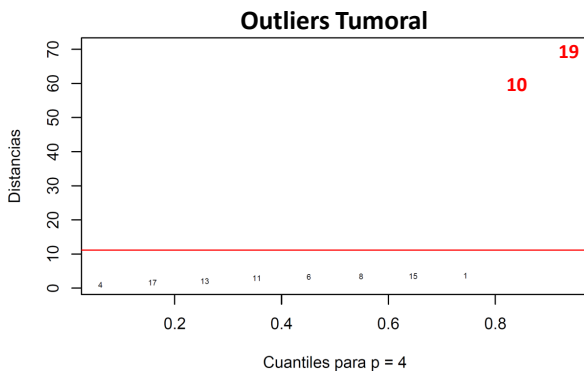
y usando como variables las componentes principales seleccionadas (figuras B.4b-c). Para más información, consultar el apéndice “A.4 Detección multivariante de outliers” y el texto Shieh and Hung (2009).

Cabe destacar que la identificación de *outliers* mediante este método combinado va a depender de los factores considerados como significativos o explicativos en los modelos lineales empleados en el análisis de expresión diferencial, puesto que serán los considerados para definir las diferentes clases en las que se agruparán las muestras. Como resultado, se han identificado como posibles *outliers* a las muestras 10, 12 y 19 (figuras B.4b-c); además, se observa una separación clara entre las muestras *Normales* y *Tumorales* (figura B.4a).

(a)



(b)



(c)

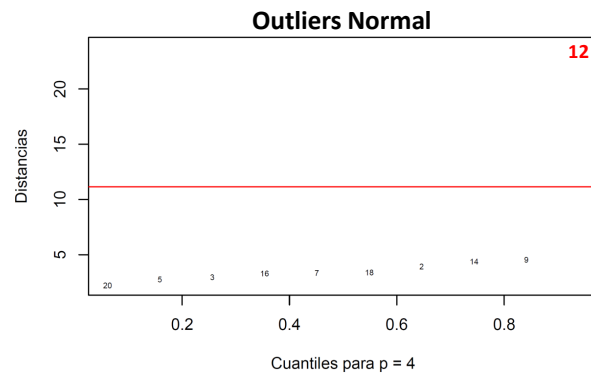


Figura B.4: Análisis de Componentes Principales y detección de *outliers* mediante la distancia de Mahalanobis. (a) Análisis de componentes principales sobre todas las muestras sin tener en cuenta las clases, coloreada en función del factor de interés *Tumor*: tumoral en negro y normal en rojo. (b-c) Detección de *outliers* mediante las distancias de Mahalanobis: En los gráficos se muestran los gráficos qq para las distancias de Mahalanobis calculadas para cada una de las clases definidas por el factor *Tumor* -tumoral y normal-, que siguen una distribución $\chi^2_{p=4}$, donde p es el número de componentes principales seleccionadas. La línea horizontal roja indica el valor crítico $\chi^2_{p=4}$ a partir del cual se considera que una observación cuya distancia de Mahalanobis es mayor será identificada como *outlier*. (ver apartado *Material y métodos*).

El otro método de aprendizaje no supervisado empleado capaz de encontrar agrupaciones entre las muestras y de detectar *outliers* es el *Análisis de Clúster Jerárquico*, donde, mediante el método de Ward, se agrupan las muestras, dos a dos, en función de su mayor o menor parecido. Como medida de disimilaridad se ha usado la distancia euclídea entre las muestras y, como variables, las 9 primeras componentes principales por ser las necesarias para explicar, al menor, un 70 % de la variabilidad de los datos -el uso de todas las variables originales puede introducir ruido y perjudicar la calidad de estos análisis exploratorios-. Así, por un lado, se identifican los microarrays 2 y 10 como posibles *outliers* y, además, se observa una clara agrupación de las muestras en función de la variable *Tumor* (Figura B.5).

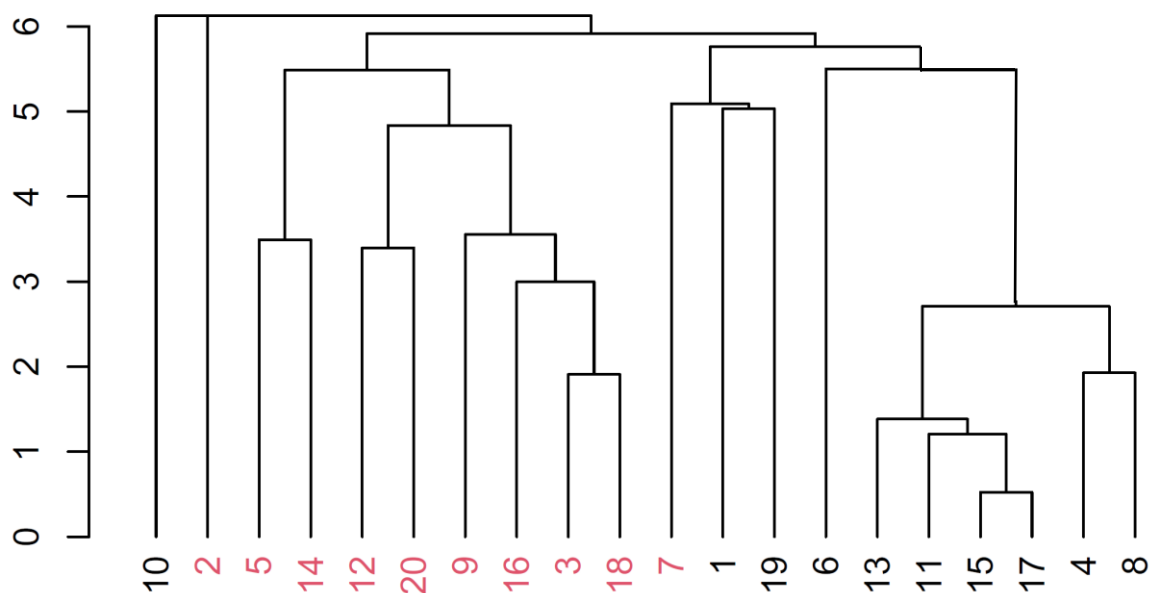


Figura B.5: Control de calidad y detección de *outliers* con métodos generales. Análisis Clúster Jerárquico. Se ha usado el método de Ward y como medida de disimilaridad se ha usado la distancia euclídea. Las muestras se han coloreado en función del factor *Tumor* (negro para muestras tumorales y rojo para normales).

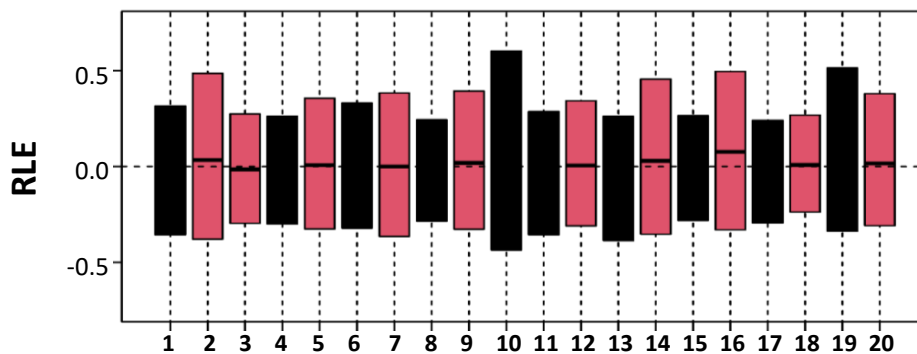
Finalmente, se emplean los modelos de bajo nivel o modelos a nivel de sonda. Estos modelos ajustan a los valores de intensidad de fluorescencia de cada una de las sondas un modelo explicativo y se obtienen los residuos del ajuste. Así, a través del análisis de estos residuos es posible evaluar qué muestra podría ser problemática, cuyo fundamento es el mismo que para los modelos lineales explicados en el apartado “A.6.1”. Así, de este modelo ajustado se obtienen dos medidas: expresión logarítmica relativa (RLE), de poca utilidad, y el error estándar normalizado sin escalar (NUSE).

RLE: una medida relativa o estandarizada de la expresión en escala logarítmica, de tal manera que su debe mostrar una expresión similar para todos los microarrays, tanto a nivel mediano como en términos de dispersión. En una representación con gráficos de cajas y bigotes, una muestra problemática debería mostrar una distribución similar a las demás.

NUSE: en la representación del error estándar normalizado no escalamado pueden detectarse microarrays problemáticos en el caso de que alguno no presente un patrón similar al resto que, en una representación de cajas y bigotes se traduce en que las cajas aparezcan desplazadas hacia arriba o hacia abajo respecto de las demás. En caso de que alguna muestra presente diferencias respecto de las demás, esta será considerada como potencial *outlier*

De los dos anteriores, el de mayor utilidad es *NUSE*. En este caso, mientras que en la representación gráfica de los valores de *RLE* no se observa ningún microarray que sea claramente diferente a los demás (figura B.6a), sí se detectan dos posibles muestras atípicas mediante la representación de los valores *NUSE*: los microarrays 2 y 10 (figura B.6b).

(a)



(b)

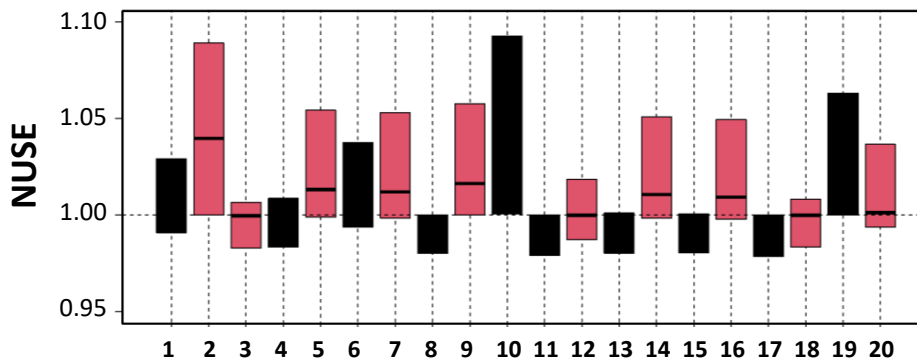


Figura B.6: Control de calidad y detección de *outliers* con métodos específicos para microarrays. a) Expresión logarítmica relativa (*RLE*): b) Error estándar normalizado sin escalar (*NUSE*). Distribuciones muy diferentes a la mayoría indicarían algún problema en la muestra.

En algunos microarrays de Affymetrix, el software empleado también permite realizar un análisis de control conocido como *análisis de degradación*, el cual se basa en la degradación del ARN. El ARN puede degradarse por diversos motivos al final de su vida útil, como por ejemplo la degradación por enzimas celulares tras ejercer su función tras la síntesis de la proteína correspondiente. Por ello, es importante asegurarse que los diferentes microarrays que van a compararse entre sí tengan un ARN de calidad similar para evitar sesgos en las inferencias

obtenidas en posteriores análisis. Debido a los mecanismos de degradación del ARN, la intensidad de las sondas es sistemáticamente menor según se aproxima al extremo 5', incluso en experimentos con una excelente calidad del RNA, por lo que a mayor pendiente mayor es la diferencia de degradación entre los extremos y peor es la calidad.

Cabe destacar que la tendencia creciente de intensidad depende del conjunto de sondas y, por ende, la distribución de estas pendientes dentro de un microarray depende del propio chip, pues para diferentes chips se han encontrado diferentes pendientes debido a las diferentes arquitecturas de los conjuntos de sondas, según estén más o menos cercanas en la secuencia de bases del gen que hibridan, incluso si en ocasiones se solapan en la propia secuencia, siendo menos pronunciada la pendiente $5' \rightarrow 3'$, o si por el contrario estuvieran más alejadas. Por lo tanto, interesa que las curvas estimadas para los diferentes chips sean lo más paralelas entre sí, pues eso significaría que el ARN de los diferentes microarrays tiene una degradación similar a lo largo de los conjuntos de sondas. A pesar de que podría tomarse como intervalo para considerar aceptables las pendientes de degradación $\leq 2 \cdot \text{pendiente}$, lo importante es lo similares que son entre sí los valores de pendiente de los diferentes microarrays de un mismo experimento, por lo que si son similares, a pesar de que el valor de pendiente esté por encima del anterior valor, las comparaciones entre los diferentes microarrays son válidas. Cabe destacar que en estos análisis se representa el logaritmo de la intensidad media a lo largo de los N conjuntos de sondas (y_j) correspondientes a la posición j de la sonda i en el conjunto de sondas, pero se estandarizan los valores de cada microarray para que tengan un error estándar 1, para lo que se dividen los valores (y_j) entre el error estándar de las intensidades en la posición j, $\hat{\sigma}_j$.

Como conclusión del control de calidad realizado a partir de los diferentes anteriores, se eliminan de los análisis posteriores los microarrays $\{2, 10, 12, 19\}$.

B.2.0.2. Preprocesado de los datos

El análisis exploratorio anterior evidenció cómo la distribución de los diferentes microarrays es diferente (figura B.3), por lo que, previo a cualquier análisis -búsqueda de patrones, análisis de expresión diferencial, análisis de conjuntos de genes...-, es necesario normalizar las diferentes muestras para hacerlos comparables entre sí. Además, en el caso de los microarrays y, en concreto, de los microarrays de Affymetrix, también es necesario corregir el ruido de fondo y resumir las intensidades de las diferentes sondas pertenecientes a un mismo *probeset* o conjunto de sondas (correspondientes a un mismo gen) a un solo valor de expresión. Para ello, se lleva a cabo un proceso que consta de tres etapas: i) corrección del ruido de fondo, ii) normalización, y iii) resumen de los diferentes valores de cada conjunto de sondas en un solo valor de expresión. En este caso, y aunque existen más de una treintena de métodos publicados para llevar a cabo estas tres etapas de forma conjunta, se ha optado por el método *robust multi-array average (RMA)* el cual se basa en la modelización de las intensidades de las mismas sondas a lo largo de todos los microarrays (Irizarry et al. (2003), B. Bolstad (2008)). Este método corrige el fondo teniendo en cuenta sólo las sondas PM y una combinación de las distribuciones exponencial para la señal y normal para el ruido, realiza

una transformación logarítmica en base dos, lleva a cabo una normalización por cuantiles usando información de todos los microarrays, y estima la intensidad de cada gen empleando el método conocido como *median polish*. Cabe destacar que método se considera el método estándar (para la mayoría de los usuarios de la plataforma Bioconductor) en el análisis de microarrays de un color. Como puede observarse en la figura B.7, se ha conseguido que unos datos que claramente no son comparables (figura B.7a) sí lo sean tras aplicar este método (figura B.7b).

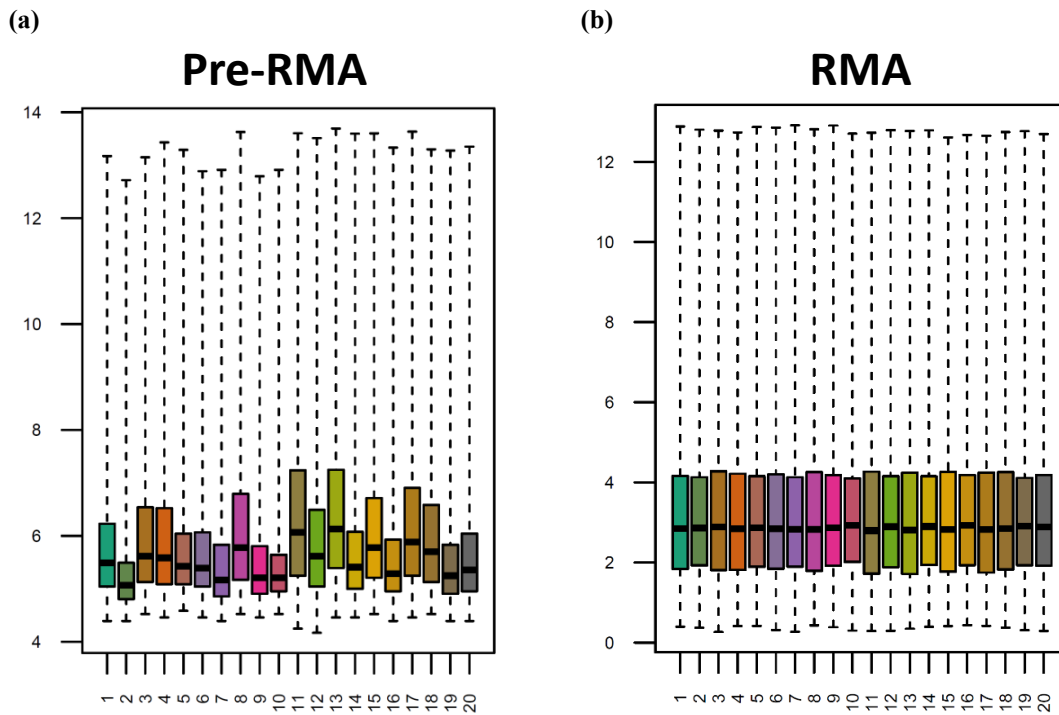


Figura B.7: Corrección del ruido de fondo, normalización y resumen mediante el método RMA. a) Expresión logarítmica previa al pre-tratamiento; b) Expresión logarítmica posterior al pre-tratamiento.

B.2.0.3. Filtrado y análisis de expresión diferencial

Posteriormente, debido a la gran dimensión de este tiempo de datos, es necesario filtrarlos para evitar que genes no relacionados con los factores estudiados introduzcan ruido y que, además, los análisis de expresión diferencial resulten muy conservativos tras corregir los p-valores al realizar comparaciones múltiples (miles de contrastes de hipótesis, uno por cada gen). Por ello, se ha llevado a cabo un filtrado no específico -sin tener en cuenta la covariable/s de estudio- de aquellos genes que, aparentemente, no muestren tener actividad alguna en las muestras estudiadas; es decir, que no se expresen, que no varíe su expresión entre los diferentes microarrays o porque no exista una equivalencia del conjunto de sondas con la anotación a usar (ALIAS, UNIGENE...).

Para ello, se ha empleado la función `genefilter::nsFilter(eset, require.entrez=FALSE, remove.dupEntrez=FALSE, var.func=IQR, var.cutoff=0.5, var.filter=TRUE, filterByQuantile=TRUE, feature.exclude="~AFFX")`. Además, dado que el interés de la presente tesis gira en torno a una serie de genes relacionados con el transporte de Ca^{2+} intracelular, dichos genes también se mantendrán tras el filtrado. Así, de un conjunto inicial de datos (ya resumidos y normalizados) de 138745 sondas, este se reduce 69439 sondas diferentes y, además, si se selecciona también sólo aquellas sondas que tienen alguna equivalencia en las anotaciones de interés (PROBE:Anotación 1:n para $n > 0$), se reduce aun más: 12835 para UNIGENE y 12944 para ALIAS.

Para llevar a cabo el análisis de expresión diferencial entre las muestras de tejido humano de cáncer de colon y normal se emplea el método limma (Ritchie et al. (2015)), similar a los explicados en el apéndice “A.6.1”, pues se basa en ajustar un modelo lineal a cada una de las sondas de los microarrays con el fin modelizar la expresión de cada una de ellas en función de una serie de variables explicativas. Además, mediante métodos bayesianos en los que se emplean para las estimaciones de cada gen información de todos los genes a la vez, se mejoraron las estimaciones y con ello la fiabilidad de las inferencias (Smyth (2004)). Al disponer de 3 potenciales variables explicativas, *Tejido* (fija), *Sexo* (fija) y *Paciente* (aleatoria), y, dado que al ser la variable *Tejido* es la única de interés, esta ha de estar en todos los modelos evaluados, se han planteado hasta un total de 6 posibles modelos diferentes: 3 fijos (sin efectos aleatorios, *paciente*), y otros 3 mixtos (con efectos aleatorios), mostrados a continuación:

- *Modelo completo o sobreóptimo Mixto*, en el que se consideran todas las posibles variables explicativas, sus interacciones y los efectos aleatorios:

$$\log(\text{expresin})_{ijk} = \text{Tejido}_i \alpha_i + \text{sexo}_j \beta_j + \text{Tejido}_i \text{sexo}_j \gamma_{ij} + A_k + \epsilon_{ijk} \quad (\text{B.1})$$

donde los parámetros del modelo son: α_i el efecto del nivel $i \in \{\text{NORMAL}, \text{TUMORAL}\}$ del factor *Tejido*; β_j el efecto del nivel $j \in \{\text{HOMBRE}, \text{MUJER}\}$ del factor *sexo*; γ_{ij} el efecto de la interacción entre el nivel i del factor *Tejido* y el nivel j del factor *sexo*; A_k el efecto aleatorio del paciente k , y ϵ_{ijk} es el error aleatorio correspondiente a la observación para la muestra del paciente k cuyo sexo es j y el nivel del factor *Tejido* es i .

- *Modelo completo o sobreóptimo Fijo*, en el que se consideran todas las posibles variables explicativas, sus interacciones y sin efectos aleatorios:

$$\log(\text{expresin})_{ij} = \text{Tejido}_i\alpha_i + \text{sexo}_j\beta_j + \text{Tejido}_i\text{sexo}_j\gamma_{ij} + \epsilon_{ij} \quad (\text{B.2})$$

- *Modelo TS Mixto*, en el que se consideran todas las posibles variables explicativas, sin interacciones, y los efectos aleatorios:

$$\log(\text{expresin})_{ijk} = \text{Tejido}_i\alpha_i + \text{sexo}_j\beta_j + A_k + \epsilon_{ijk} \quad (\text{B.3})$$

- *Modelo TS Fijo*, en el que se consideran todas las posibles variables explicativas, sin interacciones, y sin efectos aleatorios:

$$\log(\text{expresin})_{ij} = \text{Tejido}_i\alpha_i + \text{sexo}_j\beta_j + \epsilon_{ij} \quad (\text{B.4})$$

- *Modelo T Mixto*, en el que se considera la variable explicativa *Tejido* y efectos aleatorios:

$$\log(\text{expresin})_{ik} = \text{Tejido}_i\alpha_i + A_k + \epsilon_{ik} \quad (\text{B.5})$$

- *Modelo T Fijo*, en el que se considera la variable explicativa *Tejido* y efectos aleatorios:

$$\log(\text{expresin})_i = \text{Tejido}_i\alpha_i + \epsilon_i \quad (\text{B.6})$$

Para seleccionar el modelo más adecuado, que será el más simple -menos parámetros- que mejor explique los datos se han empleado dos criterios que permiten comparar entre sí diferentes modelos, basados en la verosimilitud en la penalización de aquellos modelos sobreparametrizados: la *Información de Akaike (AIC)* y la *Información Bayesiana (BIC)* (función `limma::selectModel()`). No obstante, puesto que cada uno de los modelos evaluados se ajusta a cada una de las sondas del microarray, se tiene un valor de AIC y BIC por sonda y modelo evaluado. Así, el modelo seleccionado fue aquel que mejor explica a la mayoría de las sondas (menor valor de AIC y BIC). Además, para tener en cuenta las estructuras de correlación entre las medidas repetidas realizadas sobre el mismo individuo (muestras pareadas), sólo se han considerado aquellos modelos que incluyen al factor aleatorio *paciente* (modelos mixtos).

Además, también se evaluó, en el *modelo completo mixto*, el efecto de cada una de las potenciales variables explicatorias y sus interacciones sobre la expresión de las diferentes sondas para cerciorarse de que aquellas variables susceptibles de ser eliminadas realmente tienen un efecto testimonial sobre la mayoría de las sondas, pues, en caso de para alguna de dichas variables no hubiera sido así, se habría contemplado la posibilidad de mantenerla en el modelo final. Por ello, en primer lugar, se evaluó el efecto de cada una de las variables sobre las diferentes sondas a través del *modelo completo mixto* (ecuación B.1) el cual, para una manipulación más sencilla, puede ser reparametrizado sin que esto implique diferencias a la hora de realizar las estimaciones y

los contrastes de hipótesis de interés. En particular, a partir de las variables explicativas es posible crear un único factor fijo que es la combinación de las variables explicativas *Tejido* y *sexo*, cuyos niveles son *combinación* = {NORMALHOMBRE, TUMORALHOMBRE, NORMALMUJER, TUMORALMUJER} (ecuación B.7):

$$\log(\text{expresin})_{ik} = \text{combinacin}_i \alpha_i + A_k + \epsilon_{ik} \quad (\text{B.7})$$

donde α_i es el efecto del nivel $i \in \{ \text{NORMALHOMBRE, TUMORALHOMBRE, NORMALMUJER, TUMORALMUJER} \}$ del factor *combinado*, A_k el efecto aleatorio del paciente k y ϵ_{ik} el error aleatorio correspondiente a la observación para la muestra del paciente k cuya muestra se corresponde con el nivel i del factor combinación. Por lo tanto, sobre el *modelo completo mixto reparametrizado* se realiza una serie de preguntas con el fin de evaluar el efecto de las posibles variables sobre el total de las sondas: ¿Cuál es el efecto del fenotipo tumoral en la expresión de cada uno de los genes? ¿Y del sexo? ¿Son aditivos los efectos *Tejido* y *Sexo* (no existe interacción entre ambos)? ¿Qué genes están diferencialmente expresados entre el fenotipo tumoral y el normal? ¿Es diferente el patrón de expresión diferencial entre el fenotipo tumoral y sano para los hombres y para las mujeres?... Así, a partir de la representación matricial del modelo anterior (ecuación B.8):

$$\underbrace{\begin{pmatrix} \log(\text{expresin})_1 \\ \vdots \\ \log(\text{expresin})_k \\ \vdots \\ \log(\text{expresin})_{20} \end{pmatrix}}_{\text{Matriz expresin: } 18 \times 1} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{Matriz diseo: } 18 \times 4} \underbrace{\begin{pmatrix} \alpha_{\text{NORMALHOMBRE}} \\ \alpha_{\text{TUMORALHOMBRE}} \\ \alpha_{\text{NORMALMUJER}} \\ \alpha_{\text{TUMORALMUJER}} \end{pmatrix}}_{\text{Efectos: } 4 \times 1} + \underbrace{\begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_k \\ \vdots \\ \epsilon_{20} \end{pmatrix}}_{\text{Matriz errores } 18 \times 1} \quad (\text{B.8})$$

o, de su forma matricial simplificada (ecuación B.9):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{A} + \boldsymbol{\epsilon} \quad (\text{B.9})$$

se ha creado una matriz de contrastes de hipótesis (ecuación B.10) capaz de dar respuesta a las siguientes preguntas:

- β_1 : ¿Existe expresión diferencial entre el fenotipo tumoral y el normal? o ¿Cuál es el efecto del fenotipo tumoral en la expresión del gen g respecto del fenotipo sano?
- β_2 : ¿Existe expresión diferencial entre el sexos?
- β_3 : ¿Existe interacción entre los factores *Tejido* y *Sexo*?
- β_4 : ¿Existe expresión diferencial en hombres entre el fenotipo tumoral y el normal?
- β_5 : ¿Existe expresión diferencial en mujeres entre el fenotipo tumoral y el normal?

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} -0.5 & 0.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 & -0.5 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{\text{NORMALHOMBRE}} \\ \alpha_{\text{TUMORALHOMBRE}} \\ \alpha_{\text{NORMALMUJER}} \\ \alpha_{\text{TUMORALMUJER}} \end{pmatrix} \quad (\text{B.10})$$

Para realizar los contrastes de hipótesis hay que tener en cuenta el problema de las comparaciones múltiples y, en vez controlar el error de tipo I con un nivel de significación 0.05; es decir, que un gen sea significativo para un p-valor inferior al nivel de significación, se controla la *tasa de hipótesis falsamente rechazadas (FDR)* mediante el método propuesto por Benjamini and Hochberg (1995). Además, debido a la gran variabilidad que presentan las muestras, el nivel de significación escogido ha sido 0.1¹, en lugar del 0.05. Como resultado, de las más de 69000 sondas que han quedado tras el filtrado, se ha identificado un 30.84 % (3996) de ellas como diferencialmente expresadas entre el tejido humano de cáncer de colon y el normal (factor *Tejido*), mientras que sólo lo son un 0.02 % (3) entre hombre y mujer (factor *sexo*) y un 0 % (0) para la interacción entre ambos factores, lo que evidencia la necesidad de eliminar tanto la interacción como el factor *sexo* del modelo. Como curiosidad, los genes que muestran expresión diferencial entre sexos están localizados en los cromosomas sexuales: DDX3Y y RPS4Y1 en el cromosoma Y, y XIST en el X. Además, ninguno de ellos muestra expresión diferencial entre las muestras de tejido de cáncer de colon y normal.

Para llevar a cabo la selección del modelo, como ya se ha comentado anteriormente, se va a considerar sólo aquellos modelos en los que se tenga en cuenta el efecto del factor aleatorio *Paciente*, puesto que una de las condiciones de *Gauss Markov* es que no debe existir correlación entre los residuos y, por ello, de esta manera se controlaría la estructura de correlación entre las medidas repetidas para los mismos individuos. Así, de entre los modelos que contienen dicho factor aleatorio, se ha seleccionado como modelo final aquel que, según los criterios de información AIC/BIC, se ajusta mejor a los datos para un mayor porcentaje de genes que, en este caso, resulta ser el modelo *T Mixto* (ecuación B.5) (tablas B.11-B.12).

¹En otros análisis de expresión diferencial realizados en la tesis el nivel de significación se ha fijado en $\alpha_{\text{FDR}} = 0.05$. Sin embargo, estudiando la gran variabilidad de los datos de este experimento, se ha considerado que una mejor opción es fijar dicho nivel en $\alpha_{\text{FDR}} = 0.1$ para que, a pesar de cometer algún falso positivo más, detectar más genes diferencialmente expresados. Además, puesto que controlar la *tasa de hipótesis falsamente rechazadas (FDR)* significa que del total de DEG se espera un $\alpha_{\text{FDR}} \cdot 100$ % de falsos positivos, dado el bajo número de DEG se ha considerado aceptable que el 10 % de los DEG sean falsos positivos frente a un 5 % que se obtendría de un modo más convencional. Es importante siempre llegar a una situación de consenso entre la ganancia de potencia del test y el error de tipo I (falsos positivos) cometidos.

Tabla B.11: Porcentaje del mejor ajuste de todos los modelos evaluados.

| Modelo | n° Sondas (%) | | Criterio mínimo | |
|---|---------------|---------------|-----------------|--------|
| | AIC | BIC | AIC | BIC |
| Tejido: Tejido + (1 paciente) | 9887 (76.33) | 11080 (85.54) | -14.02 | -11.89 |
| TejidoSexo: Tejido + Sexo + (1 paciente) | 1777 (13.72) | 1180 (9.11) | -12.92 | -10.09 |
| Completo: Tejido + Sexo + Tejido:sexo + (1 paciente) | 1289 (9.95) | 693 (5.35) | -12.04 | -8.50 |

Tabla B.12: Porcentaje del mejor ajuste de todos los modelos evaluados (genes de interés).

| Modelo | n° Sondas (%) | | Criterio mínimo | |
|---|---------------|-------------|-----------------|--------|
| | AIC | BIC | AIC | BIC |
| Tejido + (1 paciente) | 218 (80.74) | 235 (87.04) | -17.71 | -15.59 |
| Tejido+Sexo + (1 paciente) | 29 (10.74) | 21 (7.78) | -16.77 | -13.94 |
| Tejido+Sexo+Tejido:Sexo + (1 paciente) | 23 (8.52) | 14 (5.19) | -16.42 | -12.88 |

Aunque pueda parecer que llevar a cabo la selección de variables anterior es rizar demasiado rizo, su importancia se hace aun más evidente al comparar comparan los resultados globales obtenidos tanto con el modelo *completo mixto* como con el *T mixto* a nivel de cada una de las variables que contienen:

- *Modelo Completo Mixto*: se identifica un 30.84 % (3996) de las sondas diferencialmente expresadas entre tumoral y normal, mientras que sólo lo son un 0.02 % (3) para el factor sexo y un 0 % (0) para la interacción. Como curiosidad, los genes que muestran expresión diferencial entre sexos están localizados en los cromosomas sexuales: DDX3Y y RPS4Y1 en el cromosoma Y, y XIST en el X. Además, ninguno de ellos muestra expresión diferencial entre las muestras de tejido de cáncer de colon y normal.
- *Modelo T Mixto*: a diferencia del modelo anterior, ahora se identifica un 43.64 % de las sondas (5649) como diferencialmente expresadas, casi un 50 % más.

Como resultado del análisis de expresión diferencial, de las 12944 analizadas (las que han quedado tras el filtrado) se han identificado 4435 sondas diferencialmente expresadas para un $\alpha_{\text{FDR}} = 0.05$ y 5649 para un $\alpha_{\text{FDR}} = 0.1$. Por lo tanto, es evidente que la potencia que se ha ganado (rechazar más hipótesis alternativas verdaderas) al identificar 1214 sondas a mayores al fijar $\alpha_{\text{FDR}} = 0.1$ en lugar de $\alpha_{\text{FDR}} = 0.05$ compensa los falsos positivos esperados (222 para $\alpha_{\text{FDR}} = 0.05$ frente a 565 fijando $\alpha_{\text{FDR}} = 0.1$). Como curiosidad, si no se hubiese tenido en cuenta el factor aleatorio, para $\alpha_{\text{FDR}} = 0.1$ se hubieran identificado sólo 4420 sondas como diferencialmente expresadas.

Por lo tanto, es evidente la importancia de un adecuado diseño experimental, modelado y selección de las variables, pues, en ocasiones, se pierde una gran capacidad de identificar genes diferencialmente expresados ante la falsa creencia de que más es mejor; es decir, de que cuantas más variables tenga el modelo mejor funciona y, realmente, esto puede generar más problemas que beneficios.

Puesto que el interés de la presente tesis gira en torno a los genes implicados en la homeostasis del Ca^{2+} intracelular, se mostrarán, a continuación, los resultados del análisis de expresión diferencial para estos genes, divididos por las familias en las que han sido previamente agrupados. En concreto, mediante diagramas de barras, se representa el logFC estimado en el eje X y los genes evaluados en el Y, de tal manera que las barras dirigidas hacia la izquierda y en azul indican una menor expresión diferencial significativa en el fenotipo tumoral respecto del tumoral, mientras que en rojo y hacia la derecha indican una sobreexpresión significativa en las muestras de tejido humano de cáncer de colon.

- *VOCC* y *CP*: En las figuras B.8 y B.9 se puede observar cómo ninguno de los genes de los conjuntos *VOCC* y *CP*, respectivamente, estás diferencialmente expresado entre el fenotipo tumoral y el normal.

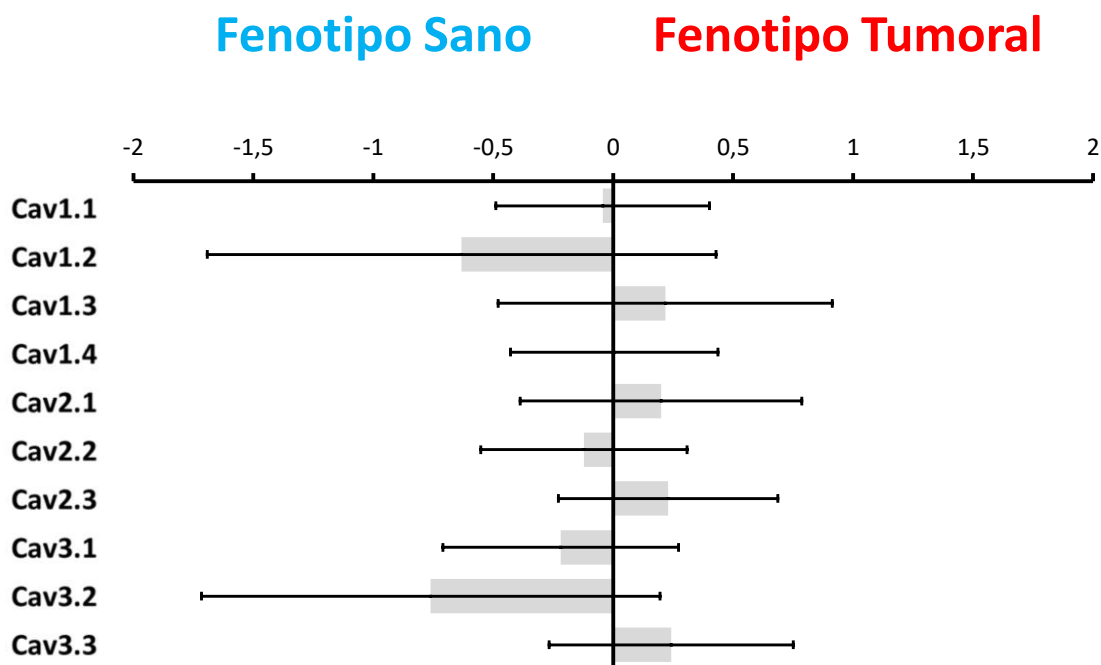


Figura B.8: Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes VOCC. Los valores de expresión diferencial, en concreto los valores log(FC) estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($\text{FDR} < 0.1$) en el tejido tumoral y una mayor expresión para las barras rojas hacia la derecha.

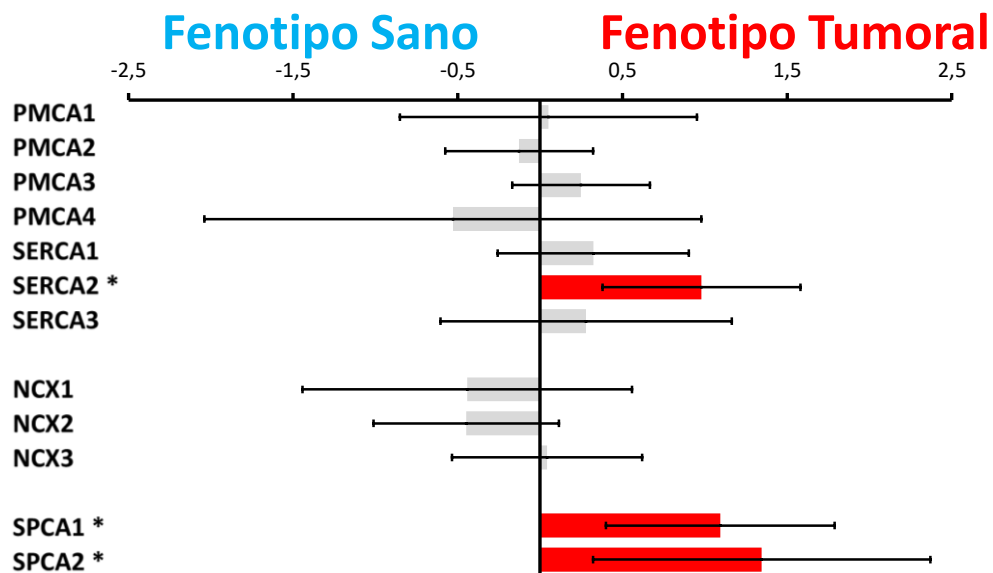


Figura B.9: Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes CP. Los valores de expresión diferencial, en concreto los valores log(FC) estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa (FDR<0.1) en el tejido tumoral y una mayor expresión para las barras rojas hacia la derecha.

- *CLC*: En la figura B.10 se puede observar cómo, de entre todos los *CLC*, sólo el IP3R3 muestra una sobreexpresión diferencial en el tejido tumoral, con un $FDR = 0.079$.

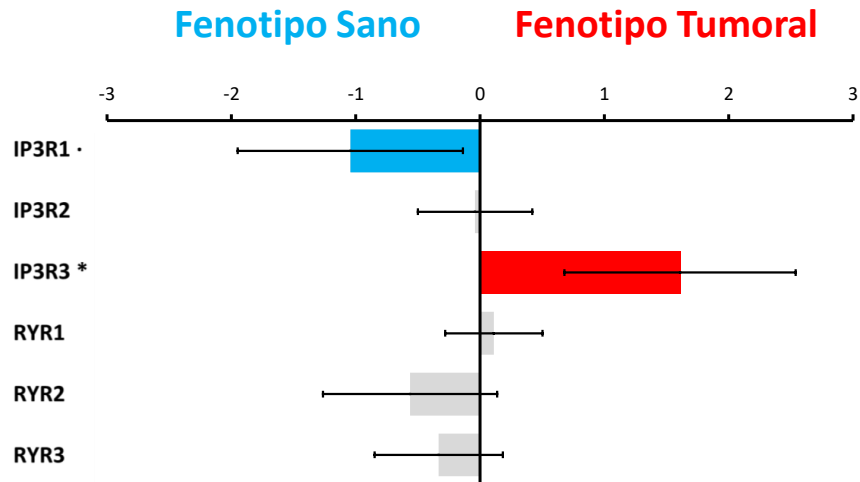


Figura B.10: Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes *CLC*. Los valores de expresión diferencial, en concreto los valores $\log_2(FC)$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($FDR < 0.1$) en el tejido tumoral y una mayor expresión para las barras rojas hacia la derecha.

- *SOCE*: En la figura B.11 se puede observar cómo el gen *MS4A12* está significativamente infraexpresado en el tejido tumoral.

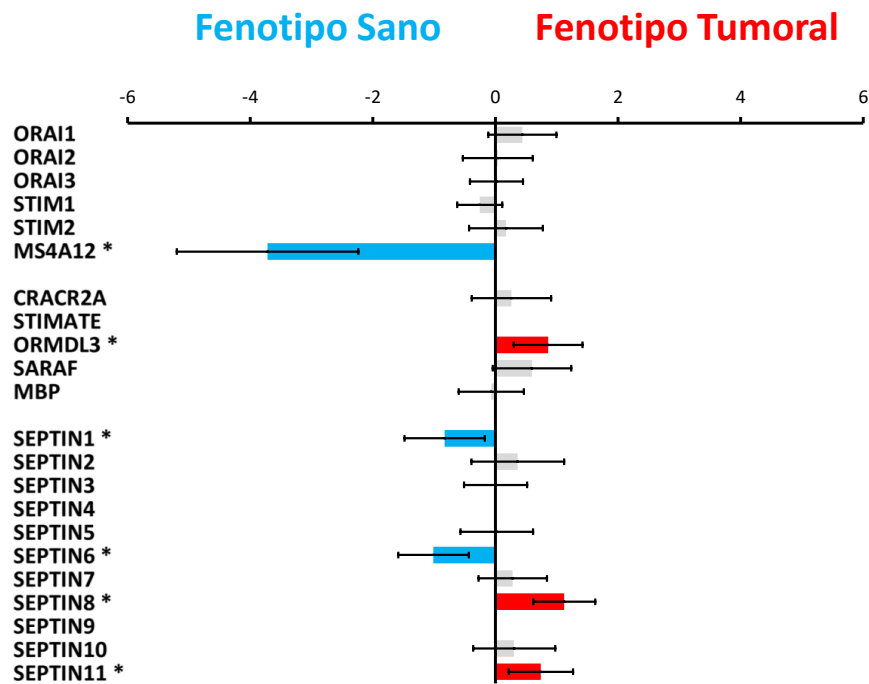


Figura B.11: Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes SOCE. Los valores de expresión diferencial, en concreto los valores log₂(FC) estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa (FDR<0.1) en el tejido tumoral y una mayor expresión para las barras rojas hacia la derecha.

- *TRP*: Como puede observarse, se detecta una infraexpresión diferencial de los genes TRPM2 y TRPP5, y una sobreexpresión del gen TRPV1 (figura B.12).

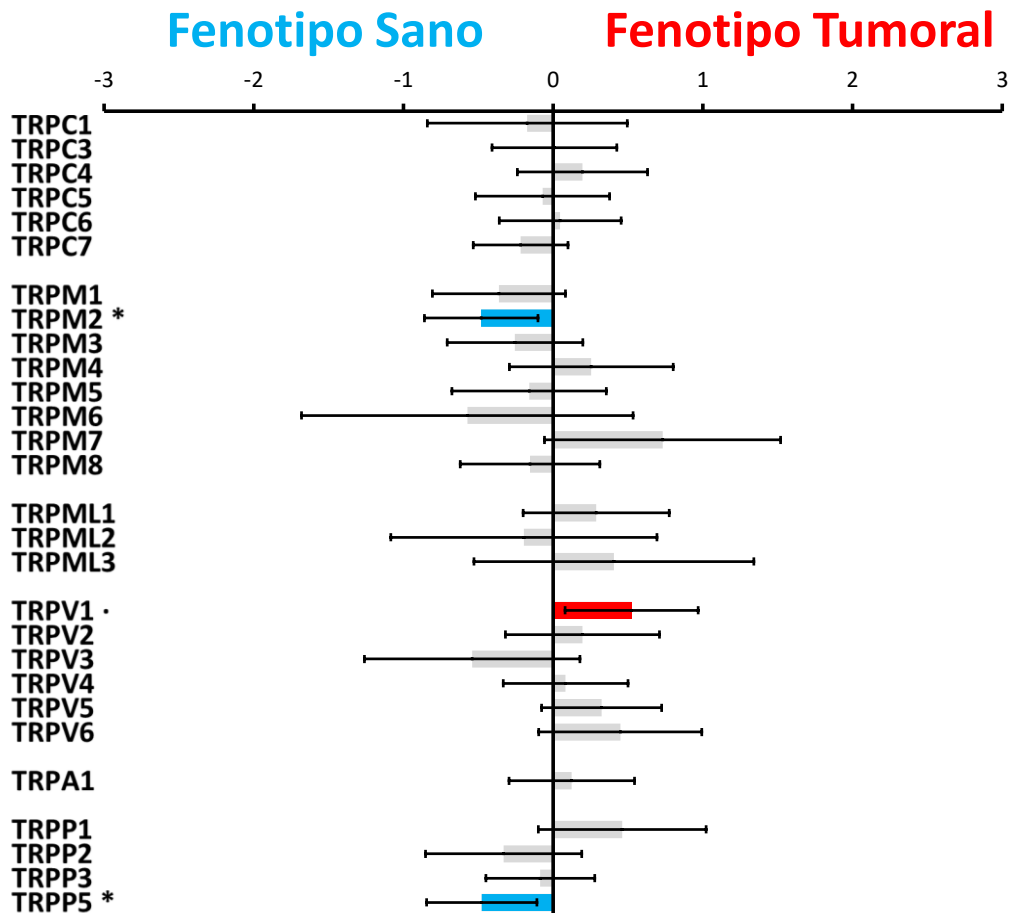


Figura B.12: Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes TRP. Los valores de expresión diferencial, en concreto los valores $\log_2(FC)$ estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa ($FDR < 0.1$) en el tejido tumoral y una mayor expresión para las barras rojas hacia la derecha.

- *MCT*: existe una sobreexpresión de la mayoría de los genes de este grupo; en concreto, de los genes *MCU*, *MICU1*, *MICU2*, *MCUR11*, *VDAC1*, *VDAC3* y *NCLX* (figura B.13).

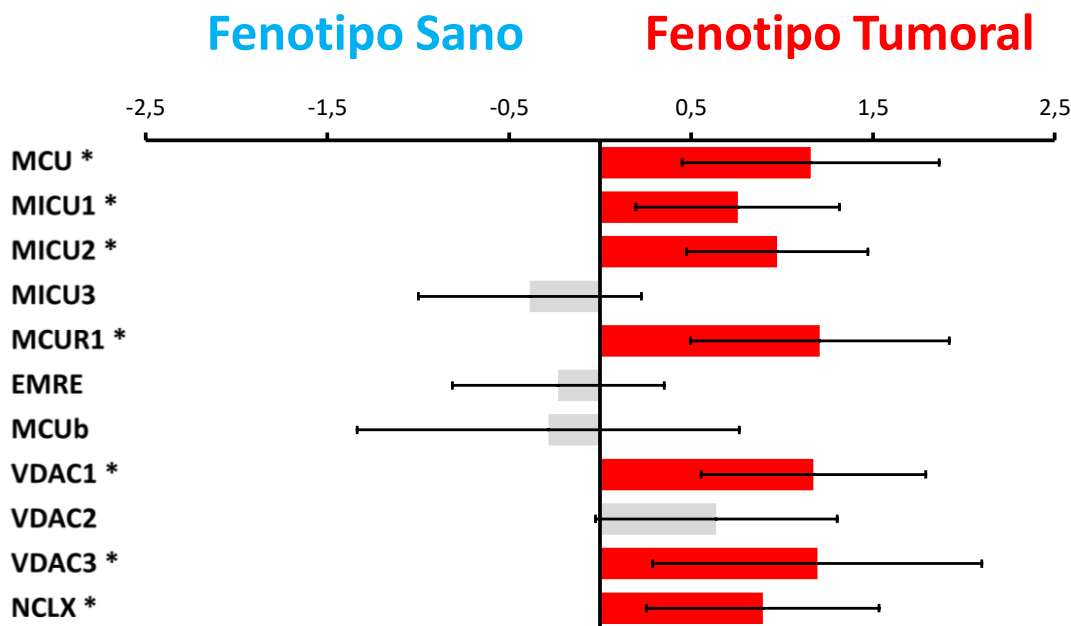


Figura B.13: Expresión diferencial en tejido humano de cáncer de colon vs. sano del conjunto de genes MCT. Los valores de expresión diferencial, en concreto los valores log₂(FC) estimados mediante el método limma, se representan en un diagrama de barras. Las barras azules hacia la izquierda indican una menor expresión significativa (FDR<0.1) en el tejido tumoral y una mayor expresión para las barras rojas hacia la derecha.

B.2.0.4. Análisis de significación biológica

A pesar de haber centrado el análisis anterior en los genes de interés, en realidad, el análisis se ha llevado a cabo sobre 12944 sondas y, de ellas, el 43.64 % muestran expresión diferencial. No obstante, manejar esta información no es tarea sencilla, pues hay que extraer información de una lista de 5649 genes diferencialmente expresados. Además, no siempre se tiene la suerte de tener un grupo de genes de interés, como es el presente caso con los genes de la homeostasis del Ca²⁺ intracelular en torno a los que gira la tesis. En consecuencia, manejar datos con tan alta dimensión puede acarrear diferentes problemas, como las comparaciones múltiples (un test por gen) o el sobreajuste en caso de clasificar la muestra en diferentes fenotipos o clases. Por lo tanto, el siguiente paso lógico es convertir esta lista de genes en una información biológicamente relevante, lo que se conoce en el campo de la bioinformática como análisis de significación biológica. Así, y debido a las características del experimento, como es el caso de una única comparación posible, *tejido de cáncer de colon humano vs. normal*, se van a emplear dos metodologías: *análisis de perfiles de expresión* y *análisis de enriquecimiento*.

B.2.0.4.1. Búsqueda de patrones: perfiles de expresión

Mediante la ayuda de un mapa de calor y el análisis clúster jerárquico, es posible representar y agrupar las diferentes muestras en el eje X y los genes diferencialmente expresados en el eje Y, de tal manera que se agruparán en función de su mayor o menor parecido mediante una técnica de aprendizaje no supervisado muy habitual: análisis clúster jerárquico aplicando el método de Ward (figura B.14). Así, por un lado, se observan dos conjuntos de muestras claramente diferenciadas y agrupadas entre sí sin información previa sobre su fenotipo: a la izquierda se agrupan las muestras tumorales y a la derecha las normales. Esta agrupación es una muy buena señal que, entre otras cosas, indica que la calidad de las muestras son buenas y que la mayor fuente de variabilidad es el tipo de tejido. Por otro lado, los genes se agrupan en dos conjuntos: un conjunto más reducido que muestra una menor expresión en las muestras de tejido tumoral y otro muy superior que muestra una clara sobreexpresión en dicho tejido respecto del normal; es decir, que se observa una clara sobreexpresión de genes en el tejido humano de cáncer de colon (figura B.14).

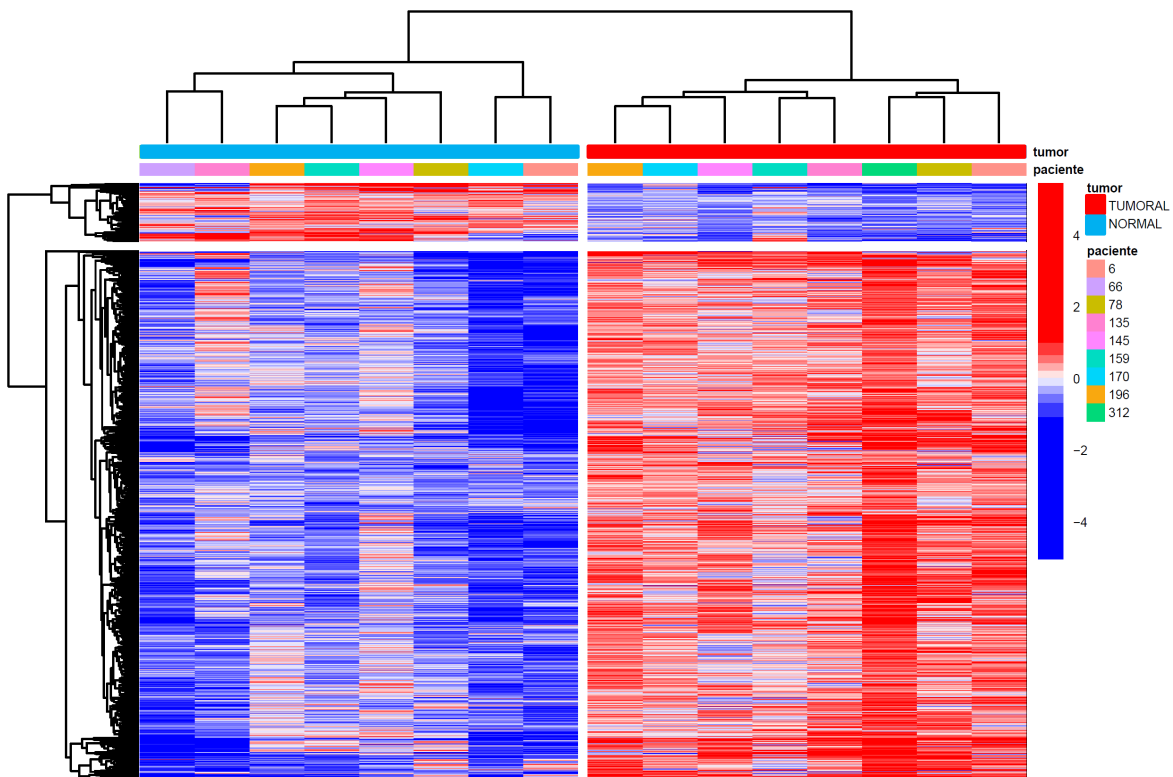


Figura B.14: Perfiles de expresión. En el mapa de calor se representan los valores centrados de la expresión en escala logarítmica. En el eje X se agrupan las diferentes muestras mediante el análisis clúster jerárquico basado en el método de Ward, y en el eje Y lo mismo pero con los genes diferencialmente expresados.

B.2.0.4.2. Búsqueda de patrones: componentes principales y algoritmos de clasificación

Una de las estrategias cuando se trabaja con datos de tan alta dimensionalidad es reducir su dimensión. Esto, cuando se trabaja con algoritmos de clasificación puede ser muy útil, pues puede ayudar a evitar problemas como el sobreajuste. En este caso, a partir de las variables originales, miles de genes, se calculan los componentes principales para reducir la dimensión y poder representar las muestras en dos dimensiones. No obstante, para encontrar qué componentes principales (PC) son las más relevantes a la hora de explicar las diferencias entre el tejido tumoral y el normal se ha empleado el algoritmo de aprendizaje automático supervisado *árboles de clasificación C5.0* implementado por la función `C50 : : C5.0 ()` (figura B.15). Afortunadamente, la PC1 es la que mejor explica estas diferencias, lo que supone que la mayor fuente de variabilidad en los datos es precisamente el tipo de tejido ya que la CP1 es capaz, por sí sola, de explicar las diferencias entre el tejido humano de CCR y el normal. Es más, las medidas de rendimiento indican una clasificación con una precisión del 100 %, una sensibilidad del 100 % o una tasa de error del 0 %.

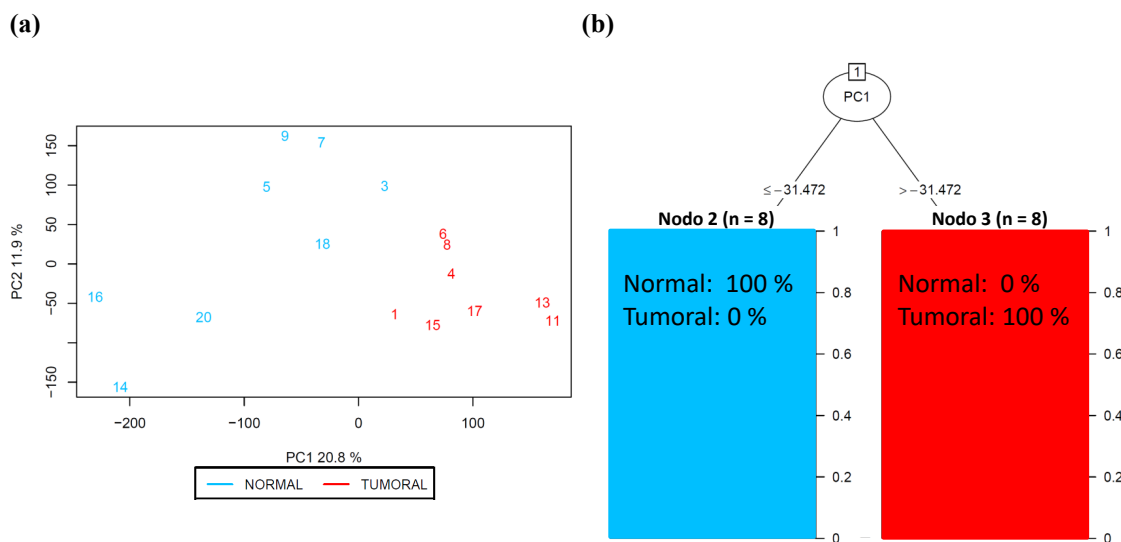


Figura B.15: Análisis de componentes principales y algoritmo de clasificación supervisada (árboles de clasificación). a) Análisis de componentes principales: sobre las muestras analizadas se a realizado un análisis de componentes principales para reducir su dimensión. b) Árboles de clasificación: sobre las componentes principales se aplica el algoritmo de clasificación supervisada *árboles de clasificación C5.0* para obtener qué PC es la más relevante en la clasificación de las muestras en función del tipo de tejido.

Tras encontrar la PC más relevante, PC1, y para darle más peso al hecho de que la PC1 es capaz de explicar las diferencias entre el tejido, se realiza un análisis de la varianza para un modelo lineal en el que el factor fijo sea el tipo de tejido y la variable respuesta la PC1, del cual se ha obtenido que, efectivamente, la PC1 se puede explicar por el factor *Tejido* (figura B.16a). Por otro lado, usando las variables PC1 y PC2 como explicativas se realiza un análisis cuadrático discriminante, otro algoritmo de clasificación no supervisado, donde puede observarse cómo también se obtiene una clasificación perfecta de las muestras en función del tipo de tejido con esta PC y la siguiente que más variabilidad de los datos explica, PC2 (figura B.16b).

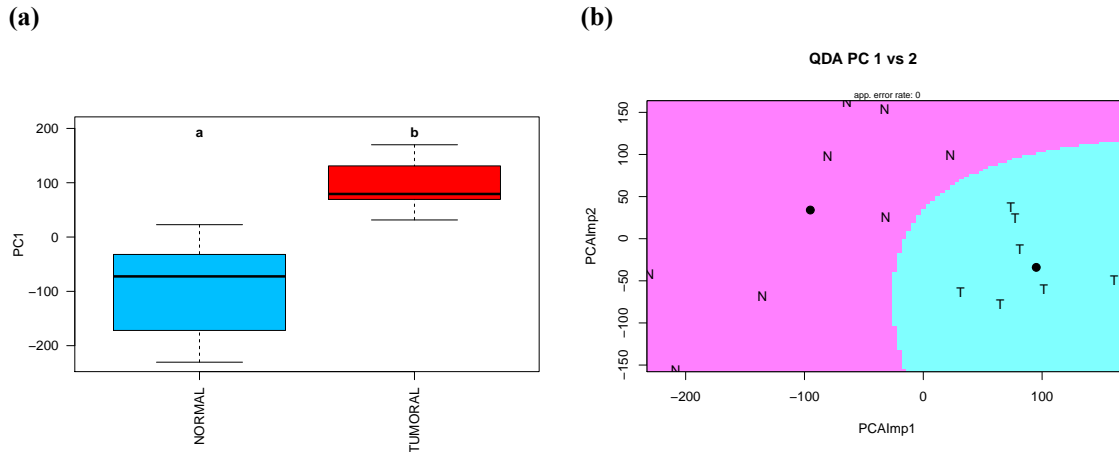


Figura B.16: Análisis de la varianza de la variable respuesta PC1 en función del tipo de tejido y algoritmo de clasificación supervisada (análisis cuadrático discriminante). a) Análisis de la varianza de la PC1: se hace un ANOVA del modelo lineal cuya variable respuesta es PC1 y el factor fijo es el tipo de tejido. b) Análisis cuadrático discriminante: usando las PC1 y PC2 como variables clasificadoras para explicar las diferencias entre las muestras de diferentes tejidos.

Tras estos resultados, parece lógico estudiar qué variables originales, genes, son las que más peso tienen o más correlacionadas están con esta PC1, pues serán aquellos que mejor expliquen las diferencias entre el tejido tumoral y el sano. Para ello, se estima la correlación para cada uno de los genes con la PC1 y se contrasta la hipótesis nula $H_0 : \rho_{\text{gen},PC1} = 0$ frente a la $H_1 : \rho_{\text{gen},PC1} \neq 0$, tal y como se explica en el apéndice “A.5 Coeficiente de correlación de Pearson”. Posteriormente, se corrigen los p valores controlando el FDR mediante la aproximación de Benjamini and Hochberg (1995). Este enfoque podría pensarse como un procedimiento alternativo de análisis de expresión diferencial. Así, de un modo similar a como se han mostrado los resultados de expresión diferencial, en el apartado “7.2.1.2.1 Búsqueda de patrones: componentes principales y algoritmos de clasificación” se han mostrado los valores de correlación entre la PC1 y la expresión de los genes relacionados con la homeostasis del Ca^{2+} intracelular (figuras 7.38 a 7.40).

B.2.0.4.3. Análisis de enriquecimiento

Otro enfoque para estudiar la lista de genes diferencialmente expresados obtenida anteriormente es, en vez de estudiar sonda a sonda o gen a gen qué significado tiene la expresión diferencial de cada uno de ellos, recurrir a información previamente generada por la comunidad científica sobre grupos de genes con significado biológico. Existen diferentes bases de datos con miles de conjuntos de genes y, de todas ellas, se han consultado las más usadas por la comunidad científica: KEGG PATHWAY, GO y MSigDB. Además, cada una de ellas tiene diferentes colecciones de conjuntos de genes en función de la temática que los relaciona. En concreto, se han consultado hasta 8 bases de datos/colecciones diferentes que contienen, en total, cerca de 30000 conjuntos de genes (tabla B.13). Por lo tanto, se cruzan los datos de la lista de genes generada por el análisis de expresión diferencial

y se contrastan la *hipótesis nula autocontenida*² -usando el método CAMERA- y la *hipótesis nula competitiva*³ -mediante el método ORA- para inferir qué conjuntos de genes están asociados con el tejido de cáncer de colon respecto del normal según su perfil de expresión diferencial. Sin entrar en mucho detalle, el primer método, ORA, comprueba si en la lista de genes hay algún conjunto sobrerrepresentado; es decir, si aparece con mayor frecuencia que otros conjuntos y, en tal caso, posiblemente esté asociado con el estudio que ha generado la lista de DEG. Para ello, emplea la distribución hipergeométrica y realiza el test unilateral de Fisher. En cuanto al segundo, el método CAMERA (del inglés *Correlation Adjusted MEan RANk gene set test*), el cual, a pesar basarse en la hipótesis nula de permutación de las etiquetas de los genes, considera la correlación entre los genes a partir de la estimación del factor de inflación de la varianza asociado a la correlación entre los genes y lo incorpora en los procedimientos tanto paramétricos como basados en pruebas de rangos. Además, cabe destacar que no asumen ninguna estructura de correlación y es adecuado para cualquier experimento que pueda ser representado no solo por comparar dos poblaciones, sino que además puede ser empleado en experimentos genómicos más complejos que han de ser modelizados a través de modelos lineales. Puesto que sólo estima un parámetro de correlación, este procedimiento permanece estable y es capaz de controlar de forma adecuada la tasa de error tipo I incluso cuando el tamaño muestral es reducido.

Puesto que la cantidad de conjuntos es demasiado elevada como para estudiar el significado de cada uno de ellos, de los que rechazan alguna de las hipótesis nulas se han buscado cuántos pertenecen a bases de datos relacionadas con el cáncer (*Oncogenic Signature Gene Sets* o *Computational Gene Sets*) y conjuntos que en su nombre contienen alguna de las siguientes palabras de interés: *calcium, mitochondria, reticulum, cancer, wnt* y *polyamine*. De esta forma, es posible deducir qué tipo de conjuntos de interés, muy a grandes rasgos, podrían estar asociados con el fenotipo tumoral. En la tabla B.13 se muestra un resumen numérico de los conjuntos de genes que rechazan alguna de las dos hipótesis nulas para cada una de las bases de datos y, además, cuántos de ellos contienen alguna de las palabras clave.

²la hipótesis nula autocontenida afirma que ningún gen en el conjunto S está diferencialmente expresado (Goeman and Bühlmann (2007)). Es decir, que examina un conjunto de genes de forma aislada sin considerar otros genes en el genoma. Esta hipótesis es adecuada para evaluar la relevancia de un único proceso biológico en un experimento dado que es fácilmente calculado a mano.

³la hipótesis nula competitiva plantea que los genes en un grupo dado S están como mucho tan frecuentemente expresados de un modo diferencial como los genes en el universo de genes (Goeman and Bühlmann (2007)). A diferencia de la hipótesis nula autocontenida, la competitiva sí usa información de otros genes del genoma, pues compara los genes del conjunto considerado respecto al resto de genes. Esta es adecuada para encontrar aquellos procesos biológicos (u otros conjuntos de genes) más relevantes respecto de aquellos que lo son menos. Además, esta es la más usada en el análisis de datos genómicos (Di Wu and Smyth (2012)). Por lo general, las hipótesis nulas competitivas asumen independencia entre los genes (gran error), pues o bien aplican la distribución de permutación de las etiquetas de los genes para estimar el p valor, o bien emplean aproximaciones paramétricas que, de forma asintótica, se acercan a la distribución de permutación. No obstante, se ha encontrado que estos test son muy sensibles a la correlación entre genes que puede dar lugar a inflar los valores de FDR. Por ello, muchos autores sugieren, en lugar de permutar las etiquetas de los genes, hacerlo con las de las muestras, como hace el método GSEA, ampliamente usado, propuesto por Mootha et al. (2003) (ver guía de usuario GSEA) ya que, de esta manera, sí se tiene en consideración la correlación entre los diferentes genes. No obstante, cuando el número de muestras es reducida no es posible aplicar estos métodos.

Tabla B.13: Análisis de conjuntos de genes para la expresión diferencial entre el tejido humano de CCR y el normal.

| Base de Datos | Conjuntos | | Conjuntos Significativos (FDR <0.05) | | | | | | | |
|---------------|--------------|---------------------|--------------------------------------|-------------|---------------------|--------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | Métodos | Totales | Calcio ¹ | Mitocondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA | 84 | 1 | 0 | 1 | 0 | 0 | 0 |
| | | | ORA | 11 | 0 | 0 | 1 | 0 | 0 | 0 |
| GO BP | 12342 | 146 | CAMERA | 198 | 8 | 3 | 2 | 0 | 0 | 0 |
| | | | ORA | 62 | 0 | 2 | 0 | 0 | 0 | 0 |
| GO MF | 4245 | 66 | CAMERA | 66 | 2 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| MH | 50 | 0 | CAMERA | 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 9 | 0 | 0 | 0 | 0 | 0 | 0 |
| MCDB | 6290 | 12 | CAMERA | 1193 | 3 | 6 | 0 | 130 | 5 | 1 |
| | | | ORA | 767 | 0 | 6 | 0 | 89 | 4 | 1 |
| MCC | 858 | 0 | CAMERA | 424 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 227 | 0 | 0 | 0 | 0 | 0 | 0 |
| MOG | 189 | 0 | CAMERA | 46 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| MIM | 5219 | 0 | CAMERA | 842 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 1124 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 29537 | 225 | CAMERA | 2870 | 14 | 9 | 3 | 130 | 5 | 1 |
| | | | ORA | 2232 | 0 | 8 | 1 | 89 | 4 | 1 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

Como puede observarse, de los 29537 conjuntos de genes analizados, 2870 rechazan la hipótesis nula autocontenida y 2232 la competitiva. De estos que rechazan la hipótesis nula autocontenida/competitiva, 424/227 pertenecen a la base de datos MCC, 46/15 a la base MOG, 14/0 contienen la palabra *calcium*, 9/8 la palabra *mitochondria*, 3/1 *reticulum*, 130/89 *cancer*, 5/4 *wnt*, y 1/1 *polyamine*.

Algunos de estos conjuntos están relacionados con:

- El transporte de Ca^{2+} : *hsa04020 Calcium signaling pathway*, *GO:0051924 regulation of calcium ion transport*, *GO:0010880 regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum*, *GO:0019722 calcium-mediated signaling*,...
- El cáncer de colon: *M16740 GRADE COLON AND RECTAL CANCER UP*, *M17423 LAIHO COLORECTAL CANCER SERRATED DN*, *M15780 GRADE COLON AND RECTAL CANCER DN*,...
- El procesamiento de proteínas en el retículo: *Protein processing in endoplasmic reticulum*.
- La maquinaria de traducción mitocondrial: *GO:0070126 mitochondrial translational termination*, *GO:0070125 mitochondrial translational elongation*...
- La ruta de señalización WNT : *GO:0060071 Wnt signaling pathway*, *planar cell polarity pathway*, *M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT*, *M1758 SANSOM WNT PATHWAY REQUIRE MYC*, *M7847 REACTOME SIGNALING BY WNT*, *M39731 WP LNCRNA INVOLVEMENT IN CANONICAL WNT SIGNALING AND COLORECTAL CANCER*...
- ...

No obstante, como se refleja la tabla 7.7, el número de conjuntos de interés es mucho mayor. Por ello, a continuación, se pone a disposición del lector dividido en función, de la base de datos a la que pertenecen, aquellos conjuntos con un FDR < 0.1 cuyos nombres contienen alguna de las palabras clave (tablas B.14-B.21).

Tabla B.14: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA).

| Conjuntos | nº Genes | Sentido | P-valor | FDR |
|--|----------|---------|----------|-----------|
| GO:0010881 regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion | 17 | -1 | 4.00e-07 | 0.0002036 |
| GO:0070126 mitochondrial translational termination | 71 | 1 | 1.10e-06 | 0.0002965 |
| GO:0070125 mitochondrial translational elongation | 69 | 1 | 2.80e-06 | 0.0005091 |
| GO:0016339 calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules | 26 | -1 | 2.43e-05 | 0.0025042 |
| GO:0051480 regulation of cytosolic calcium ion concentration | 20 | -1 | 1.26e-04 | 0.0075910 |
| GO:0051924 regulation of calcium ion transport | 17 | -1 | 1.66e-04 | 0.0087845 |

Tabla B.14: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| GO:0006890 retrograde vesicle-mediated transport, Golgi to endoplasmic reticulum | 63 | 1 | 2.46e-04 | 0.0120417 |
| GO:0010880 regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum | 13 | -1 | 3.55e-04 | 0.0160241 |
| GO:0010524 positive regulation of calcium ion transport into cytosol | 8 | -1 | 5.20e-04 | 0.0199639 |
| GO:0032543 mitochondrial translation | 27 | 1 | 5.57e-04 | 0.0211768 |
| GO:0006816 calcium ion transport | 38 | -1 | 1.15e-03 | 0.0341167 |
| GO:0007204 positive regulation of cytosolic calcium ion concentration | 83 | -1 | 1.72e-03 | 0.0408213 |
| GO:0019722 calcium-mediated signaling | 46 | -1 | 2.78e-03 | 0.0538935 |
| GO:0051209 release of sequestered calcium ion into cytosol | 21 | -1 | 5.12e-03 | 0.0801997 |
| GO:0006888 endoplasmic reticulum to Golgi vesicle-mediated transport | 123 | 1 | 5.23e-03 | 0.0806145 |
| GO:0005513 detection of calcium ion | 10 | -1 | 6.06e-03 | 0.0892141 |
| GO:0060071 Wnt signaling pathway, planar cell polarity pathway | 70 | 1 | 6.97e-03 | 0.0976223 |
| GO:0060315 negative regulation of ryanodine-sensitive calcium-release channel activity | 7 | -1 | 7.00e-03 | 0.0976223 |

Tabla B.15: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (ORA).

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| GO:0070126 mitochondrial translational termination | 71 | 54 | 1.00e-06 | 0.0003178 |
| GO:0070125 mitochondrial translational elongation | 69 | 52 | 2.50e-06 | 0.0006553 |
| GO:0006888 endoplasmic reticulum to Golgi vesicle-mediated transport | 123 | 76 | 1.13e-03 | 0.0670467 |
| GO:0032543 mitochondrial translation | 27 | 21 | 1.40e-03 | 0.0772169 |
| GO:0060071 Wnt signaling pathway, planar cell polarity pathway | 70 | 46 | 1.79e-03 | 0.0947249 |
| GO:0006890 retrograde vesicle-mediated transport, Golgi to endoplasmic reticulum | 63 | 42 | 1.84e-03 | 0.0960725 |

Tabla B.16: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOMF (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|---|---------|---------|---------|-----------|
| GO:0005509 calcium ion binding | 406 | -1 | 0.00203 | 0.0413780 |
| GO:0005246 calcium channel regulator activity | 17 | -1 | 0.00254 | 0.0475566 |

Tabla B.17: Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| hsa04020 Calcium signaling pathway | 126 | -1 | 4.17e-05 | 0.0006985 |
| hsa04141 Protein processing in endoplasmic reticulum | 133 | 1 | 1.54e-04 | 0.0017790 |

Tabla B.18: Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (ORA).

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| hsa04141 Protein processing in endoplasmic reticulum | 133 | 88 | 7.38e-05 | 0.0024723 |

Tabla B.19: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| M15472 VECCHI GASTRIC CANCER EARLY DN | 234 | -1 | 0.00e+00 | 0.0000000 |
| M8513 SMID BREAST CANCER NORMAL LIKE UP | 335 | -1 | 0.00e+00 | 0.0000000 |
| M16740 GRADE COLON AND RECTAL CANCER UP | 267 | 1 | 0.00e+00 | 0.0000000 |
| M18855 VECCHI GASTRIC CANCER EARLY UP | 358 | 1 | 0.00e+00 | 0.0000000 |
| M3766 SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP | 132 | 1 | 0.00e+00 | 0.0000000 |
| M15664 ROSTY CERVICAL CANCER PROLIFERATION CLUSTER | 123 | 1 | 0.00e+00 | 0.0000000 |
| M10319 WALLACE PROSTATE CANCER RACE UP | 212 | -1 | 0.00e+00 | 0.0000000 |
| M7854 WHITEFORD PEDIATRIC CANCER MARKERS | 105 | 1 | 0.00e+00 | 0.0000000 |
| M8520 SHEDDEN LUNG CANCER POOR SURVIVAL A6 | 346 | 1 | 0.00e+00 | 0.0000000 |
| M8365 RHODES UNDIFFERENTIATED CANCER | 62 | 1 | 0.00e+00 | 0.0000000 |
| M4995 LINDGREN BLADDER CANCER CLUSTER 2B | 303 | -1 | 0.00e+00 | 0.0000000 |
| M1345 RHODES CANCER META SIGNATURE | 58 | 1 | 0.00e+00 | 0.0000000 |
| M13273 DELYS THYROID CANCER DN | 169 | -1 | 0.00e+00 | 0.0000000 |
| M7517 SMID BREAST CANCER LUMINAL A UP | 55 | -1 | 0.00e+00 | 0.0000000 |
| M7987 LEE LIVER CANCER SURVIVAL DN | 156 | 1 | 0.00e+00 | 0.0000000 |
| M5792 LI AMPLIFIED IN LUNG CANCER | 156 | 1 | 0.00e+00 | 0.0000000 |
| M10986 CHIANG LIVER CANCER SUBCLASS UNANNOTATED DN | 163 | 1 | 0.00e+00 | 0.0000000 |
| M39678 WP RETINOBLASTOMA GENE IN CANCER | 76 | 1 | 0.00e+00 | 0.0000000 |

Tabla B.19: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| M18436 BOYAULT LIVER CANCER SUBCLASS G3 UP | 153 | 1 | 0.00e+00 | 0.0000001 |
| M3268 CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP | 149 | 1 | 0.00e+00 | 0.0000001 |
| M17333 FARMER BREAST CANCER CLUSTER 2 | 31 | 1 | 0.00e+00 | 0.0000003 |
| M854 LINDGREN BLADDER CANCER CLUSTER 3 UP | 259 | 1 | 0.00e+00 | 0.0000005 |
| M9826 VANTVEER BREAST CANCER METASTASIS DN | 82 | 1 | 0.00e+00 | 0.0000007 |
| M11504 TOMLINS PROSTATE CANCER DN | 34 | -1 | 0.00e+00 | 0.0000009 |
| M2043 RICKMAN HEAD AND NECK CANCER F | 33 | -1 | 0.00e+00 | 0.0000009 |
| M19391 LIU PROSTATE CANCER DN | 356 | -1 | 0.00e+00 | 0.0000018 |
| M7281 KONDO PROSTATE CANCER WITH H3K27ME3 | 65 | -1 | 1.00e-07 | 0.0000028 |
| M17572 SMID BREAST CANCER LUMINAL B DN | 401 | -1 | 1.00e-07 | 0.0000030 |
| M1311 WONG ENDMETRIUM CANCER DN | 50 | -1 | 1.00e-07 | 0.0000034 |
| M336 VILLANUEVA LIVER CANCER KRT19 UP | 132 | 1 | 2.00e-07 | 0.0000071 |
| M17423 LAIHO COLORECTAL CANCER SERRATED DN | 66 | 1 | 3.00e-07 | 0.0000084 |
| M13072 SMID BREAST CANCER LUMINAL A DN | 15 | 1 | 3.00e-07 | 0.0000089 |
| M27446 REACTOME MITOCHONDRIAL TRANSLATION | 74 | 1 | 4.00e-07 | 0.0000110 |
| M19766 WU SILENCED BY METHYLATION IN BLADDER CANCER | 41 | -1 | 4.00e-07 | 0.0000125 |
| M19982 WAMUNYOKOLI OVARIAN CANCER LMP UP | 217 | 1 | 5.00e-07 | 0.0000143 |
| M15356 PUJANA BREAST CANCER LIT INT NETWORK | 80 | 1 | 1.20e-06 | 0.0000301 |
| M15150 ZHANG BREAST CANCER PROGENITORS UP | 352 | 1 | 1.40e-06 | 0.0000357 |
| M39740 WP METABOLIC REPROGRAMMING IN COLON CANCER | 33 | 1 | 1.50e-06 | 0.0000372 |
| M12016 OUELLET OVARIAN CANCER INVASIVE VS LMP UP | 98 | 1 | 2.00e-06 | 0.0000472 |
| M13831 BOYAULT LIVER CANCER SUBCLASS G123 UP | 38 | 1 | 2.20e-06 | 0.0000530 |
| M2082 WANG METASTASIS OF BREAST CANCER ESRI UP | 16 | 1 | 2.40e-06 | 0.0000556 |
| M2180 PUJANA BREAST CANCER WITH BRCA1 MUTATED UP | 45 | 1 | 2.90e-06 | 0.0000663 |
| M11629 MONTERO THYROID CANCER POOR SURVIVAL UP | 9 | 1 | 3.90e-06 | 0.0000849 |

Tabla B.19: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| M11616 ACEVEDO LIVER CANCER WITH H3K27ME3 UP | 157 | -1 | 4.60e-06 | 0.0000993 |
| M39347 WP GASTRIC CANCER NETWORK 1 | 25 | 1 | 4.60e-06 | 0.0000994 |
| M7995 HOSHIDA LIVER CANCER SUBCLASS S2 | 93 | 1 | 4.60e-06 | 0.0000994 |
| M17471 SCHUETZ BREAST CANCER DUCTAL INVASIVE UP | 273 | -1 | 4.70e-06 | 0.0000999 |
| M4572 LOCKWOOD AMPLIFIED IN LUNG CANCER | 143 | 1 | 5.30e-06 | 0.0001105 |
| M9389 FINETTI BREAST CANCER KINOME RED | 15 | 1 | 7.30e-06 | 0.0001429 |
| M2435 WAMUNYOKOLI OVARIAN CANCER GRADES 1 2 UP | 127 | 1 | 8.90e-06 | 0.0001680 |
| M7396 SWEET LUNG CANCER KRAS DN | 290 | -1 | 9.10e-06 | 0.0001701 |
| M8655 DAIRKEE CANCER PRONE RESPONSE BPA | 34 | 1 | 1.00e-05 | 0.0001834 |
| M6862 NADERI BREAST CANCER PROGNOSIS UP | 38 | 1 | 1.22e-05 | 0.0002129 |
| M18973 NIKOLSKY BREAST CANCER 20Q12 Q13 AMPLICON | 86 | 1 | 1.26e-05 | 0.0002186 |
| M11302 SMID BREAST CANCER RELAPSE IN LUNG DN | 25 | -1 | 1.43e-05 | 0.0002451 |
| M4448 BOYVAULT LIVER CANCER SUBCLASS G23 UP | 41 | 1 | 1.51e-05 | 0.0002566 |
| M2890 KEGG CALCIUM SIGNALING PATHWAY | 96 | -1 | 2.39e-05 | 0.0003712 |
| M2427 NAKAMURA CANCER MICROENVIRONMENT DN | 36 | 1 | 3.56e-05 | 0.0005301 |
| M13736 PYEON CANCER HEAD AND NECK VS CERVICAL UP | 152 | 1 | 3.84e-05 | 0.0005630 |
| M1687 ZHONG SECRETOME OF LUNG CANCER AND FIBROBLAST | 103 | 1 | 3.87e-05 | 0.0005660 |
| M707 WAMUNYOKOLI OVARIAN CANCER GRADES 1 2 DN | 50 | -1 | 4.24e-05 | 0.0006063 |
| M14228 RICKMAN HEAD AND NECK CANCER A | 68 | -1 | 4.28e-05 | 0.0006089 |
| M15780 GRADE COLON AND RECTAL CANCER DN | 56 | -1 | 4.65e-05 | 0.0006520 |
| M15112 WONG MITOCHONDRIA GENE MODULE | 177 | 1 | 6.40e-05 | 0.0008578 |
| M19837 RICKMAN HEAD AND NECK CANCER E | 46 | -1 | 9.51e-05 | 0.0011956 |
| M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT | 181 | 1 | 9.56e-05 | 0.0011979 |
| M16542 YAMASHITA LIVER CANCER WITH EPCAM UP | 41 | 1 | 9.58e-05 | 0.0011979 |
| M12692 FINETTI BREAST CANCER KINOME GREEN | 13 | -1 | 1.10e-04 | 0.0013407 |

Tabla B.19: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|---|---------|---------|----------|-----------|
| M15015 FARMER BREAST CANCER CLUSTER 1 | 27 | -1 | 1.29e-04 | 0.0015209 |
| M3710 WALLACE PROSTATE CANCER UP | 14 | 1 | 1.30e-04 | 0.0015264 |
| M6754 ACEVEDO LIVER CANCER WITH H3K9ME3 UP | 77 | -1 | 1.46e-04 | 0.0016768 |
| M17094 GINESTIER BREAST CANCER ZNF217 AMPLIFIED DN | 207 | 1 | 1.66e-04 | 0.0018621 |
| M9577 MOOTHA MITOCHONDRIA | 321 | 1 | 1.85e-04 | 0.0020474 |
| M1685 ZHONG SECRETOME OF LUNG CANCER AND MACROPHAGE | 63 | 1 | 1.95e-04 | 0.0021457 |
| M416 ANDERSEN LIVER CANCER KRT19 UP | 30 | 1 | 2.14e-04 | 0.0023108 |
| M1260 WATANABE RECTAL CANCER RADIOTHERAPY RESPONSIVE UP | 76 | 1 | 2.17e-04 | 0.0023301 |
| M7727 RICKMAN HEAD AND NECK CANCER B | 24 | -1 | 2.29e-04 | 0.0024408 |
| M10091 WAMUNYOKOLI OVARIAN CANCER LMP DN | 149 | -1 | 2.68e-04 | 0.0027846 |
| M747 REACTOME METABOLISM OF POLYAMINES | 51 | 1 | 2.84e-04 | 0.0029298 |
| M17299 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL UP | 338 | 1 | 3.33e-04 | 0.0033225 |
| M1255 INAMURA LUNG CANCER SCC DN | 9 | -1 | 3.40e-04 | 0.0033691 |
| M19148 TOMLINS PROSTATE CANCER UP | 36 | 1 | 3.47e-04 | 0.0034150 |
| M10382 IIZUKA LIVER CANCER EARLY RECURRENCE | 10 | -1 | 3.59e-04 | 0.0035152 |
| M836 OUELLET CULTURED OVARIAN CANCER INVASIVE VS LMP UP | 50 | 1 | 4.48e-04 | 0.0041743 |
| M39497 WP GASTRIC CANCER NETWORK 2 | 26 | 1 | 4.77e-04 | 0.0043879 |
| M16374 IIZUKA LIVER CANCER PROGRESSION G1 G2 DN | 22 | 1 | 4.87e-04 | 0.0044586 |
| M14670 DING LUNG CANCER EXPRESSION BY COPY NUMBER | 84 | 1 | 4.95e-04 | 0.0045247 |
| M378 MINGUEZ LIVER CANCER VASCULAR INVASION UP | 11 | 1 | 4.97e-04 | 0.0045358 |
| M39329 WP CALCIUM REGULATION IN THE CARDIAC CELL | 95 | -1 | 6.36e-04 | 0.0055838 |
| M18761 SMITH LIVER CANCER | 39 | 1 | 6.96e-04 | 0.0059656 |
| M7547 JIANG HYPOXIA CANCER | 46 | 1 | 7.25e-04 | 0.0061412 |
| M4619 LIU PROSTATE CANCER UP | 78 | 1 | 7.45e-04 | 0.0062922 |
| M12795 CHARAFE BREAST CANCER BASAL VS MESENCHYMAL UP | 93 | 1 | 7.80e-04 | 0.0065023 |
| M10837 NIKOLSKY BREAST CANCER 20Q11 AMPLICON | 23 | 1 | 7.98e-04 | 0.0066332 |
| M8526 CHEN LUNG CANCER SURVIVAL | 20 | 1 | 1.02e-03 | 0.0080499 |

Tabla B.19: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| M39628 WP INTEGRATED CANCER PATHWAY | 36 | 1 | 1.10e-03 | 0.0085758 |
| M8779 LI LUNG CANCER | 28 | 1 | 1.18e-03 | 0.0090529 |
| M16932 CHIANG LIVER CANCER SUBCLASS PROLIFERATION DN | 89 | -1 | 1.20e-03 | 0.0091942 |
| M19988 GINESTIER BREAST CANCER 20Q13 AMPLIFICATION DN | 145 | 1 | 1.31e-03 | 0.0098926 |
| M13007 GLINSKY CANCER DEATH UP | 6 | 1 | 1.39e-03 | 0.0103881 |
| M38998 REACTOME ABERRANT REGULATION OF MITOTIC G1 S TRANSITION IN CANCER DUE TO RB1 DEFECTS | 12 | 1 | 1.67e-03 | 0.0119106 |
| M1758 SANSOM WNT PATHWAY REQUIRE MYC | 52 | 1 | 1.81e-03 | 0.0127206 |
| M924 REACTOME PLATELET CALCIUM HOMEOSTASIS | 18 | -1 | 1.82e-03 | 0.0127599 |
| M867 MEINHOLD OVARIAN CANCER LOW GRADE DN | 18 | 1 | 1.83e-03 | 0.0127989 |
| M672 CHIN BREAST CANCER COPY NUMBER UP | 20 | 1 | 1.87e-03 | 0.0130314 |
| M15096 REACTOME MITOCHONDRIAL TRNA AMINOACYLATION | 18 | 1 | 2.03e-03 | 0.0140111 |
| M15147 BOYALT LIVER CANCER SUBCLASS G12 UP | 30 | 1 | 2.07e-03 | 0.0141687 |
| M9565 AGUIRRE PANCREATIC CANCER COPY NUMBER UP | 207 | 1 | 2.36e-03 | 0.0157071 |
| M1253 INAMURA LUNG CANCER SCC UP | 8 | 1 | 2.38e-03 | 0.0158037 |
| M15425 VANTVEER BREAST CANCER ESR1 DN | 161 | 1 | 2.46e-03 | 0.0161234 |
| M2837 BARIS THYROID CANCER UP | 17 | 1 | 2.53e-03 | 0.0164592 |
| M1273 SHETH LIVER CANCER VS TXNIP LOSS PAM1 | 182 | 1 | 2.62e-03 | 0.0169304 |
| M11288 SCHLESINGER METHYLATED DE NOVO IN CANCER | 51 | -1 | 2.84e-03 | 0.0180690 |
| M590 REACTOME MITOCHONDRIAL PROTEIN IMPORT | 47 | 1 | 2.98e-03 | 0.0187942 |
| M12533 WALLACE PROSTATE CANCER DN | 5 | -1 | 2.98e-03 | 0.0187942 |
| M38983 MEBARKI HCC PROGENITOR WNT DN | 50 | -1 | 3.08e-03 | 0.0192567 |
| M11896 LINDGREN BLADDER CANCER CLUSTER 2A DN | 110 | -1 | 3.20e-03 | 0.0198139 |
| M1256 INAMURA LUNG CANCER SCC SUBTYPES UP | 11 | 1 | 3.32e-03 | 0.0204256 |
| M2429 SHEDDEN LUNG CANCER GOOD SURVIVAL A12 | 181 | -1 | 3.43e-03 | 0.0210128 |
| M7371 LOPES METHYLATED IN COLON CANCER UP | 13 | -1 | 3.56e-03 | 0.0216258 |
| M19097 SWEET LUNG CANCER KRAS UP | 346 | 1 | 3.63e-03 | 0.0219357 |

Tabla B.19: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|---|---------|---------|----------|-----------|
| M13462 LUI THYROID CANCER CLUSTER 5 | 12 | -1 | 3.69e-03 | 0.0222517 |
| M38989 MEBARKI HCC PROGENITOR WNT DN BLOCKED BY FZD8CRD | 40 | -1 | 3.71e-03 | 0.0223257 |
| M8692 NIKOLSKY BREAST CANCER 19Q13.1 AMPLICON | 10 | -1 | 3.74e-03 | 0.0224361 |
| M4290 LUI THYROID CANCER CLUSTER 3 | 22 | 1 | 3.88e-03 | 0.0231080 |
| M15807 BONOME OVARIAN CANCER POOR SURVIVAL DN | 16 | 1 | 4.02e-03 | 0.0237460 |
| M6875 LANDIS BREAST CANCER PROGRESSION UP | 37 | 1 | 4.05e-03 | 0.0238500 |
| M3102 SMID BREAST CANCER ERBB2 UP | 100 | 1 | 4.26e-03 | 0.0247089 |
| M12895 CHARAFE BREAST CANCER BASAL VS MESENCHYMAL DN | 35 | -1 | 4.46e-03 | 0.0255863 |
| M4333 LUI THYROID CANCER CLUSTER 4 | 12 | -1 | 4.87e-03 | 0.0273943 |
| M17788 LINDGREN BLADDER CANCER CLUSTER 1 DN | 279 | 1 | 5.10e-03 | 0.0284115 |
| M14146 BOYVAULT LIVER CANCER SUBCLASS G1 UP | 92 | 1 | 5.94e-03 | 0.0321004 |
| M1686 ZHONG SECRETOME OF LUNG CANCER AND ENDOTHELIUM | 53 | 1 | 7.00e-03 | 0.0368633 |
| M18694 LIU COMMON CANCER GENES | 40 | 1 | 7.10e-03 | 0.0373218 |
| M8411 LINDGREN BLADDER CANCER WITH LOH IN CHR9Q | 88 | 1 | 7.44e-03 | 0.0387215 |
| M9192 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL DN | 352 | -1 | 7.75e-03 | 0.0398680 |
| M7847 REACTOME SIGNALING BY WNT | 244 | 1 | 8.02e-03 | 0.0410310 |
| M1362 SCHAEFFER PROSTATE DEVELOPMENT AND CANCER BOX5 UP | 9 | -1 | 8.51e-03 | 0.0429740 |
| M18086 DAIRKEE CANCER PRONE RESPONSE BPA E2 | 85 | 1 | 8.64e-03 | 0.0434785 |
| M39360 WP INTEGRATED BREAST CANCER PATHWAY | 112 | 1 | 8.67e-03 | 0.0435915 |
| M39417 WP ELECTRON TRANSPORT CHAIN OXPHOS SYSTEM IN MITOCHONDRIA | 75 | 1 | 8.90e-03 | 0.0445926 |
| M17372 LEE LIVER CANCER MYC TGFA DN | 32 | -1 | 9.54e-03 | 0.0469043 |
| M12596 SETLUR PROSTATE CANCER TMPRSS2 ERG FUSION DN | 10 | -1 | 9.92e-03 | 0.0484019 |
| M15193 GRUETZMANN PANCREATIC CANCER UP | 293 | 1 | 1.00e-02 | 0.0486279 |
| M9224 SAKAI CHRONIC HEPATITIS VS LIVER CANCER UP | 66 | 1 | 1.01e-02 | 0.0490729 |
| M19164 RICKMAN HEAD AND NECK CANCER D | 23 | -1 | 1.04e-02 | 0.0502348 |

Tabla B.19: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|---|---------|---------|----------|-----------|
| M9908 LANDIS BREAST CANCER PROGRESSION DN | 48 | -1 | 1.10e-02 | 0.0523021 |
| M12824 SHEDDEN LUNG CANCER GOOD SURVIVAL A4 | 140 | -1 | 1.14e-02 | 0.0535429 |
| M26973 REACTOME MITOCHONDRIAL BIOGENESIS | 66 | 1 | 1.21e-02 | 0.0558767 |
| M8817 ACEVEDO LIVER CANCER WITH H3K9ME3 DN | 83 | 1 | 1.23e-02 | 0.0565291 |
| M9914 YAMASHITA LIVER CANCER WITH EPCAM DN | 10 | -1 | 1.23e-02 | 0.0565291 |
| M19675 FARMER BREAST CANCER CLUSTER 3 | 11 | 1 | 1.25e-02 | 0.0571301 |
| M15411 BARIS THYROID CANCER DN | 43 | 1 | 1.29e-02 | 0.0584955 |
| M2202 JAZAERI BREAST CANCER BRCA1 VS BRCA2 UP | 35 | 1 | 1.36e-02 | 0.0606228 |
| M12461 FINAK BREAST CANCER SDPP SIGNATURE | 17 | -1 | 1.46e-02 | 0.0639456 |
| M17936 NIKOLSKY BREAST CANCER 8Q23 Q24 AMPLICON | 90 | 1 | 1.47e-02 | 0.0641404 |
| M19508 HATADA METHYLATED IN LUNG CANCER UP | 191 | -1 | 1.48e-02 | 0.0644793 |
| M9940 HONRADO BREAST CANCER BRCA1 VS BRCA2 | 13 | 1 | 1.49e-02 | 0.0647198 |
| M5989 IVANOV MUTATED IN COLON CANCER | 9 | 1 | 1.51e-02 | 0.0653918 |
| M816 BONOME OVARIAN CANCER POOR SURVIVAL UP | 23 | -1 | 1.58e-02 | 0.0675123 |
| M6046 BARRIER CANCER RELAPSE TUMOR SAMPLE UP | 12 | 1 | 1.64e-02 | 0.0692564 |
| M5369 POOLA INVASIVE BREAST CANCER UP | 202 | -1 | 1.72e-02 | 0.0721077 |
| M8689 CHIANG LIVER CANCER SUBCLASS CTNNB1 DN | 123 | -1 | 1.72e-02 | 0.0721077 |
| M8760 MCGARVEY SILENCED BY METHYLATION IN COLON CANCER | 22 | -1 | 1.76e-02 | 0.0733058 |
| M1359 SCHAEFFER PROSTATE DEVELOPMENT AND CANCER BOX4 DN | 30 | 1 | 1.81e-02 | 0.0744756 |
| M1177 KOINUMA COLON CANCER MSI DN | 8 | 1 | 1.89e-02 | 0.0768911 |
| M39892 WP MITOCHONDRIAL CIII ASSEMBLY | 11 | 1 | 1.90e-02 | 0.0771349 |
| M398 MINGUEZ LIVER CANCER VASCULAR INVASION DN | 13 | -1 | 1.91e-02 | 0.0773778 |
| M9206 YAMASHITA LIVER CANCER STEM CELL DN | 49 | -1 | 2.04e-02 | 0.0815012 |
| M9728 BARRIER COLON CANCER RECURRENCE UP | 32 | 1 | 2.06e-02 | 0.0822434 |
| M14181 YAMASHITA METHYLATED IN PROSTATE CANCER | 37 | -1 | 2.08e-02 | 0.0826990 |

Tabla B.19: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|---|---------|---------|----------|-----------|
| M13596 YANAGISAWA LUNG CANCER RECURRENCE | 8 | 1 | 2.15e-02 | 0.0850140 |
| M39731 WP LNCRNA INVOLVEMENT IN CANONICAL WNT SIGNALING AND COLORECTAL CANCER | 66 | 1 | 2.17e-02 | 0.0855122 |
| M5542 SATO SILENCED EPIGENETICALLY IN PANCREATIC CANCER | 28 | -1 | 2.22e-02 | 0.0870077 |
| M1065 MIDORIKAWA AMPLIFIED IN LIVER CANCER | 40 | 1 | 2.27e-02 | 0.0884274 |
| M10922 WANG RECURRENT LIVER CANCER UP | 17 | 1 | 2.27e-02 | 0.0884274 |
| M38985 MEBARKI HCC PROGENITOR WNT DN CTNNB1 DEPENDENT | 19 | -1 | 2.34e-02 | 0.0903018 |
| M4960 SMID BREAST CANCER BASAL DN | 440 | -1 | 2.50e-02 | 0.0945809 |
| M27163 REACTOME CONSTITUTIVE SIGNALING BY ABERRANT PI3K IN CANCER | 45 | -1 | 2.50e-02 | 0.0945809 |
| M5652 FARMER BREAST CANCER BASAL VS LUMINAL | 244 | 1 | 2.52e-02 | 0.0951506 |
| M8897 AGUIRRE PANCREATIC CANCER COPY NUMBER DN | 168 | 1 | 2.52e-02 | 0.0951506 |
| M39840 WP CANCER IMMUNOTHERAPY BY CTLA4 BLOCKADE | 9 | -1 | 2.57e-02 | 0.0965966 |
| M1293 YANG BREAST CANCER ESR1 DN | 18 | 1 | 2.66e-02 | 0.0992048 |
| M40043 WP MITOCHONDRIAL FATTY ACID SYNTHESIS PATHWAY | 4 | 1 | 2.66e-02 | 0.0992048 |

Tabla B.20: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA).

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M16740 GRADE COLON AND RECTAL CANCER UP | 267 | 238 | 0.00e+00 | 0.0000000 |
| M18855 VECCHI GASTRIC CANCER EARLY UP | 358 | 295 | 0.00e+00 | 0.0000000 |
| M15664 ROSTY CERVICAL CANCER PROLIFERATION CLUSTER | 123 | 115 | 0.00e+00 | 0.0000000 |
| M8520 SHEDDEN LUNG CANCER POOR SURVIVAL A6 | 346 | 260 | 0.00e+00 | 0.0000000 |
| M3766 SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP | 132 | 119 | 0.00e+00 | 0.0000000 |
| M7854 WHITEFORD PEDIATRIC CANCER MARKERS | 105 | 92 | 0.00e+00 | 0.0000000 |
| M5792 LI AMPLIFIED IN LUNG CANCER | 156 | 123 | 0.00e+00 | 0.0000000 |
| M854 LINDGREN BLADDER CANCER CLUSTER 3 UP | 259 | 184 | 0.00e+00 | 0.0000000 |

Tabla B.20: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M15150 ZHANG BREAST CANCER PROGENITORS UP | 352 | 235 | 0.00e+00 | 0.0000000 |
| M10986 CHIANG LIVER CANCER SUBCLASS UNANNOTATED DN | 163 | 124 | 0.00e+00 | 0.0000000 |
| M8365 RHODES UNDIFFERENTIATED CANCER | 62 | 57 | 0.00e+00 | 0.0000000 |
| M1345 RHODES CANCER META SIGNATURE | 58 | 54 | 0.00e+00 | 0.0000000 |
| M3268 CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP | 149 | 114 | 0.00e+00 | 0.0000000 |
| M7987 LEE LIVER CANCER SURVIVAL DN | 156 | 116 | 0.00e+00 | 0.0000000 |
| M18436 BOYVAULT LIVER CANCER SUBCLASS G3 UP | 153 | 114 | 0.00e+00 | 0.0000000 |
| M17423 LAIHO COLORECTAL CANCER SERRATED DN | 66 | 58 | 0.00e+00 | 0.0000000 |
| M39678 WP RETINOBLASTOMA GENE IN CANCER | 76 | 62 | 0.00e+00 | 0.0000000 |
| M13831 BOYVAULT LIVER CANCER SUBCLASS G123 UP | 38 | 35 | 0.00e+00 | 0.0000001 |
| M18973 NIKOLSKY BREAST CANCER 20Q12 Q13 AMPLICON | 86 | 66 | 0.00e+00 | 0.0000003 |
| M19982 WAMUNYOKOLI OVARIAN CANCER LMP UP | 217 | 142 | 0.00e+00 | 0.0000004 |
| M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT | 181 | 121 | 0.00e+00 | 0.0000007 |
| M17788 LINDGREN BLADDER CANCER CLUSTER 1 DN | 279 | 174 | 0.00e+00 | 0.0000015 |
| M15356 PUJANA BREAST CANCER LIT INT NETWORK | 80 | 61 | 0.00e+00 | 0.0000015 |
| M9826 VANTVEER BREAST CANCER METASTASIS DN | 82 | 62 | 1.00e-07 | 0.0000020 |
| M27446 REACTOME MITOCHONDRIAL TRANSLATION | 74 | 57 | 1.00e-07 | 0.0000022 |
| M17299 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL UP | 338 | 204 | 1.00e-07 | 0.0000040 |
| M12016 OUELLET OVARIAN CANCER INVASIVE VS LMP UP | 98 | 71 | 1.00e-07 | 0.0000042 |
| M17094 GINESTIER BREAST CANCER ZNF217 AMPLIFIED DN | 207 | 133 | 2.00e-07 | 0.0000044 |
| M336 VILLANUEVA LIVER CANCER KRT19 UP | 132 | 90 | 3.00e-07 | 0.0000090 |
| M17333 FARMER BREAST CANCER CLUSTER 2 | 31 | 28 | 4.00e-07 | 0.0000093 |
| M1687 ZHONG SECRETOME OF LUNG CANCER AND FIBROBLAST | 103 | 73 | 4.00e-07 | 0.0000109 |
| M2427 NAKAMURA CANCER MICROENVIRONMENT DN | 36 | 31 | 9.00e-07 | 0.0000220 |
| M16858 OSMAN BLADDER CANCER DN | 302 | 181 | 1.30e-06 | 0.0000320 |

Tabla B.20: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | nºGenes | nºGenes Significativos | P-valor | FDR |
|---|---------|------------------------|----------|-----------|
| M7995 HOSHIDA LIVER CANCER SUBCLASS S2 | 93 | 66 | 1.40e-06 | 0.0000328 |
| M7847 REACTOME SIGNALING BY WNT | 244 | 150 | 1.40e-06 | 0.0000329 |
| M15193 GRUETZMANN PANCREATIC CANCER UP | 293 | 174 | 4.70e-06 | 0.0001002 |
| M4572 LOCKWOOD AMPLIFIED IN LUNG CANCER | 143 | 93 | 5.30e-06 | 0.0001097 |
| M2435 WAMUNYOKOLI OVARIAN CANCER GRADES 1 2 UP | 127 | 84 | 5.50e-06 | 0.0001149 |
| M39740 WP METABOLIC REPROGRAMMING IN COLON CANCER | 33 | 28 | 5.70e-06 | 0.0001172 |
| M15112 WONG MITOCHONDRIA GENE MODULE | 177 | 111 | 8.70e-06 | 0.0001721 |
| M1685 ZHONG SECRETOME OF LUNG CANCER AND MACROPHAGE | 63 | 46 | 1.65e-05 | 0.0003045 |
| M9577 MOOTHA MITOCHONDRIA | 321 | 186 | 1.89e-05 | 0.0003477 |
| M3710 WALLACE PROSTATE CANCER UP | 14 | 14 | 2.17e-05 | 0.0003905 |
| M1260 WATANABE RECTAL CANCER RADIOTHERAPY RESPONSIVE UP | 76 | 53 | 3.29e-05 | 0.0005673 |
| M6862 NADERI BREAST CANCER PROGNOSIS UP | 38 | 30 | 4.23e-05 | 0.0007041 |
| M12795 CHARAFE BREAST CANCER BASAL VS MESENCHYMAL UP | 93 | 62 | 6.39e-05 | 0.0010252 |
| M13736 PYEON CANCER HEAD AND NECK VS CERVICAL UP | 152 | 94 | 8.99e-05 | 0.0013959 |
| M19148 TOMLINS PROSTATE CANCER UP | 36 | 28 | 1.24e-04 | 0.0018225 |
| M39628 WP INTEGRATED CANCER PATHWAY | 36 | 28 | 1.24e-04 | 0.0018225 |
| M16374 IIZUKA LIVER CANCER PROGRESSION G1 G2 DN | 22 | 19 | 1.26e-04 | 0.0018425 |
| M16542 YAMASHITA LIVER CANCER WITH EPCAM UP | 41 | 31 | 1.37e-04 | 0.0019832 |
| M4448 BOYALT LIVER CANCER SUBCLASS G23 UP | 41 | 31 | 1.37e-04 | 0.0019832 |
| M13072 SMID BREAST CANCER LUMINAL A DN | 15 | 14 | 1.84e-04 | 0.0025548 |
| M9565 AGUIRRE PANCREATIC CANCER COPY NUMBER UP | 207 | 122 | 1.87e-04 | 0.0025671 |
| M18086 DAIRKEE CANCER PRONE RESPONSE BPA E2 | 85 | 56 | 2.30e-04 | 0.0030540 |
| M19097 SWEET LUNG CANCER KRAS UP | 346 | 193 | 2.63e-04 | 0.0034214 |
| M590 REACTOME MITOCHONDRIAL PROTEIN IMPORT | 47 | 34 | 2.83e-04 | 0.0036487 |
| M8655 DAIRKEE CANCER PRONE RESPONSE BPA | 34 | 26 | 3.53e-04 | 0.0044615 |

Tabla B.20: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|---|---------|------------------------|----------|-----------|
| M747 REACTOME METABOLISM OF POLYAMINES | 51 | 36 | 4.14e-04 | 0.0050986 |
| M14146 BOYVAULT LIVER CANCER SUBCLASS G1 UP | 92 | 59 | 4.65e-04 | 0.0055966 |
| M1686 ZHONG SECRETOME OF LUNG CANCER AND ENDOTHELIUM | 53 | 37 | 4.89e-04 | 0.0058117 |
| M8779 LI LUNG CANCER | 28 | 22 | 5.30e-04 | 0.0061844 |
| M3102 SMID BREAST CANCER ERBB2 UP | 100 | 63 | 6.15e-04 | 0.0070481 |
| M39347 WP GASTRIC CANCER NETWORK 1 | 25 | 20 | 6.32e-04 | 0.0072286 |
| M5369 POOLA INVASIVE BREAST CANCER UP | 202 | 117 | 6.44e-04 | 0.0073223 |
| M2180 PUJANA BREAST CANCER WITH BRCA1 MUTATED UP | 45 | 32 | 7.02e-04 | 0.0079040 |
| M15147 BOYVAULT LIVER CANCER SUBCLASS G12 UP | 30 | 23 | 7.36e-04 | 0.0081910 |
| M416 ANDERSEN LIVER CANCER KRT19 UP | 30 | 23 | 7.36e-04 | 0.0081910 |
| M2082 WANG METASTASIS OF BREAST CANCER ESR1 UP | 16 | 14 | 8.38e-04 | 0.0091154 |
| M15425 VANTVEER BREAST CANCER ESR1 DN | 161 | 95 | 8.72e-04 | 0.0093968 |
| M11629 MONTERO THYROID CANCER POOR SURVIVAL UP | 9 | 9 | 1.01e-03 | 0.0105493 |
| M19988 GINESTIER BREAST CANCER 20Q13 AMPLIFICATION DN | 145 | 86 | 1.20e-03 | 0.0122028 |
| M1273 SHETH LIVER CANCER VS TXNIP LOSS PAM1 | 182 | 105 | 1.43e-03 | 0.0142163 |
| M867 MEINHOLD OVARIAN CANCER LOW GRADE DN | 18 | 15 | 1.49e-03 | 0.0147874 |
| M26902 REACTOME ACTIVATION OF BAD AND TRANSLOCATION TO MITOCHONDRIA | 12 | 11 | 1.50e-03 | 0.0148356 |
| M14670 DING LUNG CANCER EXPRESSION BY COPY NUMBER | 84 | 53 | 1.54e-03 | 0.0150254 |
| M39497 WP GASTRIC CANCER NETWORK 2 | 26 | 20 | 1.54e-03 | 0.0150254 |
| M9389 FINETTI BREAST CANCER KINOME RED | 15 | 13 | 1.59e-03 | 0.0154610 |
| M39360 WP INTEGRATED BREAST CANCER PATHWAY | 112 | 68 | 1.63e-03 | 0.0157703 |
| M26973 REACTOME MITOCHONDRIAL BIOGENESIS | 66 | 43 | 1.67e-03 | 0.0161303 |
| M836 OUELLET CULTURED OVARIAN CANCER INVASIVE VS LMP UP | 50 | 34 | 1.70e-03 | 0.0162839 |
| M8411 LINDGREN BLADDER CANCER WITH LOH IN CHR9Q | 88 | 55 | 1.74e-03 | 0.0166395 |
| M10837 NIKOLSKY BREAST CANCER 20Q11 AMPLICON | 23 | 18 | 1.90e-03 | 0.0180500 |

Tabla B.20: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | n ^o Genes | n ^o Genes Significativos | P-valor | FDR |
|---|----------------------|-------------------------------------|----------|-----------|
| M13596 YANAGISAWA LUNG CANCER RECURRENCE | 8 | 8 | 2.17e-03 | 0.0202487 |
| M5275 OSMAN BLADDER CANCER UP | 310 | 169 | 2.41e-03 | 0.0220955 |
| M4619 LIU PROSTATE CANCER UP | 78 | 49 | 2.60e-03 | 0.0235386 |
| M2837 BARIS THYROID CANCER UP | 17 | 14 | 2.71e-03 | 0.0243436 |
| M1096 KOINUMA COLON CANCER MSI UP | 11 | 10 | 2.97e-03 | 0.0261107 |
| M378 MINGUEZ LIVER CANCER VASCULAR INVASION UP | 11 | 10 | 2.97e-03 | 0.0261107 |
| M5652 FARMER BREAST CANCER BASAL VS LULMINAL | 244 | 135 | 3.11e-03 | 0.0270533 |
| M9728 BARRIER COLON CANCER RECURRENCE UP | 32 | 23 | 3.19e-03 | 0.0276243 |
| M7678 CHIANG LIVER CANCER SUBCLASS POLYSOMY7 DN | 22 | 17 | 3.26e-03 | 0.0281461 |
| M8817 ACEVEDO LIVER CANCER WITH H3K9ME3 DN | 83 | 51 | 4.20e-03 | 0.0347553 |
| M15411 BARIS THYROID CANCER DN | 43 | 29 | 4.42e-03 | 0.0362432 |
| M1758 SANSOM WNT PATHWAY REQUIRE MYC | 52 | 34 | 4.56e-03 | 0.0369405 |
| M7880 SPIRA SMOKERS LUNG CANCER UP | 31 | 22 | 5.03e-03 | 0.0397446 |
| M9940 HONRADO BREAST CANCER BRCA1 VS BRCA2 | 13 | 11 | 5.61e-03 | 0.0434361 |
| M1065 MIDORIKAWA AMPLIFIED IN LIVER CANCER | 40 | 27 | 5.85e-03 | 0.0449500 |
| M27288 REACTOME BETA CATENIN INDEPENDENT WNT SIGNALING | 105 | 62 | 6.27e-03 | 0.0475269 |
| M2631 FARMER BREAST CANCER APOCRINE VS BASAL | 244 | 133 | 6.62e-03 | 0.0499281 |
| M3645 DELYS THYROID CANCER UP | 320 | 171 | 6.66e-03 | 0.0500888 |
| M15096 REACTOME MITOCHONDRIAL TRNA AMINOACYLATION | 18 | 14 | 6.96e-03 | 0.0517387 |
| M39731 WP LNCRNA INVOLVEMENT IN CANONICAL WNT SIGNALING AND COLORECTAL CANCER | 66 | 41 | 7.46e-03 | 0.0546814 |
| M7547 JIANG HYPOXIA CANCER | 46 | 30 | 7.95e-03 | 0.0576874 |
| M18761 SMITH LIVER CANCER | 39 | 26 | 8.69e-03 | 0.0620991 |
| M8526 CHEN LUNG CANCER SURVIVAL | 20 | 15 | 9.20e-03 | 0.0649623 |
| M13007 GLINSKY CANCER DEATH UP | 6 | 6 | 1.01e-02 | 0.0702017 |
| M38998 REACTOME ABERRANT REGULATION OF MITOTIC G1 S TRANSITION IN CANCER DUE TO RB1 DEFECTS | 12 | 10 | 1.03e-02 | 0.0709092 |
| M4290 LUI THYROID CANCER CLUSTER 3 | 22 | 16 | 1.15e-02 | 0.0773271 |
| M5989 IVANOV MUTATED IN COLON CANCER | 9 | 8 | 1.15e-02 | 0.0773271 |

Tabla B.20: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M523 KEGG THYROID CANCER | 17 | 13 | 1.19e-02 | 0.0793700 |
| M14383 SMID BREAST CANCER RELAPSE IN BRAIN UP | 24 | 17 | 1.38e-02 | 0.0894599 |
| M9150 GRESHOCK CANCER COPY NUMBER UP | 221 | 119 | 1.58e-02 | 0.0994126 |

Tabla B.21: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MH (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|---------|-----------|
| M5895 HALLMARK WNT BETA CATENIN SIGNALING | 26 | 1 | 0.0421 | 0.0956818 |

B.3. Análisis transcriptómico de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las líneas HT29 y NCM460

En los siguientes subapartados de este apéndice se presentan las tablas y figuras suplementarias correspondientes al apartado “7.1.3.2 Análisis transcriptómico de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO”

B.3.1. Control de calidad y preprocesado de los datos transcriptómicos de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO

Aquí se presenta la selección del mejor modelo (tabla B.22), que resultó en la selección del modelo completo; el control de calidad y detección de *outliers* mediante métodos generales (figura B.17) y específicos de microarrays (B.18), a partir de los cuales no se detectó ninguna muestra atípica, y el preprocesado de los datos mediante el método RMA para quitar el ruido de fondo, normalizar los microarrays y resumir las intensidades de las diferentes sondas pertenecientes a un mismo *probeset* o conjunto de sondas (correspondientes a un mismo gen) a un solo valor de expresión (figura B.19). Por último, se realizan los análisis de expresión diferencial, búsqueda de patrones y enriquecimiento de conjuntos de genes, mostrado en el apartado “7.1.3.2 Análisis transcriptómico de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO” correspondiente.

Tabla B.22: Porcentaje del mejor ajuste de todos los modelos evaluados.

| Modelo | n° Sondas (%) | | Criterio mínimo | |
|--|---------------|--------------|-----------------|--------|
| | AIC | BIC | AIC | BIC |
| Línea + (1 Día) | 3945 (30.81) | 4523 (30.86) | -22.26 | -20.80 |
| Tratamiento + (1 Día) | 1528 (11.78) | 1747 (13.47) | -16.61 | -15.15 |
| Línea + Tratamiento + (1 Día) | 2252 (17.36) | 2216 (17.08) | -23.72 | -21.78 |
| Línea + Tratamiento + Línea:Tratamiento + (1 Día) | 5248 (40.45) | 4487 (34.59) | -24.54 | -22.11 |

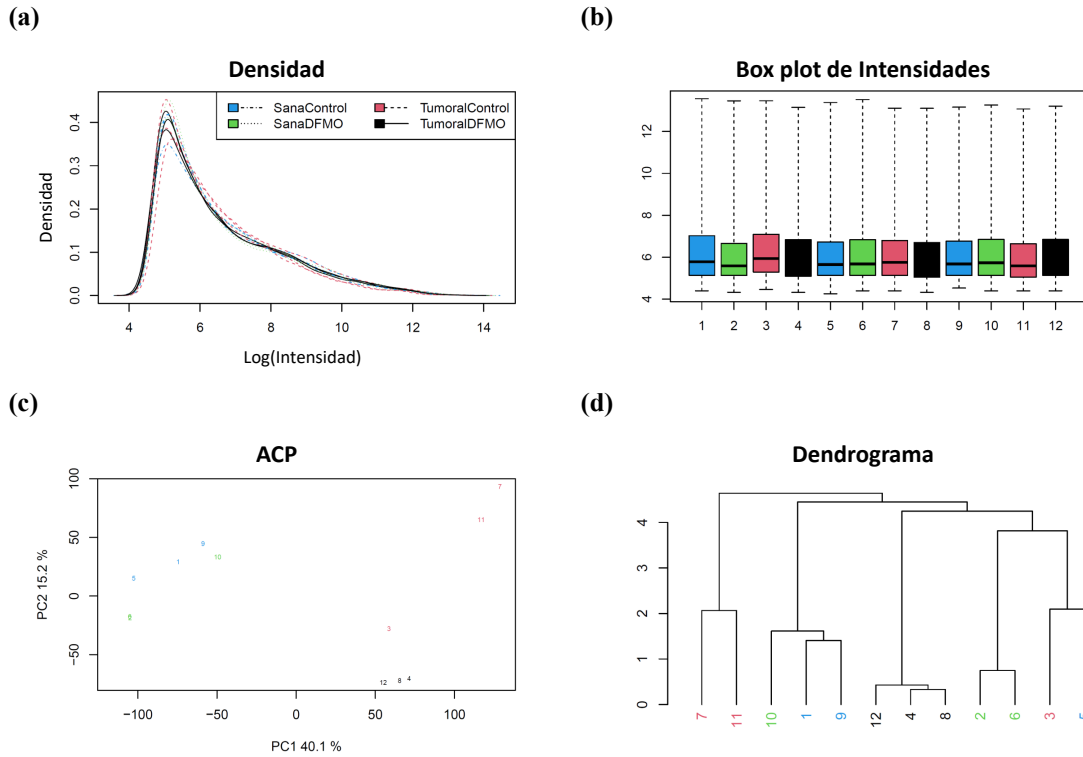


Figura B.17: Control de calidad y detección de *outliers* con métodos generales. b) Gráfico de densidad; b) Gráfico de densidad; (c) Análisis de componentes principales coloreada en función de los factores de interés *Línea* y *Tratamiento*; Análisis Clúster Jerárquico, se ha usado el método de Ward y como medida de disimilaridad se ha usado la distancia euclídea. Las muestras se han coloreado en función del factor *Tumor* (negro para muestras tumorales y rojo para normales).

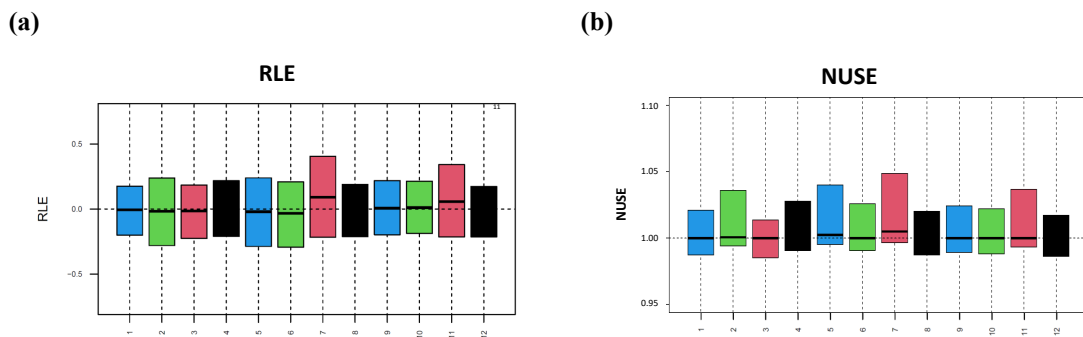


Figura B.18: Control de calidad y detección de *outliers* con métodos específicos para microarrays. a) Expresión logarítmica relativa (RLE); b) Error estándar normalizado sin escalar (NUSE). Distribuciones muy diferentes a la mayoría indicarían algún problema en la muestra.

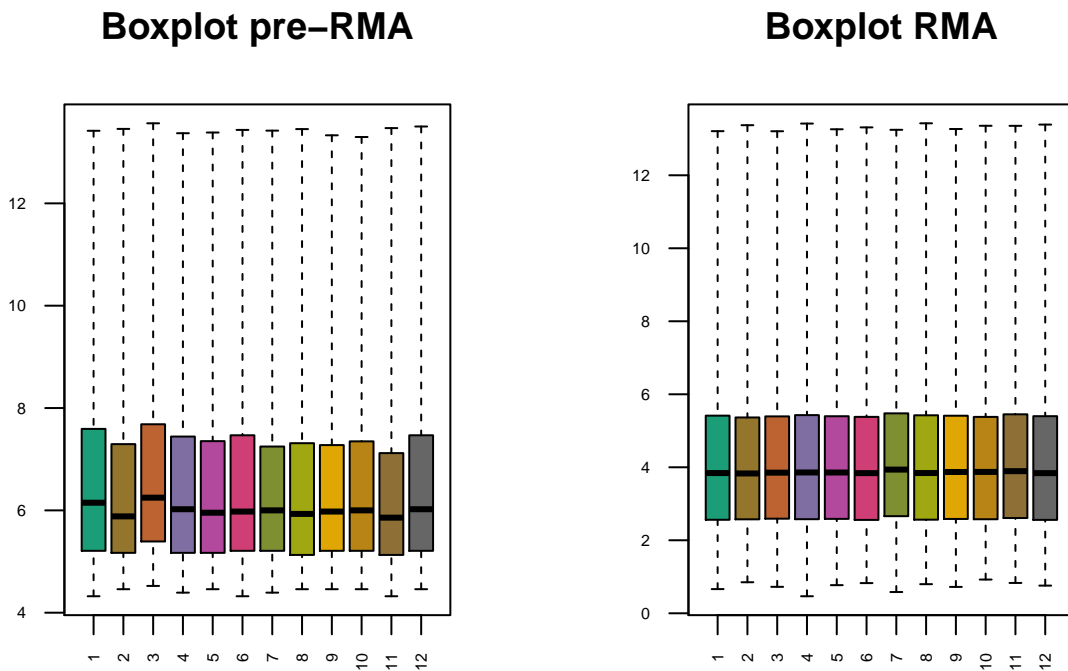


Figura B.19: Corrección del ruido de fondo, normalización y resumen mediante el método RMA. Izquierda) Expresión logarítmica previa al pre-tratamiento; Derecha) Expresión logarítmica posterior al pre-tratamiento.

B.3.2. Análisis de expresión diferencial de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO

A continuación se muestran los valores de expresión diferencial de los genes implicados en los sistemas de transporte de Ca^{2+} intracelular dividido en las diferentes familias en las que se agrupan.

**Tabla B.23: DEG de la familia VOCC
(NCM460 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|---------------|---------|--------|-------|----------|
| Cav1.1 | 0.111 | -0.481 | 0.703 | 9.37E-01 |
| Cav1.2 | 0.054 | -0.475 | 0.583 | 9.74E-01 |
| Cav1.3 | 0.324 | -0.288 | 0.936 | 7.25E-01 |
| Cav1.4 | -0.172 | -0.617 | 0.274 | 8.26E-01 |
| Cav2.1 | -0.202 | -0.582 | 0.178 | 7.25E-01 |
| Cav2.2 | 0.121 | -0.539 | 0.780 | 9.38E-01 |
| Cav2.3 | 0.106 | -0.381 | 0.592 | 9.27E-01 |
| Cav3.1 | 0.126 | -0.418 | 0.671 | 9.19E-01 |
| Cav3.2 | -0.095 | -0.538 | 0.349 | 9.28E-01 |
| Cav3.3 | -0.277 | -0.766 | 0.212 | 6.98E-01 |

**Tabla B.24: DEG de la familia VOCC
(HT29 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|---------------|---------|--------|-------|----------|
| Cav1.1 | 0.180 | -0.412 | 0.772 | 6.74E-01 |
| Cav1.2 | 0.085 | -0.444 | 0.614 | 8.32E-01 |
| Cav1.3 | 0.124 | -0.488 | 0.736 | 7.85E-01 |
| Cav1.4 | 0.114 | -0.332 | 0.559 | 7.28E-01 |
| Cav2.1 | -0.322 | -0.702 | 0.058 | 1.97E-01 |
| Cav2.2 | 0.132 | -0.528 | 0.791 | 7.88E-01 |
| Cav2.3 | 0.199 | -0.287 | 0.686 | 5.58E-01 |
| Cav3.1 | -0.026 | -0.571 | 0.519 | 9.53E-01 |
| Cav3.2 | -0.382 | -0.826 | 0.061 | 1.88E-01 |
| Cav3.3 | -0.155 | -0.644 | 0.334 | 6.59E-01 |

**Tabla B.25: DEG de la familia SOCE
(NCM460 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|----------------|---------|--------|-------|----------|
| ORAI1 | -0.032 | -0.405 | 0.342 | 9.80E-01 |
| ORAI2 | 0.177 | -0.206 | 0.561 | 7.75E-01 |
| ORAI3 | 0.195 | -0.133 | 0.523 | 6.78E-01 |
| STIM1 | -0.206 | -0.549 | 0.138 | 6.75E-01 |
| STIM2 | -0.149 | -0.610 | 0.311 | 8.67E-01 |
| MS4A12 | 0.022 | -0.364 | 0.408 | 9.86E-01 |
| CRACR2A | 0.198 | -0.254 | 0.649 | 7.95E-01 |

**Tabla B.26: DEG de la familia SOCE
(HT29 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|----------------|---------|--------|-------|----------|
| ORAI1 | 0.313 | -0.060 | 0.687 | 2.02E-01 |
| ORAI2 | 0.315 | -0.069 | 0.698 | 2.11E-01 |
| ORAI3 | -0.142 | -0.469 | 0.186 | 5.37E-01 |
| STIM1 | 0.451 | 0.107 | 0.794 | 5.04E-02 |
| STIM2 | 0.841 | 0.380 | 1.302 | 1.20E-02 |
| MS4A12 | 0.031 | -0.355 | 0.417 | 9.19E-01 |
| CRACR2A | 0.570 | 0.119 | 1.021 | 5.83E-02 |

**Tabla B.27: DEG de la familia CLC
(NCM460 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|--------------|---------|--------|-------|----------|
| IP3R1 | 0.790 | 0.360 | 1.219 | 9.85E-02 |
| IP3R2 | -0.078 | -0.747 | 0.591 | 9.69E-01 |
| IP3R3 | -0.368 | -0.878 | 0.141 | 5.77E-01 |
| RYR1 | -0.196 | -0.659 | 0.267 | 8.02E-01 |
| RYR2 | 0.630 | -0.155 | 1.416 | 5.24E-01 |
| RYR3 | 0.272 | -0.151 | 0.694 | 6.42E-01 |

**Tabla B.28: DEG de la familia CLC
(HT29 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|--------------|---------|--------|-------|----------|
| IP3R1 | 0.434 | 0.005 | 0.863 | 1.24E-01 |
| IP3R2 | -0.114 | -0.783 | 0.555 | 8.21E-01 |
| IP3R3 | -0.313 | -0.822 | 0.197 | 3.61E-01 |
| RYR1 | 0.037 | -0.426 | 0.500 | 9.20E-01 |
| RYR2 | -0.050 | -0.836 | 0.735 | 9.37E-01 |
| RYR3 | 0.039 | -0.384 | 0.461 | 9.08E-01 |

**Tabla B.29: DEG de la familia CP
(NCM460 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|--------|---------|--------|-------|----------|
| PMCA1 | 0.135 | -0.345 | 0.615 | 8.95E-01 |
| PMCA2 | 0.069 | -0.426 | 0.564 | 9.57E-01 |
| PMCA3 | 0.222 | -0.201 | 0.644 | 7.28E-01 |
| PMCA4 | -0.072 | -0.503 | 0.358 | 9.46E-01 |
| SERCA1 | 0.187 | -0.412 | 0.786 | 8.74E-01 |
| SERCA2 | 0.068 | -0.275 | 0.411 | 9.33E-01 |
| SERCA3 | -0.122 | -0.563 | 0.320 | 8.99E-01 |
| NCX1 | 0.219 | -0.224 | 0.661 | 7.51E-01 |
| NCX2 | -0.309 | -0.688 | 0.070 | 5.14E-01 |
| NCX3 | 0.172 | -0.218 | 0.562 | 7.92E-01 |

**Tabla B.30: DEG de la familia CP
(HT29 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|--------|---------|--------|-------|----------|
| PMCA1 | 0.158 | -0.321 | 0.638 | 6.44E-01 |
| PMCA2 | -0.217 | -0.713 | 0.278 | 5.29E-01 |
| PMCA3 | -0.071 | -0.493 | 0.352 | 8.25E-01 |
| PMCA4 | 0.920 | 0.489 | 1.350 | 5.65E-03 |
| SERCA1 | -0.108 | -0.707 | 0.490 | 8.09E-01 |
| SERCA2 | 0.085 | -0.258 | 0.428 | 7.36E-01 |
| SERCA3 | 0.135 | -0.306 | 0.576 | 6.71E-01 |
| NCX1 | 0.298 | -0.144 | 0.741 | 3.12E-01 |
| NCX2 | -0.066 | -0.445 | 0.313 | 8.16E-01 |
| NCX3 | -0.139 | -0.529 | 0.251 | 6.16E-01 |

**Tabla B.31: DEG de la familia MCT
(NCM460 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|-------|---------|--------|-------|----------|
| MCU | -0.006 | -0.400 | 0.388 | 9.94E-01 |
| MICU1 | -0.092 | -0.555 | 0.371 | 9.33E-01 |
| MICU2 | -0.288 | -0.899 | 0.322 | 7.68E-01 |
| MICU3 | 0.176 | -0.219 | 0.572 | 7.87E-01 |
| MCUR1 | 0.050 | -0.399 | 0.498 | 9.71E-01 |
| EMRE | 0.351 | -0.017 | 0.719 | 4.30E-01 |
| MCUb | -0.378 | -1.785 | 1.028 | 9.04E-01 |
| VDAC1 | 0.032 | -0.375 | 0.439 | 9.82E-01 |
| VDAC2 | -0.048 | -0.397 | 0.302 | 9.60E-01 |
| VDAC3 | 0.081 | -0.678 | 0.841 | 9.72E-01 |
| NCLX | -0.112 | -0.441 | 0.216 | 8.53E-01 |

**Tabla B.32: DEG de la familia MCT
(HT29 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|-------|---------|--------|-------|----------|
| MCU | 0.589 | 0.195 | 0.983 | 2.94E-02 |
| MICU1 | 0.481 | 0.019 | 0.944 | 1.14E-01 |
| MICU2 | 0.577 | -0.033 | 1.188 | 1.50E-01 |
| MICU3 | 0.202 | -0.193 | 0.597 | 4.55E-01 |
| MCUR1 | 0.056 | -0.393 | 0.505 | 8.73E-01 |
| EMRE | 0.393 | 0.025 | 0.761 | 1.05E-01 |
| MCUb | 1.345 | -0.062 | 2.751 | 1.45E-01 |
| VDAC1 | 0.420 | 0.013 | 0.827 | 1.16E-01 |
| VDAC2 | -0.140 | -0.489 | 0.209 | 5.67E-01 |
| VDAC3 | 1.022 | 0.262 | 1.781 | 4.55E-02 |
| NCLX | 0.124 | -0.204 | 0.452 | 5.92E-01 |

**Tabla B.33: DEG de la familia TRP
(NCM460 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|--------|---------|--------|-------|----------|
| TRPC1 | -0.028 | -0.640 | 0.584 | 9.88E-01 |
| TRPC3 | -0.063 | -0.455 | 0.329 | 9.47E-01 |
| TRPC4 | 0.372 | -0.112 | 0.855 | 5.50E-01 |
| TRPC5 | 0.279 | -0.167 | 0.725 | 6.57E-01 |
| TRPC6 | -0.099 | -0.607 | 0.410 | 9.34E-01 |
| TRPC7 | -0.071 | -0.380 | 0.237 | 9.20E-01 |
| TRPM1 | -0.140 | -0.664 | 0.384 | 9.04E-01 |
| TRPM2 | 0.150 | -0.194 | 0.494 | 7.96E-01 |
| TRPM3 | -0.180 | -0.720 | 0.360 | 8.62E-01 |
| TRPM5 | -0.262 | -0.736 | 0.211 | 7.08E-01 |
| TRPM6 | -0.099 | -0.609 | 0.411 | 9.34E-01 |
| TRPM7 | -0.250 | -0.816 | 0.315 | 7.91E-01 |
| TRPM8 | -0.142 | -0.583 | 0.300 | 8.70E-01 |
| TRPML1 | 0.076 | -0.341 | 0.494 | 9.38E-01 |
| TRPML2 | 0.136 | -0.319 | 0.591 | 8.81E-01 |
| TRPML3 | -0.375 | -1.013 | 0.264 | 6.83E-01 |
| TRPV1 | -0.231 | -0.684 | 0.222 | 7.40E-01 |
| TRPV2 | -0.061 | -0.485 | 0.362 | 9.55E-01 |
| TRPV3 | 0.300 | -0.237 | 0.837 | 7.05E-01 |
| TRPV4 | -0.056 | -0.547 | 0.436 | 9.70E-01 |
| TRPV5 | -0.055 | -0.495 | 0.385 | 9.65E-01 |
| TRPV6 | -0.292 | -1.130 | 0.545 | 8.50E-01 |
| TRPA1 | -0.155 | -0.636 | 0.327 | 8.69E-01 |
| TRPP1 | -0.257 | -0.713 | 0.200 | 7.03E-01 |
| TRPP2 | -0.060 | -0.827 | 0.707 | 9.82E-01 |
| TRPP3 | -0.171 | -0.701 | 0.358 | 8.67E-01 |
| TRPP5 | -0.288 | -0.862 | 0.286 | 7.44E-01 |

**Tabla B.34: DEG de la familia TRP
(HT29 DFMO vs. Control).**

| GEN | log(FC) | LI | LS | FDR |
|--------|---------|--------|--------|----------|
| TRPC1 | -0.827 | -1.439 | -0.215 | 4.46E-02 |
| TRPC3 | 0.136 | -0.256 | 0.528 | 6.25E-01 |
| TRPC4 | 0.012 | -0.471 | 0.496 | 9.78E-01 |
| TRPC5 | -0.653 | -1.099 | -0.207 | 3.23E-02 |
| TRPC6 | 0.070 | -0.439 | 0.578 | 8.58E-01 |
| TRPC7 | -0.243 | -0.552 | 0.065 | 2.30E-01 |
| TRPM1 | -0.257 | -0.781 | 0.267 | 4.74E-01 |
| TRPM2 | 0.218 | -0.126 | 0.561 | 3.45E-01 |
| TRPM3 | -0.332 | -0.873 | 0.208 | 3.59E-01 |
| TRPM5 | -0.213 | -0.686 | 0.261 | 5.19E-01 |
| TRPM6 | -0.287 | -0.797 | 0.223 | 4.04E-01 |
| TRPM7 | -0.221 | -0.786 | 0.345 | 5.79E-01 |
| TRPM8 | 0.197 | -0.245 | 0.638 | 5.22E-01 |
| TRPML1 | 0.231 | -0.187 | 0.648 | 4.14E-01 |
| TRPML2 | 0.134 | -0.321 | 0.589 | 6.84E-01 |
| TRPML3 | 0.224 | -0.414 | 0.862 | 6.22E-01 |
| TRPV1 | 0.322 | -0.131 | 0.775 | 2.83E-01 |
| TRPV2 | -0.238 | -0.661 | 0.186 | 4.06E-01 |
| TRPV3 | -0.046 | -0.583 | 0.491 | 9.14E-01 |
| TRPV4 | -0.153 | -0.645 | 0.339 | 6.65E-01 |
| TRPV5 | -0.298 | -0.738 | 0.142 | 3.09E-01 |
| TRPV6 | -2.567 | -3.405 | -1.729 | 8.46E-04 |
| TRPA1 | -0.127 | -0.608 | 0.355 | 7.18E-01 |
| TRPP1 | -1.032 | -1.489 | -0.575 | 4.26E-03 |
| TRPP2 | 0.868 | 0.101 | 1.635 | 8.64E-02 |
| TRPP3 | -0.026 | -0.555 | 0.503 | 9.51E-01 |
| TRPP5 | -0.087 | -0.661 | 0.487 | 8.42E-01 |

B.3.3. Análisis de enriquecimiento de genes de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO

En este apéndice se muestra el análisis de enriquecimiento de conjuntos de genes correspondiente al análisis transcriptómico de los experimentos de inhibición de la ruta de síntesis de poliaminas en dos líneas celulares: NCM460 y HT29. En concreto, se van a mostrar, divididos en diferentes tablas según la base de datos a la que pertenezcan y el método de análisis que se haya empleado, aquellos conjuntos con un FDR < 0.1 que contienen las palabras de interés: *calcium*, *mitochondria*, *reticulum*, *cancer*, *wnt* y *polyamine*. En el primer subapartado se mostrarán las correspondientes al efecto del DFMO en las células NCM460 y en segundo, en las HT29.

B.3.3.1. Análisis de enriquecimiento de genes de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las células NCM460

En la tabla B.35 se muestra un resumen numérico del análisis de enriquecimiento realizado a partir del análisis de expresión diferencial entre las células NCM460 tratadas con DFMO y sin tratar. Para ello, se han cruzado estos datos con 8 colecciones de conjuntos de genes que han sido empeladas a lo largo de la presente tesis, de tal manera que en dicha tabla se muestra no sólo cuántos conjuntos de genes rechazan alguna de las dos hipótesis nulas evaluadas (autocontenida y competitiva), sino también cuántos de aquellos que rechazan alguna de dichas hipótesis nulas pertenecen a las bases de datos relacionadas con el cáncer (MCC y MOG) y/o contienen en su nombre alguna de las siguientes palabras clave de interés para la la temática de la tesis: *calcium*, *mitochondria*, *reticulum*, *cancer*, *wnt* y *polyamines*. Posteriormente, en las tablas B.36-B.37 se muestra en detalle el análisis de enriquecimiento de conjuntos de genes para aquellos conjuntos con un FDR < 0.1 que contienen alguna de las palabras de interés divididos en diferentes tablas en función tanto de la base de datos a la que pertenecen como del método empleado para su análisis.

Tabla B.35: Análisis de enriquecimiento de genes para el tratamiento con DFMO en NCM460.

| Base de Datos | Conjuntos | | Conjuntos Significativos (FDR <0.05) | | | | | | | |
|---------------|--------------|---------------------|--------------------------------------|------------|----------------------|---------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | Métodos | Totales | Calcium ¹ | Micochondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA ORA | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO BP | 12342 | 146 | CAMERA ORA | 19 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO MF | 4245 | 66 | CAMERA ORA | 7 | 0 | 0 | 0 | 0 | 0 | 0 |
| MH | 50 | 0 | CAMERA ORA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| MCDB | 6290 | 12 | CAMERA ORA | 222 | 0 | 0 | 0 | 19 | 0 | 0 |
| MCC | 858 | 0 | CAMERA ORA | 61 | 0 | 0 | 0 | 0 | 0 | 0 |
| MOG | 189 | 0 | CAMERA ORA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| MIM | 5219 | 0 | CAMERA ORA | 48 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 29537 | 225 | CAMERA ORA | 380 | 0 | 0 | 0 | 19 | 0 | 0 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology definidas por nuestro grupo.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

Tabla B.36: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA).

| Conjuntos | nº Genes | DIR | P-valor | FDR |
|---|----------|-----|----------|-----------|
| GO:0006596 polyamine biosynthetic process | 7 | -1 | 0.000601 | 0.0825664 |

Tabla B.37: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA).

| Conjuntos | nº Genes | DIR | P-valor | FDR |
|---|----------|-----|----------|-----------|
| M15664 ROSTY CERVICAL CANCER PROLIFERATION CLUSTER | 121 | -1 | 0.00e+00 | 0.0000000 |
| M7854 WHITEFORD PEDIATRIC CANCER MARKERS | 91 | -1 | 0.00e+00 | 0.0000054 |
| M3766 SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP | 133 | -1 | 0.00e+00 | 0.0000102 |
| M17333 FARMER BREAST CANCER CLUSTER 2 | 31 | -1 | 0.00e+00 | 0.0000131 |
| M2180 PUJANA BREAST CANCER WITH BRCA1 MUTATED UP | 43 | -1 | 1.50e-06 | 0.0001876 |
| M8520 SHEDDEN LUNG CANCER POOR SURVIVAL A6 | 332 | -1 | 2.10e-06 | 0.0002352 |
| M39678 WP RETINOBLASTOMA GENE IN CANCER | 74 | -1 | 4.30e-06 | 0.0004247 |
| M336 VILLANUEVA LIVER CANCER KRT19 UP | 140 | -1 | 1.15e-05 | 0.0009958 |
| M2427 NAKAMURA CANCER MICROENVIRONMENT DN | 36 | -1 | 3.61e-05 | 0.0023990 |
| M11629 MONTERO THYROID CANCER POOR SURVIVAL UP | 10 | -1 | 4.88e-05 | 0.0029669 |
| M13072 SMID BREAST CANCER LUMINAL A DN | 15 | -1 | 7.63e-05 | 0.0043513 |
| M8365 RHODES UNDIFFERENTIATED CANCER | 48 | -1 | 7.69e-05 | 0.0043513 |
| M6862 NADERI BREAST CANCER PROGNOSIS UP | 33 | -1 | 9.29e-05 | 0.0051050 |
| M854 LINDGREN BLADDER CANCER CLUSTER 3 UP | 240 | -1 | 1.48e-04 | 0.0077598 |
| M17788 LINDGREN BLADDER CANCER CLUSTER 1 DN | 250 | -1 | 3.49e-04 | 0.0149965 |
| M3268 CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP | 143 | -1 | 4.40e-04 | 0.0173421 |
| M5369 POOLA INVASIVE BREAST CANCER UP | 165 | -1 | 5.69e-04 | 0.0207123 |
| M16586 LAIHO COLORECTAL CANCER SERRATED UP | 76 | 1 | 1.12e-03 | 0.0331648 |
| M39628 WP INTEGRATED CANCER PATHWAY | 37 | -1 | 1.46e-03 | 0.0413064 |
| M27100 REACTOME WNT MEDIATED ACTIVATION OF DVL | 6 | -1 | 2.14e-03 | 0.0538771 |
| M39740 WP METABOLIC REPROGRAMMING IN COLON CANCER | 28 | -1 | 3.08e-03 | 0.0682256 |
| M8629 NADERI BREAST CANCER PROGNOSIS DN | 13 | 1 | 3.64e-03 | 0.0762000 |

Tabla B.37: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA).
(continuación)

| Conjuntos | n° Genes | DIR | P-valor | FDR |
|---|-----------------|------------|----------------|------------|
| M4448 BOYALT LIVER CANCER SUBCLASS G23 UP | 40 | -1 | 3.86e-03 | 0.0787854 |
| M39881 WP MITOCHONDRIAL CIV ASSEMBLY | 24 | 1 | 4.54e-03 | 0.0870674 |
| M14132 SLEBOS HEAD AND NECK CANCER WITH HPV UP | 57 | -1 | 5.16e-03 | 0.0945173 |

B.3.3.2. Análisis de enriquecimiento de genes de la inhibición de la ruta de síntesis de las poliaminas mediante DFMO en las células HT29

En la tabla B.38 se muestra un resumen numérico del análisis de enriquecimiento realizado a partir del análisis de expresión diferencial entre las células HT29 tratadas con DFMO y sin tratar. Para ello, se han cruzado estos datos con 8 colecciones de conjuntos de genes que han sido empeladas a lo largo de la presente tesis, de tal manera que en dicha tabla se muestra no sólo cuántos conjuntos de genes rechazan alguna de las dos hipótesis nulas evaluadas (autocontenida y competitiva), sino también cuántos de aquellos que rechazan alguna de dichas hipótesis nulas pertenecen a las bases de datos relacionadas con el cáncer (MCC y MOG) y/o contienen en su nombre alguna de las siguientes palabras clave de interés para la la temática de la tesis: *calcium*, *mitochondria*, *reticulum*, *cancer*, *wnt* y *polyamines*. Posteriormente, en las tablas B.39-B.44 se muestra en detalle el análisis de enriquecimiento de conjuntos de genes para aquellos conjuntos con un FDR < 0.1 y que contienen alguna de las palabras de interés divididos en diferentes tablas en función tanto de la base de datos a la que pertenecen como del método empleado para su análisis.

Tabla B.38: Análisis de conjuntos de genes para el tratamiento con DFMO en HT29.

| Base de Datos | Conjuntos | | Conjuntos Significativos (FDR <0.05) | | | | | | | |
|---------------|--------------|---------------------|--------------------------------------|-------------|---------------------|--------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | Métodos | Totales | Calcio ¹ | Mitocondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA | 13 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | | ORA | 21 | 0 | 0 | 0 | 0 | 0 | 0 |
| GO BP | 12342 | 146 | CAMERA | 38 | 0 | 2 | 3 | 0 | 0 | 0 |
| | | | ORA | 35 | 0 | 2 | 0 | 0 | 2 | 0 |
| GO MF | 4245 | 66 | CAMERA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| MH | 50 | 0 | CAMERA | 7 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 8 | 0 | 0 | 0 | 0 | 0 | 0 |
| MCDB | 6290 | 12 | CAMERA | 447 | 0 | 4 | 0 | 16 | 3 | 1 |
| | | | ORA | 791 | 0 | 7 | 1 | 55 | 2 | 1 |
| MCC | 858 | 0 | CAMERA | 120 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 131 | 0 | 0 | 0 | 0 | 0 | 0 |
| MOG | 189 | 0 | CAMERA | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| MIM | 5219 | 0 | CAMERA | 107 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | ORA | 1167 | 0 | 0 | 0 | 0 | 1 | 0 |
| Total | 29537 | 225 | CAMERA | 740 | 0 | 6 | 4 | 16 | 3 | 1 |
| | | | ORA | 2160 | 0 | 9 | 1 | 55 | 5 | 1 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

Tabla B.39: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA).

| Conjuntos | nº Genes | DIR | P-valor | FDR |
|--|----------|-----|----------|-----------|
| GO:0006120 mitochondrial electron transport, NADH to ubiquinone | 35 | 1 | 1.00e-07 | 0.0001629 |
| GO:0032981 mitochondrial respiratory chain complex I assembly | 47 | 1 | 1.00e-07 | 0.0001629 |
| GO:0030968 endoplasmic reticulum unfolded protein response | 27 | -1 | 1.24e-04 | 0.0233804 |
| GO:0070059 intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress | 22 | -1 | 1.41e-04 | 0.0233804 |
| GO:0034976 response to endoplasmic reticulum stress | 46 | -1 | 2.66e-04 | 0.0346560 |
| GO:0060071 Wnt signaling pathway, planar cell polarity pathway | 52 | 1 | 7.23e-04 | 0.0578129 |

Tabla B.40: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (ORA).

| Conjuntos | nº Genes | NR.SIG.GENES | P-valor | FDR |
|---|----------|--------------|----------|-----------|
| GO:0032981 mitochondrial respiratory chain complex I assembly | 47 | 34 | 0.00e+00 | 0.0000030 |
| GO:0006120 mitochondrial electron transport, NADH to ubiquinone | 35 | 27 | 0.00e+00 | 0.0000102 |
| GO:0060071 Wnt signaling pathway, planar cell polarity pathway | 52 | 33 | 9.00e-07 | 0.0003693 |
| GO:0090090 negative regulation of canonical Wnt signaling pathway | 105 | 50 | 1.64e-04 | 0.0239309 |

Tabla B.41: Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (CAMERA).

| Conjuntos | nº Genes | DIR | P-valor | FDR |
|--|----------|-----|---------|-----------|
| hsa04141 Protein processing in endoplasmic reticulum | 91 | -1 | 1.5e-06 | 0.0002572 |

Tabla B.42: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA).

| Conjuntos | nº Genes | DIR | P-valor | FDR |
|--|----------|-----|----------|-----------|
| M39417 WP ELECTRON TRANSPORT CHAIN OXPHOS SYSTEM IN MITOCHONDRIA | 65 | 1 | 0.00e+00 | 0.0000077 |
| M15664 ROSTY CERVICAL CANCER PROLIFERATION CLUSTER | 121 | 1 | 1.00e-07 | 0.0000130 |
| M3766 SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP | 133 | 1 | 4.00e-07 | 0.0000439 |

Tabla B.42: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n° Genes | DIR | P-valor | FDR |
|--|----------|-----|----------|-----------|
| M39781 WP MITOCHONDRIAL COMPLEX I ASSEMBLY MODEL OXPHOS SYSTEM | 39 | 1 | 3.00e-06 | 0.0001929 |
| M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT | 152 | 1 | 6.00e-06 | 0.0003324 |
| M7854 WHITEFORD PEDIATRIC CANCER MARKERS | 91 | 1 | 7.50e-06 | 0.0003841 |
| M18855 VECCHI GASTRIC CANCER EARLY UP | 322 | 1 | 1.01e-05 | 0.0004892 |
| M39678 WP RETINOBLASTOMA GENE IN CANCER | 74 | 1 | 9.28e-05 | 0.0030199 |
| M15112 WONG MITOCHONDRIA GENE MODULE | 154 | 1 | 1.12e-04 | 0.0034229 |
| M19982 WAMUNYOKOLI OVARIAN CANCER LMP UP | 173 | 1 | 1.34e-04 | 0.0039272 |
| M8520 SHEDDEN LUNG CANCER POOR SURVIVAL A6 | 332 | 1 | 2.65e-04 | 0.0066717 |
| M27100 REACTOME WNT MEDIATED ACTIVATION OF DVL | 6 | -1 | 4.05e-04 | 0.0096040 |
| M747 REACTOME METABOLISM OF POLYAMINES | 40 | 1 | 4.96e-04 | 0.0112514 |
| M7987 LEE LIVER CANCER SURVIVAL DN | 107 | 1 | 5.02e-04 | 0.0112514 |
| M7847 REACTOME SIGNALING BY WNT | 214 | 1 | 7.06e-04 | 0.0143587 |
| M2180 PUJANA BREAST CANCER WITH BRCA1 MUTATED UP | 43 | 1 | 9.45e-04 | 0.0179424 |
| M39881 WP MITOCHONDRIAL CIV ASSEMBLY | 24 | 1 | 1.19e-03 | 0.0213863 |
| M16586 LAIHO COLORECTAL CANCER SERRATED UP | 76 | 1 | 1.21e-03 | 0.0214756 |
| M17372 LEE LIVER CANCER MYC TGFA DN | 40 | -1 | 1.50e-03 | 0.0250658 |
| M3268 CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP | 143 | 1 | 1.68e-03 | 0.0278296 |
| M10986 CHIANG LIVER CANCER SUBCLASS UNANNOTATED DN | 129 | 1 | 1.89e-03 | 0.0300873 |
| M9951 WANG METHYLATED IN BREAST CANCER | 32 | -1 | 2.07e-03 | 0.0321469 |
| M5636 LEE LIVER CANCER MYC E2F1 DN | 41 | -1 | 3.19e-03 | 0.0439298 |
| M15356 PUJANA BREAST CANCER LIT INT NETWORK | 77 | 1 | 3.48e-03 | 0.0464678 |
| M10922 WANG RECURRENT LIVER CANCER UP | 13 | 1 | 4.16e-03 | 0.0528320 |
| M18241 NIKOLSKY BREAST CANCER 7P15 AMPLICON | 8 | 1 | 4.46e-03 | 0.0557744 |
| M7547 JIANG HYPOXIA CANCER | 39 | 1 | 4.48e-03 | 0.0559022 |
| M424 ANDERSEN LIVER CANCER KRT19 DN | 43 | -1 | 5.49e-03 | 0.0650125 |
| M15472 VECCHI GASTRIC CANCER EARLY DN | 194 | -1 | 6.42e-03 | 0.0726541 |

Tabla B.42: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | nº Genes | DIR | P-valor | FDR |
|---|----------|-----|----------|-----------|
| M2118 SMID BREAST CANCER RELAPSE IN PLEURA DN | 19 | -1 | 6.69e-03 | 0.0745289 |
| M27446 REACTOME MITOCHONDRIAL TRANSLATION | 55 | 1 | 6.76e-03 | 0.0750163 |
| M15626 OUYANG PROSTATE CANCER PROGRESSION UP | 11 | 1 | 7.40e-03 | 0.0793452 |
| M18436 BOYALT LIVER CANCER SUBCLASS G3 UP | 136 | 1 | 7.84e-03 | 0.0820160 |
| M336 VILLANUEVA LIVER CANCER KRT19 UP | 140 | 1 | 7.92e-03 | 0.0824459 |
| M4333 LUI THYROID CANCER CLUSTER 4 | 9 | -1 | 8.07e-03 | 0.0835508 |
| M13831 BOYALT LIVER CANCER SUBCLASS G123 UP | 27 | 1 | 8.29e-03 | 0.0852111 |
| M9389 FINETTI BREAST CANCER KINOME RED | 15 | 1 | 8.43e-03 | 0.0864945 |
| M836 OUELLET CULTURED OVARIAN CANCER INVASIVE VS LMP UP | 43 | 1 | 9.57e-03 | 0.0947878 |

Tabla B.43: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA).

| Conjuntos | nº Genes | NR.SIG.GENES | P-valor | FDR |
|--|----------|--------------|----------|-----------|
| M18855 VECCHI GASTRIC CANCER EARLY UP | 322 | 172 | 0.00e+00 | 0.0000000 |
| M3766 SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP | 133 | 85 | 0.00e+00 | 0.0000000 |
| M15664 ROSTY CERVICAL CANCER PROLIFERATION CLUSTER | 121 | 77 | 0.00e+00 | 0.0000000 |
| M16740 GRADE COLON AND RECTAL CANCER UP | 196 | 109 | 0.00e+00 | 0.0000000 |
| M39417 WP ELECTRON TRANSPORT CHAIN OXPHOS SYSTEM IN MITOCHONDRIA | 65 | 48 | 0.00e+00 | 0.0000000 |
| M15112 WONG MITOCHONDRIA GENE MODULE | 154 | 88 | 0.00e+00 | 0.0000000 |
| M9577 MOOTHA MITOCHONDRIA | 295 | 145 | 0.00e+00 | 0.0000000 |
| M8520 SHEDDEN LUNG CANCER POOR SURVIVAL A6 | 332 | 158 | 0.00e+00 | 0.0000000 |
| M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT | 152 | 85 | 0.00e+00 | 0.0000000 |
| M7854 WHITEFORD PEDIATRIC CANCER MARKERS | 91 | 55 | 0.00e+00 | 0.0000002 |
| M747 REACTOME METABOLISM OF POLYAMINES | 40 | 30 | 0.00e+00 | 0.0000004 |
| M17788 LINDGREN BLADDER CANCER CLUSTER 1 DN | 250 | 115 | 1.00e-07 | 0.0000037 |
| M7847 REACTOME SIGNALING BY WNT | 214 | 101 | 1.00e-07 | 0.0000045 |

Tabla B.43: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | nº Genes | NR.SIG.GENES | P-valor | FDR |
|--|----------|--------------|----------|-----------|
| M19097 SWEET LUNG CANCER KRAS UP | 276 | 124 | 2.00e-07 | 0.0000056 |
| M7987 LEE LIVER CANCER SURVIVAL DN | 107 | 58 | 2.00e-07 | 0.0000071 |
| M3645 DELYS THYROID CANCER UP | 280 | 124 | 4.00e-07 | 0.0000137 |
| M15193 GRUETZMANN PANCREATIC CANCER UP | 224 | 103 | 5.00e-07 | 0.0000148 |
| M10986 CHIANG LIVER CANCER SUBCLASS UNANNOTATED DN | 129 | 66 | 5.00e-07 | 0.0000170 |
| M14507 CHARAFE BREAST CANCER LUMINAL VS BASAL DN | 328 | 141 | 6.00e-07 | 0.0000179 |
| M39678 WP RETINOBLASTOMA GENE IN CANCER | 74 | 43 | 6.00e-07 | 0.0000188 |
| M5792 LI AMPLIFIED IN LUNG CANCER | 100 | 54 | 6.00e-07 | 0.0000191 |
| M3268 CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP | 143 | 70 | 2.00e-06 | 0.0000540 |
| M12016 OUELLET OVARIAN CANCER INVASIVE VS LMP UP | 72 | 41 | 2.30e-06 | 0.0000605 |
| M854 LINDGREN BLADDER CANCER CLUSTER 3 UP | 240 | 106 | 3.30e-06 | 0.0000833 |
| M17299 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL UP | 277 | 118 | 8.00e-06 | 0.0001810 |
| M2180 PUJANA BREAST CANCER WITH BRCA1 MUTATED UP | 43 | 27 | 1.05e-05 | 0.0002299 |
| M8124 SMID BREAST CANCER BASAL UP | 412 | 165 | 1.18e-05 | 0.0002498 |
| M39781 WP MITOCHONDRIAL COMPLEX I ASSEMBLY MODEL OXPHOS SYSTEM | 39 | 25 | 1.32e-05 | 0.0002773 |
| M8365 RHODES UNDIFFERENTIATED CANCER | 48 | 29 | 1.47e-05 | 0.0003022 |
| M9826 VANTVEER BREAST CANCER METASTASIS DN | 81 | 42 | 4.00e-05 | 0.0007034 |
| M1345 RHODES CANCER META SIGNATURE | 41 | 25 | 4.59e-05 | 0.0007901 |
| M9565 AGUIRRE PANCREATIC CANCER COPY NUMBER UP | 169 | 73 | 2.45e-04 | 0.0032868 |
| M16586 LAIHO COLORECTAL CANCER SERRATED UP | 76 | 38 | 2.46e-04 | 0.0032925 |
| M1687 ZHONG SECRETOME OF LUNG CANCER AND FIBROBLAST | 74 | 37 | 2.95e-04 | 0.0038404 |
| M19982 WAMUNYOKOLI OVARIAN CANCER LMP UP | 173 | 74 | 3.19e-04 | 0.0041153 |
| M2631 FARMER BREAST CANCER APOCRINE VS BASAL | 232 | 95 | 3.31e-04 | 0.0042605 |
| M1167 HUMMERICH SKIN CANCER PROGRESSION UP | 52 | 28 | 3.34e-04 | 0.0042798 |
| M2427 NAKAMURA CANCER MICROENVIRONMENT DN | 36 | 21 | 4.42e-04 | 0.0054558 |

Tabla B.43: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | nº Genes | NR.SIG.GENES | P-valor | FDR |
|--|----------|--------------|----------|-----------|
| M5652 FARMER BREAST CANCER BASAL VS LULMINAL | 234 | 95 | 4.69e-04 | 0.0057028 |
| M9951 WANG METHYLATED IN BREAST CANCER | 32 | 19 | 6.08e-04 | 0.0070768 |
| M15150 ZHANG BREAST CANCER PROGENITORS UP | 311 | 120 | 9.83e-04 | 0.0104811 |
| M6862 NADERI BREAST CANCER PROGNOSIS UP | 33 | 19 | 1.03e-03 | 0.0109211 |
| M8817 ACEVEDO LIVER CANCER WITH H3K9ME3 DN | 89 | 41 | 1.21e-03 | 0.0125730 |
| M1685 ZHONG SECRETOME OF LUNG CANCER AND MACROPHAGE | 43 | 23 | 1.26e-03 | 0.0129980 |
| M796 REACTOME ATF4 ACTIVATES GENES IN RESPONSE TO ENDOPLASMIC RETICULUM STRESS | 20 | 13 | 1.41e-03 | 0.0143129 |
| M590 REACTOME MITOCHONDRIAL PROTEIN IMPORT | 41 | 22 | 1.51e-03 | 0.0150342 |
| M4572 LOCKWOOD AMPLIFIED IN LUNG CANCER | 139 | 59 | 1.54e-03 | 0.0153063 |
| M13831 BOYVAULT LIVER CANCER SUBCLASS G123 UP | 27 | 16 | 1.69e-03 | 0.0162873 |
| M16962 IIZUKA LIVER CANCER PROGRESSION L0 L1 DN | 10 | 8 | 1.71e-03 | 0.0163422 |
| M3102 SMID BREAST CANCER ERBB2 UP | 92 | 41 | 2.63e-03 | 0.0233193 |
| M1273 SHETH LIVER CANCER VS TXNIP LOSS PAM1 | 170 | 69 | 2.64e-03 | 0.0233193 |
| M8779 LI LUNG CANCER | 26 | 15 | 3.38e-03 | 0.0287879 |
| M2250 LINDGREN BLADDER CANCER CLUSTER 1 UP | 85 | 38 | 3.46e-03 | 0.0293817 |
| M2435 WAMUNYOKOLI OVARIAN CANCER GRADES 1 2 UP | 88 | 39 | 3.72e-03 | 0.0312185 |
| M13072 SMID BREAST CANCER LUMINAL A DN | 15 | 10 | 3.96e-03 | 0.0327388 |
| M9389 FINETTI BREAST CANCER KINOME RED | 15 | 10 | 3.96e-03 | 0.0327388 |
| M39881 WP MITOCHONDRIAL CIV ASSEMBLY | 24 | 14 | 4.01e-03 | 0.0327388 |
| M10922 WANG RECURRENT LIVER CANCER UP | 13 | 9 | 4.35e-03 | 0.0347210 |
| M18436 BOYVAULT LIVER CANCER SUBCLASS G3 UP | 136 | 56 | 4.42e-03 | 0.0351325 |
| M1218 BORLAK LIVER CANCER EGF UP | 29 | 16 | 4.57e-03 | 0.0359631 |
| M39544 WP MITOCHONDRIAL GENE EXPRESSION | 9 | 7 | 4.57e-03 | 0.0359631 |
| M12795 CHARAFE BREAST CANCER BASAL VS MESENCHYMAL UP | 95 | 41 | 5.30e-03 | 0.0402786 |

Tabla B.43: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | nº Genes | NR.SIG.GENES | P-valor | FDR |
|---|----------|--------------|----------|-----------|
| M17044 WANG HCP PROSTATE CANCER | 68 | 31 | 5.53e-03 | 0.0418042 |
| M7577 ACEVEDO LIVER CANCER DN | 337 | 124 | 5.70e-03 | 0.0428062 |
| M13736 PYEON CANCER HEAD AND NECK VS CERVICAL UP | 152 | 61 | 6.05e-03 | 0.0448453 |
| M14539 BERTUCCI MEDULLARY VS DUCTAL BREAST CANCER UP | 124 | 51 | 6.57e-03 | 0.0479276 |
| M8897 AGUIRRE PANCREATIC CANCER COPY NUMBER DN | 139 | 56 | 7.55e-03 | 0.0536670 |
| M39015 REACTOME ABERRANT REGULATION OF MITOTIC EXIT IN CANCER DUE TO RB1 DEFECTS | 16 | 10 | 7.70e-03 | 0.0542608 |
| M26973 REACTOME MITOCHONDRIAL BIOGENESIS | 51 | 24 | 8.62e-03 | 0.0593534 |
| M3710 WALLACE PROSTATE CANCER UP | 14 | 9 | 8.91e-03 | 0.0606921 |
| M38998 REACTOME ABERRANT REGULATION OF MITOTIC G1 S TRANSITION IN CANCER DUE TO RB1 DEFECTS | 14 | 9 | 8.91e-03 | 0.0606921 |
| M15425 VANTVEER BREAST CANCER ESR1 DN | 143 | 57 | 9.13e-03 | 0.0619576 |
| M9538 GRADE COLON VS RECTAL CANCER UP | 21 | 12 | 9.52e-03 | 0.0642347 |
| M1493 NIKOLSKY BREAST CANCER 11Q12 Q14 AMPLICON | 95 | 40 | 9.56e-03 | 0.0642769 |
| M11896 LINDGREN BLADDER CANCER CLUSTER 2A DN | 98 | 41 | 9.95e-03 | 0.0665079 |
| M1260 WATANABE RECTAL CANCER RADIOTHERAPY RESPONSIVE UP | 65 | 29 | 1.02e-02 | 0.0673129 |
| M18241 NIKOLSKY BREAST CANCER 7P15 AMPLICON | 8 | 6 | 1.19e-02 | 0.0746526 |
| M1275 SHETH LIVER CANCER VS TXNIP LOSS PAM2 | 116 | 47 | 1.21e-02 | 0.0756581 |
| M16955 ACEVEDO LIVER CANCER WITH H3K27ME3 DN | 119 | 48 | 1.24e-02 | 0.0771121 |
| M11504 TOMLINS PROSTATE CANCER DN | 24 | 13 | 1.26e-02 | 0.0774290 |
| M4381 LUI THYROID CANCER PAX8 PPARG UP | 29 | 15 | 1.28e-02 | 0.0783212 |
| M11238 LINDGREN BLADDER CANCER CLUSTER 3 DN | 154 | 60 | 1.31e-02 | 0.0796452 |
| M8475 WANG ESOPHAGUS CANCER VS NORMAL UP | 77 | 33 | 1.31e-02 | 0.0796452 |
| M27288 REACTOME BETA CATENIN INDEPENDENT WNT SIGNALING | 91 | 38 | 1.32e-02 | 0.0801679 |
| M11507 IIZUKA LIVER CANCER PROGRESSION L1 G1 UP | 17 | 10 | 1.37e-02 | 0.0823297 |
| M39901 WP ENDOPLASMIC RETICULUM STRESS RESPONSE IN CORONAVIRUS INFECTION | 22 | 12 | 1.52e-02 | 0.0892785 |

Tabla B.43: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | n ^o Genes | NR.SIG.GENES | P-valor | FDR |
|---|----------------------|--------------|----------|-----------|
| M8411 LINDGREN BLADDER CANCER WITH LOH IN CHR9Q | 78 | 33 | 1.62e-02 | 0.0941906 |
| M2195 LEE LIVER CANCER CIPROFIBRATE UP | 40 | 19 | 1.64e-02 | 0.0941970 |
| M4448 BOYVAULT LIVER CANCER SUBCLASS G23 UP | 40 | 19 | 1.64e-02 | 0.0941970 |
| M9192 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL DN | 314 | 113 | 1.66e-02 | 0.0949640 |

Tabla B.44: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA).

| Conjuntos | n ^o Genes | NR.SIG.GENES | P-valor | FDR |
|---|----------------------|--------------|---------|-----------|
| M8477 GSE26351 WNT VS BMP PATHWAY STIM HEMATOPOIETIC PROGENITORS UP | 124 | 51 | 0.00448 | 0.0253245 |

B.3.3.3. Análisis de enriquecimiento de genes de la expresión diferencial entre las células HT29 y NCM460

En la tabla B.45 se muestra un resumen numérico del análisis de enriquecimiento realizado a partir del análisis de expresión diferencial entre las células HT29 y NCM460. Para ello, se han cruzado estos datos con 8 colecciones de conjuntos de genes que han sido empeladas a lo largo de la presente tesis, de tal manera que en dicha tabla se muestra no sólo cuántos conjuntos de genes rechazan alguna de las dos hipótesis nulas evaluadas (autocontenida y competitiva), sino también cuántos de aquellos que rechazan alguna de dichas hipótesis nulas pertenecen a las bases de datos relacionadas con el cáncer (MCC y MOG) y/o contienen en su nombre alguna de las siguientes palabras clave de interés para la la temática de la tesis: *calcium*, *mitochondria*, *reticulum*, *cancer*, *wnt* y *polyamines*. Posteriormente, en las tablas B.46-B.50 se muestran los resultados del análisis de enriquecimiento de aquellos conjuntos con un FDR < 0.1 que contienen las palabras de interés.

Tabla B.45: Análisis de enriquecimiento de genes para la expresión diferencial entre las células HT29 y NCM460 llevada a cabo con el microarray Clariom D Human.

| Base de Datos | Conjuntos | | Conjuntos Significativos (FDR <0.05) | | | | | | | |
|---------------|--------------|---------------------|--------------------------------------|---------------------|---------------------|--------------------------|------------------------|---------------------|------------------|------------------------|
| | Totales | Calcio ¹ | Métodos | Totales | Calcio ¹ | Mitocondria ¹ | Reticulum ¹ | Cancer ¹ | WNT ¹ | Polyamine ¹ |
| KEGG | 345 | 2 | CAMERA ORA | 25 6 | 0 0 | 0 0 | 1 0 | 1 0 | 0 0 | 0 0 |
| GO BP | 12342 | 146 | CAMERA ORA | 35 10 | 0 0 | 2 0 | 0 0 | 0 0 | 0 0 | 0 0 |
| GO MF | 4245 | 66 | CAMERA ORA | 2 2 | 0 0 | 0 0 | 0 0 | 0 0 | 0 0 | 0 0 |
| MH | 50 | 0 | CAMERA ORA | 11 5 | 0 0 | 0 0 | 0 0 | 0 0 | 0 0 | 0 0 |
| MCDB | 6290 | 12 | CAMERA ORA | 567 528 | 0 0 | 4 6 | 0 0 | 27 46 | 1 3 | 0 0 |
| MCC | 858 | 0 | CAMERA ORA | 89 62 | 0 0 | 0 0 | 0 0 | 0 0 | 0 0 | 0 0 |
| MOG | 189 | 0 | CAMERA ORA | 4 6 | 0 0 | 0 0 | 0 0 | 0 0 | 0 0 | 0 0 |
| MIM | 5219 | 0 | CAMERA ORA | 234 1204 | 0 0 | 0 0 | 0 0 | 0 0 | 0 2 | 0 0 |
| Total | 29537 | 225 | CAMERA ORA | 967 1823 | 0 0 | 6 6 | 1 0 | 28 46 | 1 5 | 0 0 |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Notas:

¹ El número de conjuntos que contienen la palabra del encabezado y que rechazan la hipótesis nula con el método indicado en la columna **Métodos**.

Tabla B.46: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| GO:0032981 mitochondrial respiratory chain complex I assembly | 47 | -1 | 1.05e-05 | 0.0038304 |
| GO:0006120 mitochondrial electron transport, NADH to ubiquinone | 35 | -1 | 2.12e-05 | 0.0059490 |
| GO:0070059 intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress | 22 | 1 | 6.49e-04 | 0.0561475 |
| GO:0006888 endoplasmic reticulum to Golgi vesicle-mediated transport | 108 | 1 | 1.16e-03 | 0.0718710 |
| GO:0042776 mitochondrial ATP synthesis coupled proton transport | 12 | -1 | 1.90e-03 | 0.0855704 |
| GO:0034976 response to endoplasmic reticulum stress | 46 | 1 | 1.97e-03 | 0.0865851 |

Tabla B.47: Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| hsa04141 Protein processing in endoplasmic reticulum | 91 | 1 | 0.000014 | 0.0011690 |
| hsa05216 Thyroid cancer | 22 | 1 | 0.001180 | 0.0231835 |
| hsa05219 Bladder cancer | 31 | 1 | 0.005320 | 0.0612717 |

Tabla B.48: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA).

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| M15664 ROSTY CERVICAL CANCER PROLIFERATION CLUSTER | 121 | -1 | 0.00e+00 | 0.0000000 |
| M3766 SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP | 133 | -1 | 0.00e+00 | 0.0000004 |
| M7854 WHITEFORD PEDIATRIC CANCER MARKERS | 91 | -1 | 0.00e+00 | 0.0000007 |
| M2180 PUJANA BREAST CANCER WITH BRCA1 MUTATED UP | 43 | -1 | 0.00e+00 | 0.0000049 |
| M336 VILLANUEVA LIVER CANCER KRT19 UP | 140 | -1 | 9.00e-07 | 0.0000552 |
| M39417 WP ELECTRON TRANSPORT CHAIN OXPHOS SYSTEM IN MITOCHONDRIA | 65 | -1 | 1.10e-06 | 0.0000625 |
| M39678 WP RETINOBLASTOMA GENE IN CANCER | 74 | -1 | 1.30e-06 | 0.0000722 |
| M8520 SHEDDEN LUNG CANCER POOR SURVIVAL A6 | 332 | -1 | 3.80e-06 | 0.0001752 |
| M14132 SLEBOS HEAD AND NECK CANCER WITH HPV UP | 57 | -1 | 5.90e-06 | 0.0002641 |

Tabla B.48: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|---|---------|---------|----------|-----------|
| M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT | 152 | -1 | 1.79e-05 | 0.0006904 |
| M18855 VECCHI GASTRIC CANCER EARLY UP | 322 | -1 | 1.80e-05 | 0.0006904 |
| M8293 NIKOLSKY BREAST CANCER 8P12 P11 AMPLICON | 39 | -1 | 2.79e-05 | 0.0010192 |
| M39781 WP MITOCHONDRIAL COMPLEX I ASSEMBLY MODEL OXPHOS SYSTEM | 39 | -1 | 4.59e-05 | 0.0015589 |
| M39619 WP SENESENCE AND AUTOPHAGY IN CANCER | 70 | 1 | 1.19e-04 | 0.0033454 |
| M17333 FARMER BREAST CANCER CLUSTER 2 | 31 | -1 | 1.20e-04 | 0.0033454 |
| M9389 FINETTI BREAST CANCER KINOME RED | 15 | -1 | 3.65e-04 | 0.0077758 |
| M3268 CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP | 143 | -1 | 3.86e-04 | 0.0081402 |
| M15112 WONG MITOCHONDRIA GENE MODULE | 154 | -1 | 5.03e-04 | 0.0097777 |
| M9951 WANG METHYLATED IN BREAST CANCER | 32 | 1 | 5.51e-04 | 0.0104617 |
| M9577 MOOTHA MITOCHONDRIA | 295 | -1 | 6.30e-04 | 0.0115770 |
| M8365 RHODES UNDIFFERENTIATED CANCER | 48 | -1 | 6.63e-04 | 0.0119152 |
| M1352 NIKOLSKY BREAST CANCER 15Q26 AMPLICON | 16 | 1 | 7.66e-04 | 0.0132657 |
| M854 LINDGREN BLADDER CANCER CLUSTER 3 UP | 240 | -1 | 9.13e-04 | 0.0150369 |
| M9719 CAVARD LIVER CANCER MALIGNANT VS BENIGN | 17 | 1 | 9.33e-04 | 0.0151912 |
| M6767 POOLA INVASIVE BREAST CANCER DN | 79 | 1 | 1.14e-03 | 0.0175609 |
| M8411 LINDGREN BLADDER CANCER WITH LOH IN CHR9Q | 78 | -1 | 1.22e-03 | 0.0183482 |
| M39523 WP BLADDER CANCER | 30 | 1 | 2.11e-03 | 0.0266784 |
| M19779 FARMER BREAST CANCER CLUSTER 6 | 13 | 1 | 2.38e-03 | 0.0288783 |
| M11629 MONTERO THYROID CANCER POOR SURVIVAL UP | 10 | -1 | 2.48e-03 | 0.0297756 |
| M39347 WP GASTRIC CANCER NETWORK 1 | 20 | -1 | 3.83e-03 | 0.0409130 |
| M523 KEGG THYROID CANCER | 17 | 1 | 4.59e-03 | 0.0469264 |
| M8817 ACEVEDO LIVER CANCER WITH H3K9ME3 DN | 89 | -1 | 4.83e-03 | 0.0488557 |
| M38998 REACTOME ABERRANT REGULATION OF MITOTIC G1 S TRANSITION IN CANCER DUE TO RB1 DEFECTS | 14 | -1 | 5.11e-03 | 0.0510553 |
| M39688 WP APOPTOSISRELATED NETWORK DUE TO ALTERED NOTCH3 IN OVARIAN CANCER | 37 | 1 | 5.35e-03 | 0.0523549 |

Tabla B.48: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | n°Genes | Sentido | P-valor | FDR |
|--|---------|---------|----------|-----------|
| M14507 CHARAFE BREAST CANCER LUMINAL VS BASAL DN | 328 | 1 | 5.47e-03 | 0.0533465 |
| M14690 REACTOME MITOCHONDRIAL FATTY ACID BETA OXIDATION | 28 | -1 | 5.85e-03 | 0.0558143 |
| M3102 SMID BREAST CANCER ERBB2 UP | 92 | 1 | 6.00e-03 | 0.0567728 |
| M17788 LINDGREN BLADDER CANCER CLUSTER 1 DN | 250 | -1 | 6.07e-03 | 0.0569623 |
| M6364 HEIDENBLAD AMPLIFIED IN SOFT TISSUE CANCER | 8 | 1 | 6.22e-03 | 0.0578946 |
| M16586 LAIHO COLORECTAL CANCER SERRATED UP | 76 | 1 | 6.85e-03 | 0.0619426 |
| M3645 DELYS THYROID CANCER UP | 280 | 1 | 7.38e-03 | 0.0648872 |
| M27719 REACTOME MITOCHONDRIAL FATTY ACID BETA OXIDATION OF SATURATED FATTY ACIDS | 8 | -1 | 8.42e-03 | 0.0703513 |
| M39892 WP MITOCHONDRIAL CIII ASSEMBLY | 12 | -1 | 8.63e-03 | 0.0714169 |
| M27446 REACTOME MITOCHONDRIAL TRANSLATION | 55 | -1 | 8.66e-03 | 0.0714169 |
| M7987 LEE LIVER CANCER SURVIVAL DN | 107 | -1 | 9.52e-03 | 0.0764140 |
| M39015 REACTOME ABERRANT REGULATION OF MITOTIC EXIT IN CANCER DUE TO RB1 DEFECTS | 16 | -1 | 9.56e-03 | 0.0766275 |
| M26973 REACTOME MITOCHONDRIAL BIOGENESIS | 51 | -1 | 1.06e-02 | 0.0823084 |
| M7847 REACTOME SIGNALING BY WNT | 214 | -1 | 1.09e-02 | 0.0840668 |
| M590 REACTOME MITOCHONDRIAL PROTEIN IMPORT | 41 | -1 | 1.18e-02 | 0.0882683 |
| M15807 BONOME OVARIAN CANCER POOR SURVIVAL DN | 16 | -1 | 1.23e-02 | 0.0911732 |
| M5091 EINAV INTERFERON SIGNATURE IN CANCER | 20 | 1 | 1.24e-02 | 0.0914400 |
| M19661 WANG ESOPHAGUS CANCER PROGRESSION UP | 5 | 1 | 1.26e-02 | 0.0923192 |
| M10986 CHIANG LIVER CANCER SUBCLASS UNANNOTATED DN | 129 | -1 | 1.30e-02 | 0.0942830 |
| M19612 LEE LIVER CANCER ACOX1 UP | 43 | 1 | 1.40e-02 | 0.0996389 |

Tabla B.49: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA).

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M3766 SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP | 133 | 126 | 0.00e+00 | 0.0000000 |

Tabla B.49: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | nºGenes | nºGenes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M8520 SHEDDEN LUNG CANCER POOR SURVIVAL A6 | 332 | 268 | 0.00e+00 | 0.0000000 |
| M854 LINDGREN BLADDER CANCER CLUSTER 3 UP | 240 | 202 | 0.00e+00 | 0.0000000 |
| M15664 ROSTY CERVICAL CANCER PROLIFERATION CLUSTER | 121 | 113 | 0.00e+00 | 0.0000000 |
| M18855 VECCHI GASTRIC CANCER EARLY UP | 322 | 258 | 0.00e+00 | 0.0000000 |
| M7854 WHITEFORD PEDIATRIC CANCER MARKERS | 91 | 84 | 0.00e+00 | 0.0000000 |
| M17788 LINDGREN BLADDER CANCER CLUSTER 1 DN | 250 | 193 | 0.00e+00 | 0.0000000 |
| M39678 WP RETINOBLASTOMA GENE IN CANCER | 74 | 66 | 0.00e+00 | 0.0000001 |
| M3268 CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP | 143 | 115 | 0.00e+00 | 0.0000002 |
| M16740 GRADE COLON AND RECTAL CANCER UP | 196 | 150 | 0.00e+00 | 0.0000006 |
| M9577 MOOTHA MITOCHONDRIA | 295 | 214 | 0.00e+00 | 0.0000014 |
| M2180 PUJANA BREAST CANCER WITH BRCA1 MUTATED UP | 43 | 40 | 2.00e-07 | 0.0000094 |
| M336 VILLANUEVA LIVER CANCER KRT19 UP | 140 | 108 | 6.00e-07 | 0.0000235 |
| M14132 SLEBOS HEAD AND NECK CANCER WITH HPV UP | 57 | 49 | 2.90e-06 | 0.0001057 |
| M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT | 152 | 114 | 3.40e-06 | 0.0001207 |
| M8365 RHODES UNDIFFERENTIATED CANCER | 48 | 42 | 5.80e-06 | 0.0001919 |
| M19982 WAMUNYOKOLI OVARIAN CANCER LMP UP | 173 | 127 | 6.40e-06 | 0.0002047 |
| M17333 FARMER BREAST CANCER CLUSTER 2 | 31 | 29 | 8.50e-06 | 0.0002609 |
| M15150 ZHANG BREAST CANCER PROGENITORS UP | 311 | 214 | 1.38e-05 | 0.0003866 |
| M10986 CHIANG LIVER CANCER SUBCLASS UNANNOTATED DN | 129 | 97 | 1.49e-05 | 0.0004074 |
| M17299 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL UP | 277 | 192 | 1.88e-05 | 0.0005044 |
| M7987 LEE LIVER CANCER SURVIVAL DN | 107 | 82 | 2.01e-05 | 0.0005318 |
| M15112 WONG MITOCHONDRIA GENE MODULE | 154 | 113 | 2.12e-05 | 0.0005532 |
| M6862 NADERI BREAST CANCER PROGNOSIS UP | 33 | 30 | 2.55e-05 | 0.0006477 |
| M17936 NIKOLSKY BREAST CANCER 8Q23 Q24 AMPLICON | 102 | 78 | 3.65e-05 | 0.0008977 |
| M7847 REACTOME SIGNALING BY WNT | 214 | 150 | 6.23e-05 | 0.0014473 |

Tabla B.49: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | n°Genes | n°Genes Significativos | P-valor | FDR |
|--|---------|------------------------|----------|-----------|
| M5792 LI AMPLIFIED IN LUNG CANCER | 100 | 76 | 6.65e-05 | 0.0015202 |
| M13736 PYEON CANCER HEAD AND NECK VS CERVICAL UP | 152 | 110 | 7.31e-05 | 0.0016644 |
| M3645 DELYS THYROID CANCER UP | 280 | 191 | 8.70e-05 | 0.0019197 |
| M5652 FARMER BREAST CANCER BASAL VS LULMINAL | 234 | 162 | 9.12e-05 | 0.0019970 |
| M39417 WP ELECTRON TRANSPORT CHAIN OXPHOS SYSTEM IN MITOCHONDRIA | 65 | 52 | 9.17e-05 | 0.0019970 |
| M1273 SHETH LIVER CANCER VS TXNIP LOSS PAM1 | 170 | 121 | 1.09e-04 | 0.0023507 |
| M1345 RHODES CANCER META SIGNATURE | 41 | 35 | 1.12e-04 | 0.0024063 |
| M4572 LOCKWOOD AMPLIFIED IN LUNG CANCER | 139 | 100 | 2.21e-04 | 0.0044161 |
| M9389 FINETTI BREAST CANCER KINOME RED | 15 | 15 | 2.30e-04 | 0.0045483 |
| M8124 SMID BREAST CANCER BASAL UP | 412 | 270 | 2.75e-04 | 0.0053275 |
| M11896 LINDGREN BLADDER CANCER CLUSTER 2A DN | 98 | 73 | 2.77e-04 | 0.0053482 |
| M590 REACTOME MITOCHONDRIAL PROTEIN IMPORT | 41 | 34 | 4.43e-04 | 0.0079544 |
| M1170 HUMMERICH SKIN CANCER PROGRESSION DN | 54 | 43 | 4.52e-04 | 0.0080473 |
| M19097 SWEET LUNG CANCER KRAS UP | 276 | 185 | 4.67e-04 | 0.0082629 |
| M9826 VANTVEER BREAST CANCER METASTASIS DN | 81 | 61 | 5.28e-04 | 0.0091998 |
| M16858 OSMAN BLADDER CANCER DN | 230 | 155 | 9.09e-04 | 0.0143506 |
| M15193 GRUETZMANN PANCREATIC CANCER UP | 224 | 151 | 1.03e-03 | 0.0159093 |
| M8411 LINDGREN BLADDER CANCER WITH LOH IN CHR9Q | 78 | 58 | 1.24e-03 | 0.0183591 |
| M2544 ROESSLER LIVER CANCER METASTASIS UP | 66 | 50 | 1.32e-03 | 0.0191460 |
| M11213 SATO SILENCED BY DEACETYLATION IN PANCREATIC CANCER | 31 | 26 | 1.56e-03 | 0.0216920 |
| M15356 PUJANA BREAST CANCER LIT INT NETWORK | 77 | 57 | 1.63e-03 | 0.0223392 |
| M39628 WP INTEGRATED CANCER PATHWAY | 37 | 30 | 1.98e-03 | 0.0257761 |
| M1758 SANSOM WNT PATHWAY REQUIRE MYC | 46 | 36 | 2.34e-03 | 0.0297180 |
| M9565 AGUIRRE PANCREATIC CANCER COPY NUMBER UP | 169 | 115 | 2.39e-03 | 0.0300856 |
| M5275 OSMAN BLADDER CANCER UP | 271 | 178 | 2.46e-03 | 0.0308987 |
| M12795 CHARAFE BREAST CANCER BASAL VS MESENCHYMAL UP | 95 | 68 | 2.67e-03 | 0.0333167 |

Tabla B.49: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | n° Genes | n° Genes Significativos | P-valor | FDR |
|--|----------|-------------------------|----------|-----------|
| M39781 WP MITOCHONDRIAL COMPLEX I ASSEMBLY MODEL OXPHOS SYSTEM | 39 | 31 | 3.03e-03 | 0.0368435 |
| M26973 REACTOME MITOCHONDRIAL BIOGENESIS | 51 | 39 | 3.31e-03 | 0.0392462 |
| M11629 MONTERO THYROID CANCER POOR SURVIVAL UP | 10 | 10 | 3.76e-03 | 0.0430629 |
| M14539 BERTUCCI MEDULLARY VS DUCTAL BREAST CANCER UP | 124 | 85 | 6.16e-03 | 0.0635458 |
| M2435 WAMUNYOKOLI OVARIAN CANCER GRADES 1 2 UP | 88 | 62 | 7.19e-03 | 0.0706778 |
| M1286 HOSHIDA LIVER CANCER SUBCLASS S3 | 176 | 117 | 7.20e-03 | 0.0706778 |
| M14693 VANTVEER BREAST CANCER POOR PROGNOSIS | 33 | 26 | 8.14e-03 | 0.0765133 |
| M17044 WANG HCP PROSTATE CANCER | 68 | 49 | 8.21e-03 | 0.0769183 |
| M14631 KEGG COLORECTAL CANCER | 45 | 34 | 8.29e-03 | 0.0772877 |
| M18694 LIU COMMON CANCER GENES | 36 | 28 | 8.34e-03 | 0.0775010 |
| M6176 HOSHIDA LIVER CANCER LATE RECURRENCE UP | 36 | 28 | 8.34e-03 | 0.0775010 |
| M39347 WP GASTRIC CANCER NETWORK 1 | 20 | 17 | 8.44e-03 | 0.0776725 |
| M16586 LAIHO COLORECTAL CANCER SERRATED UP | 76 | 54 | 8.99e-03 | 0.0817727 |
| M2108 VECCHI GASTRIC CANCER ADVANCED VS EARLY DN | 106 | 73 | 9.00e-03 | 0.0817727 |
| M15425 VANTVEER BREAST CANCER ESR1 DN | 143 | 96 | 9.41e-03 | 0.0850920 |
| M39758 WP CHROMOSOMAL AND MICROSATELLITE INSTABILITY IN COLORECTAL CANCER | 50 | 37 | 1.05e-02 | 0.0921774 |
| M8817 ACEVEDO LIVER CANCER WITH H3K9ME3 DN | 89 | 62 | 1.06e-02 | 0.0926284 |
| M500 VANTVEER BREAST CANCER METASTASIS UP | 29 | 23 | 1.11e-02 | 0.0952500 |
| M2840 CHARAFE BREAST CANCER LUMINAL VS BASAL UP | 235 | 152 | 1.12e-02 | 0.0956771 |

Tabla B.50: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA).

| Conjuntos | n° Genes | n° Genes Signif. | P-valor | FDR |
|--|----------|------------------|----------|-----------|
| M8477 GSE26351 WNT VS BMP PATHWAY STIM HEMATOPOIETIC PROGENITORS UP | 124 | 89 | 0.000313 | 0.0041958 |
| M8480 GSE26351 WNT VS BMP PATHWAY STIM HEMATOPOIETIC PROGENITORS DN | 125 | 89 | 0.000487 | 0.0057123 |

Tabla B.50: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA) (continuación)

| Conjuntos | n° Genes | n° Genes Signif. | P-valor | FDR |
|---|-----------------|-------------------------|----------------|------------|
| M8476 GSE26351 UNSTIM VS WNT PATHWAY STIM HEMATOPOIETIC PROGENITORS DN | 136 | 89 | 0.020700 | 0.0731731 |

B.4. Reanálisis transcriptómico de muestras de tejido humano de cáncer de colon extraídas de repositorios públicos y metaanálisis

En este apartado se muestra el material suplementario correspondiente a los reanálisis y metaanálisis presentado en el “7.2.3 Reanálisis transcriptómico de muestras de tejido humano de cáncer de colon extraídas de repositorios públicos y metaanálisis”. Cabe destacar que los reanálisis han sido realizados a partir de los datos transcriptómicos crudos (*.CEL*) extraídos del repositorio público NCBI GEO, generados por microarrays a partir de diferentes experimentos realizados sobre tejido humano de cáncer y normal. Posteriormente, los resultados extraídos de estos reanálisis junto con los por nuestro grupo -también en tejido humano de cáncer de colon y normal- (apartado “7.2.1 Análisis transcriptómico de muestras pareadas de tejido humano de cáncer de colon”) fueron empleados para realizar un metaanálisis; es decir, que todos estos resultados, generados de forma independiente, se sintetizan de forma cuantitativa para obtener un único resultado del efecto que se está evaluando (expresión diferencial de un gen, relación de un conjunto de genes con un fenotipo...). Así, lo que se obtiene es una estimación del efecto global más precisa y exacta sobre el estudio en cuestión.

B.4.1. Reanálisis y metaanálisis de expresión diferencial

A continuación, se muestran los resultados de expresión diferencial obtenidos para cada uno de los experimentos reanalizados y de los metaanálisis realizados a partir de ellos. En concreto, se muestran únicamente los resultados correspondientes a los genes relacionados con la homeostasis del Ca^{2+} intracelular dividido en las familias de genes VOCC, SOCE, TRP, CP, CLC y MCT (tablas B.51-B.56, respectivamente).

Tabla B.51: Metaanálisis del conjunto de genes VOCC.

| Genes | Nuestras | | | GSE8671 | | | GSE39582 | | | GSE44076 | | | GSE41258 | | | GSE33113 | | | Metaanálisis | | | |
|--------|----------|---------|----------|---------|----------|---------|----------|---------|----------|----------|----------|---------|----------|---------|----------|----------|-----|--------|--------------|-----|--------|--|
| | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | |
| | | | | | | | | | | | | | | | | | | | | | | |
| Cav1.1 | -4,4E-02 | 8,9E-01 | -1,2E-01 | 5,2E-02 | 4,0E-02 | 7,4E-01 | -6,8E-02 | 1,6E-02 | -5,2E-02 | 5,2E-02 | 2,2E-01 | 9,1E-02 | -1,9E-01 | 3,0E-01 | -1,6 | 1,5E-01 | | | | | | |
| Cav1.2 | -6,3E-01 | 3,4E-01 | 8,8E-02 | 2,7E-03 | 1,8E-01 | 4,8E-03 | -2,6E-02 | 5,9E-02 | -4,7E-02 | 1,1E-01 | -4,7E-01 | 7,5E-08 | -2,9E-01 | 5,1E-01 | -6,5E-01 | 5,7E-01 | | | | | | |
| Cav1.3 | 2,2E-01 | 6,4E-01 | 6,0E-02 | 1,0E-01 | 2,1E-01 | 1,3E-02 | 2,0E-01 | 3,3E-16 | 2,6E-01 | 2,1E-10 | 6,3E-01 | 1,1E-05 | 1,1 | 8,3E-07 | 1,0E+01 | 0 | | | | | | |
| Cav1.4 | 4,1E-03 | 9,9E-01 | -1,1E-01 | 3,9E-02 | 1,7E-02 | 8,8E-01 | -1,1E-01 | 4,1E-04 | 3,4E-02 | 3,2E-01 | 1,9E-01 | 9,9E-02 | -7,5E-02 | 7,9E-01 | -8,3E-01 | 4,5E-01 | | | | | | |
| Cav2.1 | 2,0E-01 | 6,1E-01 | -1,7E-01 | 1,1E-07 | -1,8E-01 | 6,0E-07 | -2,6E-01 | 9,0E-17 | -2,2E-02 | 2,8E-01 | -5,7E-02 | 3,0E-01 | -9,0E-01 | 1,0E-02 | -7,9 | 6,6E-15 | | | | | | |
| Cav2.2 | -1,2E-01 | 6,7E-01 | 4,6E-02 | 3,2E-01 | 2,3E-02 | 7,5E-01 | -2,5E-02 | 2,8E-01 | 7,2E-02 | 3,7E-02 | -3,7E-02 | 6,7E-01 | 7,9E-02 | 6,3E-01 | 7,4E-01 | 5,1E-01 | | | | | | |
| Cav2.3 | 2,3E-01 | 4,3E-01 | 2,1E-02 | 2,7E-01 | -5,6E-02 | 4,4E-02 | -3,7E-02 | 2,5E-02 | 2,0E-03 | 9,4E-01 | 1,2E-01 | 2,4E-03 | 1,3E-01 | 6,7E-01 | -1,1 | 3,0E-01 | | | | | | |
| Cav3.1 | -2,2E-01 | 4,9E-01 | -4,4E-02 | 5,9E-02 | -1,5E-02 | 6,6E-01 | -7,2E-02 | 1,3E-06 | 2,2E-02 | 2,8E-02 | 1,2E-01 | 6,5E-03 | -5,3E-02 | 9,0E-01 | -1,1 | 3,0E-01 | | | | | | |
| Cav3.2 | -7,6E-01 | 2,0E-01 | -1,2E-01 | 2,4E-02 | 1,6E-02 | 8,7E-01 | -6,3E-02 | 4,2E-01 | -2,4E-01 | 1,1E-07 | -4,0E-01 | 2,4E-04 | -6,7E-01 | 1,9E-02 | -4,5 | 1,4E-05 | | | | | | |
| Cav3.3 | 2,4E-01 | 4,6E-01 | -9,2E-02 | 1,5E-02 | -1,7E-02 | 8,0E-01 | -9,3E-03 | 6,3E-01 | 7,6E-03 | 7,2E-01 | -1,5E-01 | 2,5E-02 | -2,1E-01 | 3,0E-01 | -1,4 | 1,9E-01 | | | | | | |

Tabla B.52: Metaanálisis del conjunto de genes SOCE.

| Genes | Nuestras | | | GSE8671 | | | GSE39582 | | | GSE44076 | | | GSE41258 | | | GSE33113 | | | Metaanálisis | | | | |
|---------|----------|---------|----------|---------|----------|---------|----------|---------|----------|----------|----------|---------|----------|---------|----------|----------|-----|--------|--------------|-----|--------|--|--|
| | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | logFC | FDR | Efecto | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
| ORAI1 | 4,4E-01 | 2,0E-01 | 4,7E-01 | 2,8E-10 | 3,2E-01 | 5,4E-02 | 5,9E-01 | 2,9E-22 | 1,9E-01 | 4,2E-01 | 6,8 | 7,6E-04 | 1,2 | 7,6E-04 | 6,8 | 2,5E-11 | | | | | | | |
| ORAI2 | 4,0E-02 | 9,2E-01 | -3,9E-02 | 1,6E-01 | 2,0E-01 | 6,7E-04 | 2,2E-01 | 5,7E-21 | 1,3E-01 | 1,9E-10 | 2,2E-01 | 1,1E-03 | 8,8E-01 | 1,1E-02 | 9 | 0 | | | | | | | |
| ORAI3 | 1,8E-02 | 9,6E-01 | -3,0E-01 | 1,4E-04 | 1,8E-01 | 2,7E-01 | 1,2E-01 | 1,2E-03 | 1,0E-01 | 3,3E-02 | 4,7E-02 | 8,6E-01 | 6,1E-02 | 8,5E-01 | 2,5 | 1,7E-02 | | | | | | | |
| STIM1 | -2,5E-01 | 2,6E-01 | -3,9E-01 | 6,2E-14 | -3,3E-01 | 5,0E-05 | -8,5E-01 | 5,5E-51 | -3,4E-01 | 1,7E-20 | -7,7E-01 | 2,0E-09 | -2,2 | 1,1E-07 | -1,3E+01 | 0 | | | | | | | |
| STIM2 | 1,7E-01 | 6,6E-01 | 2,5E-01 | 1,2E-06 | -6,3E-02 | 5,3E-01 | -1,1E-01 | 3,0E-04 | 1,1E-01 | 4,8E-01 | 1,1E-01 | 4,8E-01 | 2,9E-01 | 5,7E-01 | -8,0E-01 | 4,7E-01 | | | | | | | |
| MS4A12 | -3,7 | 6,1E-04 | -5,8 | 5,4E-20 | -7,2 | 4,6E-21 | -7 | 2,8E-80 | -4,6 | 5,7E-90 | -6 | 3,6E-07 | -3,8 | 1,3E-12 | -1,2E+01 | 0 | | | | | | | |
| CRACR2A | 2,6E-01 | 5,3E-01 | -2,3E-02 | 4,8E-01 | 7,6E-02 | 3,2E-01 | 1,1E-01 | 1,1E-04 | 1,2E-01 | 1,4E-01 | 3,3 | 3,0E-02 | 3,7E-01 | 3,0E-02 | 3,3 | 1,8E-03 | | | | | | | |

Tabla B.53: Metaanálisis del conjunto de genes TRP.

| Genes | Nuestras | | | GSE8671 | | | GSE39582 | | | GSE44076 | | | GSE41258 | | | GSE33113 | | | Metaanálisis | | | |
|--------|----------|---------|---------|----------|----------|---------|----------|---------|----------|----------|---------|----------|----------|----------|----------|----------|----------|---------|--------------|-----|--------|-----|
| | logFC | FDR | FDR | logFC | FDR | FDR | logFC | FDR | FDR | logFC | FDR | FDR | logFC | FDR | FDR | logFC | FDR | FDR | Efecto | FDR | Efecto | FDR |
| | | | | | | | | | | | | | | | | | | | | | | |
| TRPC1 | -1,7E-01 | 7,0E-01 | 6,5E-15 | 1,8E-02 | 8,5E-01 | 8,7E-04 | -3,5E-01 | 8,7E-04 | 1,1E-08 | -1,0E-01 | 1,1E-08 | 3,4E-04 | -5,7E-01 | 3,4E-04 | -1,1 | 6,7E-03 | -6,8 | 3,5E-11 | | | | |
| TRPC3 | 9,6E-03 | 9,8E-01 | 8,9E-01 | -5,3E-02 | 1,8E-01 | 1,4E-03 | -5,4E-02 | 1,4E-03 | 1,8E-02 | 1,4E-01 | 1,2E-01 | 1,0E-01 | 1,0E-01 | 1,2E-01 | -3,9E-02 | 8,9E-01 | -1,3 | 2,5E-01 | | | | |
| TRPC4 | 2,0E-01 | 4,8E-01 | 4,6E-01 | 8,9E-02 | 2,9E-02 | 2,7E-01 | -2,9E-02 | 2,7E-01 | 6,5E-02 | 4,1E-01 | 4,1E-01 | 4,1E-01 | -4,8E-02 | 4,1E-01 | -5,6E-02 | 8,0E-01 | -3,0E-02 | 9,8E-01 | | | | |
| TRPC5 | -7,0E-02 | 8,2E-01 | 3,8E-01 | 3,5E-02 | 3,8E-01 | 8,5E-08 | -9,8E-02 | 3,4E-01 | 1,6E-02 | 1,5E-01 | 6,6E-02 | 1,5E-01 | 1,5E-01 | 6,6E-02 | 2,4E-02 | 9,5E-01 | -1,2 | 2,9E-01 | | | | |
| TRPC6 | 4,7E-02 | 8,7E-01 | 1,8E-03 | -3,6E-02 | 6,3E-01 | 3,7E-01 | -2,2E-02 | 3,7E-01 | 6,0E-02 | 1,1E-04 | 9,0E-02 | -2,4E-01 | -6,0E-02 | 1,1E-04 | -4,6E-01 | 1,9E-02 | -3,8 | 2,6E-04 | | | | |
| TRPC7 | -2,2E-01 | 2,7E-01 | 1,2E-02 | 1,0E-02 | 8,5E-01 | 3,9E-46 | 3,0E-01 | 3,9E-46 | 1,5E-01 | 7,3E-09 | 8,8E-04 | 2,2E-01 | 2,2E-01 | 8,8E-04 | 9,9E-01 | 7,5E-02 | 6,8 | 3,1E-11 | | | | |
| TRPM1 | -3,6E-01 | 1,9E-01 | 7,7E-01 | -3,0E-02 | 2,9E-01 | 9,0E-06 | -5,3E-02 | 9,0E-06 | 7,2E-01 | 7,2E-01 | 2,8E-01 | 4,6E-02 | 4,6E-02 | 2,8E-01 | -2,3E-01 | 2,9E-01 | -2,7 | 1,0E-02 | | | | |
| TRPM2 | -4,8E-01 | 4,3E-02 | 9,3E-07 | 1,5E-01 | 1,1E-01 | 4,4E-59 | 4,3E-01 | 4,4E-59 | 1,5E-18 | 1,5E-18 | 7,7E-01 | -3,1E-02 | 7,7E-01 | -1,4E-01 | 9,2E-01 | -1,6 | 1,3E-01 | | | | | |
| TRPM3 | -2,5E-01 | 3,7E-01 | 3,5E-01 | -6,6E-02 | 1,1E-02 | 1,0E-02 | -3,8E-02 | 3,3E-02 | 3,3E-02 | 1,7E-02 | 9,6E-03 | 1,1E-01 | 1,1E-01 | 9,6E-03 | 3,4E-02 | 9,3E-01 | -1 | 3,5E-01 | | | | |
| TRPM4 | 2,6E-01 | 4,7E-01 | 9,2E-01 | -1,5 | 4,1E-08 | 1,3E-17 | -1 | 1,3E-17 | -9,0E-01 | 2,1E-18 | 8,7E-04 | -1,4 | 8,7E-04 | -1,1 | 1,5E-03 | -1,0E+01 | 0 | 3,7E-06 | | | | |
| TRPM5 | -1,6E-01 | 6,4E-01 | 7,7E-01 | -1,2E-01 | 3,8E-01 | 6,9E-07 | -1,6E-01 | 6,9E-07 | -5,7E-01 | 4,8E-29 | 9,0E-05 | -6,1E-01 | -6,1E-01 | 9,0E-05 | -6,3E-01 | 2,4E-02 | -4,8 | 3,7E-06 | | | | |
| TRPM6 | -5,7E-01 | 4,1E-01 | 1,1E-28 | -2,2 | 7,7E-130 | 2,3E-80 | -3,1 | 2,3E-80 | -5,7E-01 | 4,8E-29 | 1,1E-22 | -1,3 | 1,1E-22 | -5,2 | 2,2E-04 | -2,3 | 3,1E-02 | | | | | |
| TRPM7 | 7,3E-01 | 1,3E-01 | 3,8E-02 | -2,4E-01 | 6,0E-06 | 2,3E-32 | -5,3E-01 | 2,3E-32 | 3,3E-01 | 3,3E-01 | 5,9E-02 | -1,3E-01 | -1,3E-01 | 5,9E-02 | -8,6E-01 | 1,0E-01 | -6,8 | 2,7E-11 | | | | |
| TRPM8 | -1,5E-01 | 6,1E-01 | 4,0E-02 | -1,4E-01 | 6,1E-02 | 3,2E-01 | -1,3E-02 | 3,2E-01 | -2,9E-02 | 3,3E-01 | 2,8E-01 | -9,7E-02 | -9,7E-02 | 2,8E-01 | -1,9E-01 | 1,5E-01 | -2,6 | 1,6E-02 | | | | |
| TRPML1 | 2,9E-01 | 3,4E-01 | 4,0E-06 | -7,9E-01 | 1,6E-08 | 1,4E-47 | -6,0E-01 | 1,4E-47 | -1,5E-01 | 3,8E-04 | 5,8E-01 | 1,4E-01 | 1,4E-01 | 5,8E-01 | -1 | 7,6E-02 | -8,9 | 0 | | | | |
| TRPML2 | -1,9E-01 | 7,5E-01 | 4,9E-20 | -2,5 | 2,0E-30 | 7,2E-51 | -2,1 | 7,2E-51 | -1,6 | 3,1E-08 | 3,1E-08 | -1,6 | 3,1E-08 | -2,9 | 4,6E-07 | -1,1E+01 | 0 | 0 | | | | |
| TRPML3 | 4,1E-01 | 5,0E-01 | 1,3E-09 | 6,9E-02 | 4,1E-01 | 3,5E-02 | -5,5E-02 | 3,5E-02 | 5,8E-02 | 2,5E-01 | 9,0E-01 | 2,0E-02 | 2,0E-02 | 9,0E-01 | -2,6E-01 | 5,1E-01 | -8,6E-01 | 4,4E-01 | | | | |
| TRPV1 | 5,2E-01 | 5,8E-02 | 3,7E-02 | -1,1E-01 | 9,5E-02 | 9,2E-12 | -1,4E-01 | 9,2E-12 | -3,0E-02 | 2,0E-01 | 5,8E-02 | -1,7E-01 | -1,7E-01 | 5,8E-02 | -4,8E-01 | 8,4E-02 | -5,7 | 3,0E-08 | | | | |
| TRPV2 | 2,0E-01 | 5,6E-01 | 7,6E-05 | 1,5E-02 | 9,0E-01 | 7,1E-03 | -6,5E-02 | 7,1E-03 | 9,8E-02 | 3,2E-02 | 7,4E-02 | 2,3E-01 | 2,3E-01 | 7,4E-02 | -3,3E-02 | 9,3E-01 | -2,5E-01 | 8,3E-01 | | | | |
| TRPV3 | -5,4E-01 | 2,2E-01 | 2,0E-01 | 2,4E-02 | 6,7E-01 | 1,6E-11 | -8,5E-02 | 1,6E-11 | 6,7E-01 | 1,6E-11 | 8,1E-02 | -1,2E-01 | -1,2E-01 | 8,1E-02 | -5,8E-01 | 4,4E-02 | -4 | 9,5E-05 | | | | |
| TRPV4 | 8,3E-02 | 7,7E-01 | 2,1E-01 | 3,5E-01 | 9,2E-03 | 1,1E-10 | 1,2E-01 | 1,1E-10 | 1,1E-01 | 1,0E-03 | 9,0E-01 | 2,1E-02 | 2,1E-02 | 9,0E-01 | 4,5E-01 | 6,8E-02 | 6,1 | 2,6E-09 | | | | |
| TRPV5 | 3,2E-01 | 1,9E-01 | 3,8E-02 | 3,7E-03 | 9,6E-01 | 6,7E-01 | 9,0E-03 | 6,7E-01 | 2,7E-03 | 9,4E-01 | 6,4E-03 | 2,1E-01 | 2,1E-01 | 6,4E-03 | 1,7E-01 | 4,2E-01 | 1,1 | 3,0E-01 | | | | |
| TRPV6 | 4,5E-01 | 1,8E-01 | 9,0E-03 | 1,0E-01 | 2,2E-01 | 1,6E-02 | -4,8E-02 | 1,6E-02 | 5,5E-03 | 9,0E-01 | 8,1E-01 | -2,2E-02 | -2,2E-02 | 8,1E-01 | -8,7E-02 | 7,2E-01 | -4,9E-01 | 6,7E-01 | | | | |
| TRPA1 | 1,2E-01 | 6,6E-01 | 2,5E-01 | -2,5E-01 | 2,6E-04 | 8,3E-09 | -2,0E-01 | 8,3E-09 | -1,1E-01 | 3,4E-07 | 1,7E-01 | -1,6E-01 | -1,6E-01 | 1,7E-01 | -5,8E-01 | 2,4E-02 | -7,7 | 6,1E-14 | | | | |
| TRPP1 | 4,6E-01 | 1,8E-01 | 1,5E-02 | 1,1E-02 | 8,6E-01 | 1,6E-01 | -3,8E-02 | 1,6E-01 | -4,4E-02 | 7,1E-02 | 6,0E-05 | -2,4E-01 | -2,4E-01 | 6,0E-05 | -3,9E-01 | 1,5E-01 | -2,9 | 5,4E-03 | | | | |
| TRPP2 | -3,3E-01 | 3,1E-01 | 1,3E-09 | 7,9E-01 | 7,4E-03 | 5,2E-01 | -1,5E-02 | 5,2E-01 | -1,3E-01 | 1,9E-01 | 1,5E-02 | -7,5E-01 | -7,5E-01 | 1,5E-02 | -5,5E-01 | 1,6E-01 | -1,1 | 2,9E-01 | | | | |
| TRPP3 | -4,8E-01 | 4,3E-02 | 9,3E-07 | 1,5E-01 | 1,1E-01 | 4,4E-59 | -3,5E-01 | 4,4E-59 | -3,5E-01 | 1,5E-18 | 7,7E-01 | -3,1E-02 | -3,1E-02 | 7,7E-01 | -1,4E-01 | 9,2E-01 | -1,6 | 1,3E-01 | | | | |
| TRPP5 | -2,5E-01 | 3,7E-01 | 1,6E-02 | -6,6E-02 | 1,1E-02 | 1,0E-02 | -3,8E-02 | 1,0E-02 | 3,3E-02 | 1,7E-02 | 9,6E-03 | 1,1E-01 | 1,1E-01 | 9,6E-03 | 3,4E-02 | 9,3E-01 | -1 | 3,5E-01 | | | | |

Tabla B.54: Metaanálisis del conjunto de genes CP.

| Genes | Nuestras | | GSE8671 | | GSE39582 | | GSE44076 | | GSE41258 | | GSE33113 | | Metaanálisis | | Comb. p | |
|--------|----------|---------|----------|---------|----------|---------|----------|---------|----------|---------|----------|---------|--------------|---------|----------|---------|
| | logFC | FDR | logFC | FDR | logFC | FDR | logFC | FDR | logFC | FDR | logFC | FDR | Efecto | FDR | | Efecto |
| PMCA1 | 5,0E-02 | 9,4E-01 | -9,8E-01 | 7,1E-23 | -1,4 | 4,2E-29 | -1,2 | 1,4E-90 | -5,4E-01 | 1,2E-39 | -6,1E-01 | 2,3E-05 | -3,2 | 1,3E-05 | -9,6 | 0 |
| PMCA2 | -1,3E-01 | 6,7E-01 | -7,7E-02 | 5,1E-04 | 1,1E-03 | 9,8E-01 | -1,6E-02 | 1,6E-01 | 3,7E-03 | 8,3E-01 | 5,6E-03 | 9,0E-01 | -2,3E-01 | 2,3E-01 | -1,3 | 2,4E-01 |
| PMCA3 | 2,5E-01 | 3,4E-01 | -4,7E-02 | 5,3E-02 | -1,1E-01 | 2,9E-03 | -4,0E-02 | 7,8E-03 | 5,2E-04 | 9,7E-01 | 2,0E-02 | 7,1E-01 | -2,8E-01 | 1,8E-01 | -3,5 | 7,0E-04 |
| PMCA4 | -5,3E-01 | 5,9E-01 | -9,2E-02 | 7,6E-02 | 2,0E-01 | 9,3E-02 | -3,2E-02 | 2,7E-02 | -5,8E-01 | 1,7E-31 | -1,1 | 3,1E-13 | -1,2 | 6,5E-02 | -5,9 | 1,0E-08 |
| SERCA1 | 3,2E-01 | 3,7E-01 | 5,4E-03 | 9,1E-01 | 5,8E-02 | 4,4E-01 | -8,8E-03 | 7,9E-01 | 6,5E-02 | 3,7E-02 | -4,6E-02 | 5,8E-01 | 1,4E-01 | 2,6E-01 | 1,4 | 1,9E-01 |
| SERCA2 | 9,8E-01 | 1,2E-02 | -2,4E-01 | 1,2E-08 | -1,3E-01 | 6,3E-02 | -9,2E-02 | 1,2E-06 | -9,1E-02 | 1,6E-03 | -2,3E-01 | 6,5E-03 | -7,1E-01 | 3,0E-02 | -6,7 | 4,2E-11 |
| SERCA3 | 2,8E-01 | 6,3E-01 | -4,2E-01 | 1,6E-15 | -8,5E-01 | 2,5E-23 | -5,4E-01 | 4,0E-51 | -1 | 1,3E-76 | -6,0E-01 | 7,9E-09 | -2,9 | 7,8E-07 | -1,6E+01 | 0 |
| NCX1 | -4,4E-01 | 4,9E-01 | -3,1E-02 | 5,6E-02 | 5,5E-02 | 2,6E-02 | -1,2E-02 | 6,3E-01 | -2,3E-02 | 1,3E-01 | -6,0E-02 | 7,0E-02 | -2,0E-01 | 3,7E-01 | -3,8E-01 | 7,4E-01 |
| NCX2 | -4,5E-01 | 2,0E-01 | -4,6E-02 | 4,4E-01 | 9,0E-02 | 3,3E-01 | -1,7E-01 | 1,6E-27 | -1,6E-01 | 1,5E-05 | -2,9E-01 | 2,7E-02 | -7,6E-01 | 6,4E-02 | -4,2 | 4,6E-05 |
| NCX2 | 4,1E-02 | 9,2E-01 | -7,6E-02 | 8,2E-02 | 4,8E-02 | 4,1E-01 | -1,2E-01 | 2,8E-05 | -1,2E-01 | 1,6E-01 | 1,2E-01 | 1,6E-01 | -1,0E-01 | 7,6E-01 | -1,2 | 2,9E-01 |
| SPCA1 | 1,1 | 1,5E-02 | 4,6E-01 | 2,0E-17 | 3,8E-01 | 2,0E-06 | 4,0E-01 | 2,1E-55 | 2,1E-01 | 3,5E-18 | 3,6E-01 | 1,2E-03 | 2,2 | 2,5E-07 | 1,0E+01 | 0 |
| SPCA2 | 1,3 | 3,7E-02 | -1,4E-01 | 3,7E-03 | -7,0E-02 | 4,3E-01 | -2,2E-01 | 6,1E-13 | -4,9E-01 | 1,9E-18 | -1,5E-01 | 1,7E-01 | -6,9E-01 | 5,2E-02 | -7,8 | 3,1E-14 |

Tabla B.55: Metaanálisis del conjunto de genes CLC.

| Genes | Nuestras | | GSE8671 | | GSE39582 | | GSE44076 | | GSE41258 | | GSE33113 | | Metaanálisis | | Comb. p | |
|-------|----------|---------|----------|---------|----------|---------|----------|---------|----------|---------|----------|---------|--------------|---------|----------|---------|
| | logFC | FDR | logFC | FDR | logFC | FDR | logFC | FDR | logFC | FDR | logFC | FDR | Efecto | FDR | | Efecto |
| IP3R1 | -1 | 6,4E-02 | -4,0E-01 | 2,5E-15 | 1,2E-01 | 1,6E-01 | -6,8E-02 | 2,5E-02 | -3,0E-01 | 4,9E-32 | -5,0E-01 | 2,5E-05 | -1,5 | 2,3E-02 | -6,8 | 3,4E-11 |
| IP3R2 | -4,2E-02 | 9,0E-01 | -1,6E-01 | 1,4E-03 | 2,6E-01 | 8,0E-03 | 3,7E-01 | 6,5E-12 | 6,1E-02 | 2,6E-02 | 1,5E-02 | 9,2E-01 | 2,5E-01 | 5,0E-01 | 5,3 | 1,9E-07 |
| IP3R3 | 1,6 | 8,5E-03 | 7,8E-02 | 1,2E-02 | -8,5E-02 | 1,3E-01 | 2,9E-02 | 3,5E-01 | 6,7E-02 | 2,8E-02 | -2,9E-02 | 7,1E-01 | 3,0E-01 | 2,4E-01 | 1 | 3,5E-01 |
| RYR1 | 1,1E-01 | 6,7E-01 | -6,5E-02 | 3,7E-01 | 1,3E-01 | 3,0E-01 | 7,0E-03 | 7,1E-01 | -3,0E-02 | 2,8E-01 | -2,2E-01 | 9,6E-02 | -7,0E-02 | 6,7E-01 | -2,2E-01 | 8,5E-01 |
| RYR2 | -5,6E-01 | 1,9E-01 | 3,8E-02 | 2,9E-01 | 1,0E-01 | 2,0E-01 | 4,5E-02 | 2,9E-02 | -1,7E-02 | 3,5E-01 | -8,4E-02 | 6,0E-01 | 8,2E-02 | 6,8E-01 | 1,4 | 2,0E-01 |
| RYR3 | -3,3E-01 | 3,0E-01 | -9,3E-03 | 8,2E-01 | -3,8E-02 | 7,2E-01 | -2,6E-01 | 3,1E-08 | -4,1E-01 | 9,5E-24 | -1,1 | 4,2E-12 | -1,2 | 1,1E-02 | -8,4 | 0 |

Tabla B.56: Metaanálisis del conjunto de genes MCT.

| Genes | Nuestras | | GSE8671 | | GSE39582 | | GSE44076 | | GSE41258 | | GSE33113 | | Metaanálisis | | Comb. p | |
|-------|----------|---------|----------|---------|----------|---------|----------|---------|----------|---------|----------|---------|--------------|---------|---------|---------|
| | logFC | FDR | logFC | FDR | logFC | FDR | logFC | FDR | logFC | FDR | logFC | FDR | Efecto | FDR | | Efecto |
| MCU | 1,2 | 1,2E-02 | -1,3E-01 | 1,6E-02 | -1 | 4,0E-09 | -1,1 | 6,8E-63 | -1,2E-02 | 7,4E-01 | -4,1E-01 | 1,1E-01 | -1,1 | 2,2E-01 | -8,5 | 0 |
| MICU1 | 7,6E-01 | 3,2E-02 | 1,1E-01 | 6,9E-02 | -2,0E-01 | 8,7E-02 | -1,8E-01 | 2,2E-08 | -1,2E-02 | 7,4E-01 | 2,7E-01 | 1,7E-01 | 8,9E-02 | 8,2E-01 | -2,7 | 9,0E-03 |
| MICU2 | 9,7E-01 | 3,9E-03 | -1,6E-01 | 2,2E-02 | -5,4E-01 | 4,9E-03 | -3,0E-01 | 4,7E-09 | -3,0E-03 | 9,8E-01 | -2,0E-01 | 3,9E-01 | -3,1E-01 | 3,6E-01 | -5,1 | 8,1E-07 |
| MICU3 | -3,9E-01 | 3,1E-01 | -1,9E-01 | 1,4E-04 | 2,7E-02 | 8,6E-01 | -3,3E-01 | 2,6E-07 | 1,7E-01 | 5,0E-02 | -1,3 | 3,6E-04 | -8,3E-01 | 5,6E-03 | -4,7 | 5,2E-06 |
| MCURI | 1,2 | 9,6E-03 | -3,3E-01 | 7,0E-08 | -5,6E-01 | 4,0E-05 | 4,3E-02 | 5,1E-01 | 1,7E-01 | 5,0E-02 | 3,5E-01 | 4,3E-02 | -5,0E-02 | 9,3E-01 | -1,9 | 7,0E-02 |
| EMRE | -2,3E-01 | 5,4E-01 | -1,1E-01 | 1,3E-04 | -1,1E-01 | 7,3E-02 | -5,0E-01 | 2,7E-34 | 4,1E-01 | 7,5E-07 | -4,5E-01 | 4,2E-06 | -1,4 | 2,0E-03 | -6,6 | 8,9E-11 |
| MCUb | -2,9E-01 | 6,9E-01 | 7,9E-01 | 1,8E-09 | 6,4E-01 | 1,2E-02 | 4,1E-01 | 2,0E-13 | 4,1E-01 | 1,6E-10 | 6,5E-01 | 4,0E-02 | 1 | 1,2E-05 | 9,2 | 0 |
| VDAC1 | 1,2 | 4,8E-03 | 4,3E-01 | 2,8E-16 | -1,9E-01 | 1,3E-02 | 1,1E-01 | 1,5E-02 | 1,5E-01 | 1,6E-10 | 9,1E-02 | 4,5E-01 | 1,1 | 4,0E-02 | 4,1 | 6,7E-05 |
| VDAC2 | 6,4E-01 | 1,2E-01 | 2,8E-02 | 6,0E-01 | -5,6E-01 | 2,0E-06 | -3,9E-01 | 5,7E-33 | -4,4E-01 | 1,2E-12 | -1,2E-01 | 5,2E-01 | -7,6E-01 | 1,1E-01 | -9,7 | 0 |
| VDAC3 | 1,2 | 3,6E-02 | 2,3E-01 | 7,5E-10 | -1,3E-02 | 9,0E-01 | -5,6E-02 | 7,0E-02 | -1,0E-02 | 7,8E-01 | 1,8E-01 | 1,4E-01 | 5,9E-01 | 1,4E-01 | 1,1 | 3,2E-01 |
| NCLX | 8,9E-01 | 2,7E-02 | -1,3E-01 | 9,7E-04 | -6,3E-01 | 3,4E-14 | -3,1E-01 | 7,4E-50 | -6,5E-03 | 8,7E-01 | -1,7E-01 | 1,5E-01 | -9,8E-01 | 1,8E-01 | -9,8 | 0 |

B.4.2. Metaanálisis de enriquecimiento de genes

Siguiendo el mismo planteamiento que en los análisis de enriquecimiento realizados a lo largo de la tesis, se evaluó cuántos de los conjuntos de genes que muestran asociación con el fenotipo de CCR humano pertenecen a las bases de datos relacionadas directamente con el cáncer -MCC y MOG- (tabla 7.18) y/o contienen en su nombre alguna de las 6 palabras clave *calcium*, *mitochondria*, *reticulum*, *cancer*, *wnt* y *polyamine* (tablas de B.58 a B.63, respectivamente). Posteriormente, se muestran algunos de los grupos que han mostrado asociación con el fenotipo de cáncer de colon en el metaanálisis de enriquecimiento presentado en el apartado “7.2.3.2 Metaanálisis de enriquecimiento” . En concreto, se muestran, divididos en función de la base de datos a la que pertenecen y el método de análisis de enriquecimiento empleado, los resultados del análisis de enriquecimiento de aquellos conjuntos de genes con un FDR < 0.1 y que contienen alguna de las palabras de interés en su nombre (tablas B.64-B.74).

Tabla B.57: Total de conjuntos de genes que rechazan alguna de las hipótesis tras el reanálisis y metaanálisis de enriquecimiento de genes.

| Base de Datos | Método | n° de Conjuntos con un FDR < 0.05 (0.1) | | | | | | | |
|---------------|--------|---|------------|------------|------------|------------|------------|--------------|--|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis | |
| KEGG | CAMERA | 143(174) | 169(195) | 123(173) | 158(206) | 141(175) | 117(153) | 233(257) | |
| | ORA | 46(65) | 33(60) | 62(101) | 22(39) | 26(42) | 47(66) | 46(60) | |
| GOBP | CAMERA | 958(1449) | 1414(1935) | 1005(1520) | 1130(1628) | 818(1198) | 904(1383) | 1723(2128) | |
| | ORA | 452(794) | 248(532) | 366(674) | 55(174) | 137(402) | 393(737) | 121(210) | |
| GOMF | CAMERA | 323(461) | 449(619) | 347(518) | 407(584) | 294(411) | 282(437) | 604(735) | |
| | ORA | 144(228) | 70(158) | 109(210) | 14(52) | 47(127) | 107(210) | 24(48) | |
| MH | CAMERA | 26(31) | 28(32) | 21(27) | 23(27) | 21(25) | 18(24) | 33(40) | |
| | ORA | 13(15) | 9(13) | 11(14) | 9(12) | 11(12) | 13(16) | 13(22) | |
| MCDB | CAMERA | 2274(2906) | 2591(3239) | 1906(2562) | 2131(2797) | 1951(2581) | 1813(2452) | 3419(3885) | |
| | ORA | 1525(1945) | 990(1532) | 1202(1742) | 415(803) | 619(1119) | 1149(1787) | 1355(1911) | |
| MCC | CAMERA | 524(584) | 482(556) | 311(390) | 450(529) | 378(460) | 318(392) | 578(645) | |
| | ORA | 264(299) | 209(269) | 161(212) | 110(171) | 200(280) | 144(218) | 214(273) | |
| MOG | CAMERA | 85(103) | 93(111) | 52(67) | 61(91) | 60(84) | 52(74) | 106(125) | |
| | ORA | 35(44) | 22(36) | 40(47) | 19(29) | 21(42) | 27(38) | 37(46) | |
| MIM | CAMERA | 1897(2436) | 2220(2781) | 1073(1672) | 1596(2183) | 1346(1944) | 1263(1857) | 2419(2885) | |
| | ORA | 1891(2297) | 1816(2423) | 1060(1596) | 734(1314) | 596(1066) | 962(1414) | 1778(2361) | |
| Total | CAMERA | 6227(8141) | 7444(9466) | 4838(6927) | 5956(8045) | 5008(6876) | 4765(6770) | 9113(10697) | |
| | ORA | 4370(5687) | 3397(5023) | 3011(4596) | 1378(2594) | 1657(3090) | 2842(4486) | 3588(4931) | |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data
 Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Tabla B.58: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *calcium*.

| Base de Datos | Método | n° de Conjuntos con un FDR < 0.05 (0.1) | | | | | | | | | |
|---------------|--------|---|---------|----------|----------|----------|----------|--------------|--------|--------|--------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis | | | |
| KEGG | CAMERA | 2(2) | 2(2) | 1(2) | 2(2) | 2(2) | 2(2) | 2(2) | 2(2) | 2(2) | 2(2) |
| | ORA | 0(0) | 0(1) | 0(1) | 0(0) | 1(1) | 1(2) | 1(1) | 1(2) | 1(1) | 1(1) |
| GOBP | CAMERA | 28(37) | 31(37) | 20(27) | 23(29) | 19(22) | 22(28) | 30(34) | 22(28) | 30(34) | 30(34) |
| | ORA | 1(5) | 3(7) | 6(9) | 0(2) | 1(3) | 11(15) | 2(2) | 11(15) | 2(2) | 2(2) |
| GOMF | CAMERA | 8(10) | 7(10) | 3(6) | 4(8) | 7(9) | 5(6) | 11(13) | 5(6) | 11(13) | 11(13) |
| | ORA | 0(3) | 0(1) | 1(2) | 0(0) | 0(1) | 3(3) | 0(0) | 3(3) | 0(0) | 0(0) |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MCDB | CAMERA | 7(8) | 4(8) | 5(5) | 5(6) | 5(5) | 4(5) | 7(10) | 4(5) | 7(10) | 7(10) |
| | ORA | 1(1) | 0(1) | 1(1) | 0(0) | 0(0) | 2(3) | 0(0) | 2(3) | 0(0) | 0(0) |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| Total | CAMERA | 45(57) | 44(57) | 29(40) | 34(45) | 33(38) | 33(41) | 50(59) | 33(41) | 50(59) | 50(59) |
| | ORA | 2(9) | 3(10) | 8(13) | 0(2) | 2(5) | 17(23) | 3(3) | 17(23) | 3(3) | 3(3) |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Tabla B.59: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *cancer*.

| Base de Datos | Método | n° de Conjuntos con un FDR < 0.05 (0.1) | | | | | | | |
|---------------|--------|---|----------|----------|----------|----------|----------|--------------|--|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis | |
| KEGG | CAMERA | 0(3) | 4(6) | 1(6) | 1(3) | 1(3) | 1(1) | 5(8) | |
| | ORA | 0(2) | 0(0) | 4(5) | 0(0) | 0(2) | 1(1) | 1(4) | |
| GOBP | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| | ORA | 10(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| GOMF | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| MCDB | CAMERA | 214(267) | 212(261) | 168(217) | 187(241) | 202(252) | 172(226) | 291(327) | |
| | ORA | 143(184) | 92(133) | 117(165) | 44(86) | 78(126) | 122(167) | 146(196) | |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | |
| Total | CAMERA | 214(270) | 216(267) | 169(223) | 188(244) | 203(255) | 173(227) | 296(335) | |
| | ORA | 143(186) | 92(133) | 121(170) | 44(86) | 78(128) | 123(168) | 147(200) | |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Tabla B.60: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *WNT*.

| Base de Datos | Método | n° de Conjuntos con un FDR < 0.05 (0.1) | | | | | | | | | |
|---------------|--------|---|---------|----------|----------|----------|----------|--------------|------|--|--------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis | | | |
| KEGG | CAMERA | 0(0) | 0(0) | 0(1) | 0(1) | 0(0) | 0(0) | 0(0) | 0(1) | | 0(1) |
| | ORA | 0(1) | 1(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | 0(0) |
| GOBP | CAMERA | 3(5) | 5(6) | 3(3) | 5(6) | 2(3) | 3(5) | 5(7) | | | 5(7) |
| | ORA | 3(4) | 3(5) | 1(2) | 0(0) | 1(4) | 0(1) | 0(0) | | | 0(0) |
| GOMF | CAMERA | 0(0) | 1(1) | 0(1) | 0(0) | 0(1) | 0(2) | 0(0) | | | 0(0) |
| | ORA | 0(0) | 0(1) | 0(0) | 0(0) | 1(2) | 0(0) | 0(0) | | | 0(0) |
| MH | CAMERA | 1(1) | 0(0) | 1(1) | 1(1) | 0(1) | 0(0) | 1(1) | | | 1(1) |
| | ORA | 1(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| MCDB | CAMERA | 10(16) | 11(15) | 15(16) | 11(18) | 13(17) | 7(12) | 17(22) | | | 17(22) |
| | ORA | 11(17) | 13(16) | 2(4) | 2(6) | 1(5) | 4(5) | 4(9) | | | 4(9) |
| MCC | CAMERA | 0(1) | 0(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| MOG | CAMERA | 0(0) | 1(1) | 0(0) | 0(1) | 0(1) | 0(1) | 1(1) | | | 1(1) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| MIM | CAMERA | 0(3) | 1(1) | 0(0) | 0(0) | 0(1) | 0(1) | 2(2) | | | 2(2) |
| | ORA | 1(2) | 0(1) | 1(3) | 1(2) | 0(1) | 0(0) | 1(2) | | | 1(2) |
| Total | CAMERA | 14(26) | 19(25) | 19(22) | 17(27) | 15(24) | 10(21) | 26(34) | | | 26(34) |
| | ORA | 16(25) | 17(24) | 4(9) | 3(8) | 3(12) | 4(6) | 5(11) | | | 5(11) |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Tabla B.61: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *mitochondria*.

| Base de Datos | Método | n° de Conjuntos con un FDR < 0.05 (0.1) | | | | | | | | | |
|---------------|--------|---|---------|----------|----------|----------|----------|--------------|------|------|------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis | | | |
| KEGG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| GOBP | CAMERA | 16(29) | 24(29) | 21(29) | 20(29) | 16(19) | 19(24) | 32(36) | | | |
| | ORA | 14(27) | 7(19) | 5(8) | 1(5) | 1(6) | 4(8) | 0(2) | | | |
| GOMF | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| MCDB | CAMERA | 13(16) | 14(14) | 19(22) | 16(21) | 14(16) | 7(11) | 24(28) | | | |
| | ORA | 13(17) | 6(10) | 8(12) | 2(6) | 1(5) | 4(9) | 9(13) | | | |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | |
| Total | CAMERA | 29(45) | 38(43) | 40(51) | 36(50) | 30(35) | 26(35) | 56(64) | | | |
| | ORA | 27(44) | 13(29) | 13(20) | 3(11) | 2(11) | 8(17) | 9(15) | | | |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Tabla B.62: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *reticulum*.

| Base de Datos | Método | n° de Conjuntos con un FDR < 0.05(0.1) | | | | | | | | | |
|---------------|--------|--|---------|----------|----------|----------|----------|--------------|--|--|------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis | | | |
| KEGG | CAMERA | 1(1) | 0(0) | 0(0) | 1(1) | 0(1) | 0(1) | 1(1) | | | |
| | ORA | 1(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| GOBP | CAMERA | 9(15) | 3(7) | 5(6) | 6(8) | 4(5) | 7(9) | 8(12) | | | |
| | ORA | 6(10) | 2(3) | 2(3) | 1(2) | 1(2) | 4(5) | 1(1) | | | |
| GOMF | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| MCDB | CAMERA | 1(1) | 1(1) | 2(3) | 2(2) | 1(2) | 2(2) | 3(3) | | | |
| | ORA | 1(1) | 0(1) | 0(0) | 0(0) | 0(1) | 1(1) | 0(1) | | | |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | | | 0(0) |
| Total | CAMERA | 11(17) | 4(8) | 7(9) | 9(11) | 5(8) | 9(12) | 12(16) | | | |
| | ORA | 8(12) | 2(4) | 2(3) | 1(2) | 1(3) | 5(6) | 1(2) | | | |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Tabla B.63: Reanálisis y metaanálisis de conjuntos de genes y cuántos contienen la palabra de interés *polyamine*.

| Base de Datos | Método | n° de Conjuntos con un FDR < 0.05(0.1) | | | | | | | | | |
|---------------|--------|--|---------|----------|----------|----------|----------|--------------|------|------|------|
| | | Nuestras | GSE8671 | GSE39582 | GSE44076 | GSE41258 | GSE33113 | Metaanálisis | | | |
| KEGG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| GOBP | CAMERA | 0(0) | 1(1) | 1(1) | 1(1) | 2(2) | 2(2) | 2(2) | 2(2) | 2(2) | 2(2) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| GOMF | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 1(1) | 1(1) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MH | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MCDB | CAMERA | 1(1) | 1(1) | 1(1) | 1(1) | 1(1) | 1(1) | 1(1) | 1(1) | 1(1) | 1(1) |
| | ORA | 1(1) | 1(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(1) | 0(1) | 0(1) |
| MCC | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MOG | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| MIM | CAMERA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| | ORA | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) |
| Total | CAMERA | 1(1) | 2(2) | 2(2) | 2(2) | 3(3) | 4(4) | 3(3) | 4(4) | 3(3) | 3(3) |
| | ORA | 1(1) | 1(1) | 0(0) | 0(0) | 0(0) | 0(0) | 0(0) | 0(1) | 0(1) | 0(1) |

Abreviaturas:

GO: Gene Ontology; BP: Biological Process; MF: Molecular Function; MH: Msigdb Hallmark; MCDB: Msigdb Curated Data Base; MCC: Msigdb Computational Cancer; MOG: Msigdb Oncogenic; MIM: Msigdb Immunology.

Tabla B.64: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA).

| Conjuntos | pval.adj |
|---|-----------|
| GO:0010881 regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion | 0.0000000 |
| GO:0070126 mitochondrial translational termination | 0.0000000 |
| GO:0070125 mitochondrial translational elongation | 0.0000000 |
| GO:0016339 calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules | 0.0068841 |
| GO:0051480 regulation of cytosolic calcium ion concentration | 0.0000018 |
| GO:0051924 regulation of calcium ion transport | 0.0000436 |
| GO:0006890 retrograde vesicle-mediated transport, Golgi to endoplasmic reticulum | 0.0348416 |
| GO:0010880 regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum | 0.0000000 |
| GO:0010524 positive regulation of calcium ion transport into cytosol | 0.0000217 |
| GO:0032543 mitochondrial translation | 0.0000064 |
| GO:0006816 calcium ion transport | 0.0101074 |
| GO:0007204 positive regulation of cytosolic calcium ion concentration | 0.0000030 |
| GO:0019722 calcium-mediated signaling | 0.0000005 |
| GO:0051209 release of sequestered calcium ion into cytosol | 0.0052004 |
| GO:0006888 endoplasmic reticulum to Golgi vesicle-mediated transport | 0.0930906 |
| GO:0005513 detection of calcium ion | 0.0000000 |
| GO:0060071 Wnt signaling pathway, planar cell polarity pathway | 0.0000003 |
| GO:0060315 negative regulation of ryanodine-sensitive calcium-release channel activity | 0.0000000 |
| GO:0099509 regulation of presynaptic cytosolic calcium ion concentration | 0.0001886 |
| GO:0042776 mitochondrial ATP synthesis coupled proton transport | 0.0000001 |
| GO:0099566 regulation of postsynaptic cytosolic calcium ion concentration | 0.0916645 |
| GO:1901385 regulation of voltage-gated calcium channel activity | 0.0260892 |
| GO:0038018 Wnt receptor catabolic process | 0.0004504 |
| GO:0032981 mitochondrial respiratory chain complex I assembly | 0.0007712 |
| GO:0034975 protein folding in endoplasmic reticulum | 0.0057700 |

Tabla B.64: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------|
| GO:0050848 regulation of calcium-mediated signaling | 0.0002154 |
| GO:0060314 regulation of ryanodine-sensitive calcium-release channel activity | 0.0000030 |
| GO:0010882 regulation of cardiac muscle contraction by calcium ion signaling | 0.0000199 |
| GO:0051279 regulation of release of sequestered calcium ion into cytosol | 0.0087376 |
| GO:0090263 positive regulation of canonical Wnt signaling pathway | 0.0028597 |
| GO:0071786 endoplasmic reticulum tubular network organization | 0.0003846 |
| GO:1904380 endoplasmic reticulum mannose trimming | 0.0396948 |
| GO:0070588 calcium ion transmembrane transport | 0.0012255 |
| GO:0071277 cellular response to calcium ion | 0.0009234 |
| GO:0006120 mitochondrial electron transport, NADH to ubiquinone | 0.0000005 |
| GO:0060316 positive regulation of ryanodine-sensitive calcium-release channel activity | 0.0000046 |
| GO:0006122 mitochondrial electron transport, ubiquinol to cytochrome c | 0.0000002 |
| GO:0006391 transcription initiation from mitochondrial promoter | 0.0863989 |
| GO:0070131 positive regulation of mitochondrial translation | 0.0000143 |
| GO:0006123 mitochondrial electron transport, cytochrome c to oxygen | 0.0000001 |
| GO:0006851 mitochondrial calcium ion transmembrane transport | 0.0149835 |
| GO:0090090 negative regulation of canonical Wnt signaling pathway | 0.0389937 |
| GO:0007029 endoplasmic reticulum organization | 0.0510718 |
| GO:0070972 protein localization to endoplasmic reticulum | 0.0268461 |
| GO:0051481 negative regulation of cytosolic calcium ion concentration | 0.0021658 |
| GO:0032237 activation of store-operated calcium channel activity | 0.0012881 |
| GO:1900740 positive regulation of protein insertion into mitochondrial membrane involved in apoptotic signaling pathway | 0.0784717 |
| GO:0090646 mitochondrial tRNA processing | 0.0461401 |
| GO:0097250 mitochondrial respirasome assembly | 0.0257632 |

Tabla B.64: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------|
| GO:0016338 calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules | 0.0042418 |
| GO:0017156 calcium-ion regulated exocytosis | 0.0001090 |
| GO:0006264 mitochondrial DNA replication | 0.0002432 |
| GO:0070901 mitochondrial tRNA methylation | 0.0010300 |
| GO:0070129 regulation of mitochondrial translation | 0.0213896 |
| GO:0043504 mitochondrial DNA repair | 0.0237428 |
| GO:0045039 protein insertion into mitochondrial inner membrane | 0.0022072 |
| GO:0006874 cellular calcium ion homeostasis | 0.0000003 |
| GO:0043653 mitochondrial fragmentation involved in apoptotic process | 0.0481732 |
| GO:0032471 negative regulation of endoplasmic reticulum calcium ion concentration | 0.0630173 |
| GO:0016055 Wnt signaling pathway | 0.0812380 |
| GO:0051592 response to calcium ion | 0.0123834 |
| GO:0030150 protein import into mitochondrial matrix | 0.0000012 |
| GO:0061817 endoplasmic reticulum-plasma membrane tethering | 0.0629484 |
| GO:0051560 mitochondrial calcium ion homeostasis | 0.0920924 |
| GO:0061732 mitochondrial acetyl-CoA biosynthetic process from pyruvate | 0.0004296 |
| GO:0055074 calcium ion homeostasis | 0.0894362 |
| GO:0008053 mitochondrial fusion | 0.0140728 |
| GO:0032469 endoplasmic reticulum calcium ion homeostasis | 0.0001615 |
| GO:0060070 canonical Wnt signaling pathway | 0.0840150 |
| GO:0036444 calcium import into the mitochondrion | 0.0005674 |
| GO:0006595 polyamine metabolic process | 0.0001991 |
| GO:1902110 positive regulation of mitochondrial membrane permeability involved in apoptotic process | 0.0007130 |
| GO:1903573 negative regulation of response to endoplasmic reticulum stress | 0.0001339 |
| GO:0006596 polyamine biosynthetic process | 0.0469424 |
| GO:0034982 mitochondrial protein processing | 0.0480439 |
| GO:0033615 mitochondrial proton-transporting ATP synthase complex assembly | 0.0607783 |
| GO:2000051 negative regulation of non-canonical Wnt signaling pathway | 0.0039489 |

Tabla B.65: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOBP (ORA).

| Conjuntos | pval.adj |
|--|-----------|
| GO:0006123 mitochondrial electron transport, cytochrome c to oxygen | 0.0820680 |
| GO:0090201 negative regulation of release of cytochrome c from mitochondria | 0.0697452 |
| GO:0060315 negative regulation of ryanodine-sensitive calcium-release channel activity | 0.0311613 |
| GO:0010880 regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum | 0.0047555 |

Tabla B.66: Análisis de enriquecimiento de los conjuntos de genes de la base de datos GOMF (CAMERA).

| Conjuntos | pval.adj |
|--|-----------|
| GO:0005246 calcium channel regulator activity | 0.0000001 |
| GO:0005219 ryanodine-sensitive calcium-release channel activity | 0.0701328 |
| GO:0005245 voltage-gated calcium channel activity | 0.0657777 |
| GO:0015085 calcium ion transmembrane transporter activity | 0.0122135 |
| GO:0019855 calcium channel inhibitor activity | 0.0080350 |
| GO:0047498 calcium-dependent phospholipase A2 activity | 0.0217885 |
| GO:0005262 calcium channel activity | 0.0063465 |
| GO:0005229 intracellular calcium activated chloride channel activity | 0.0006080 |
| GO:0072345 NAADP-sensitive calcium-release channel activity | 0.0210934 |
| GO:0048306 calcium-dependent protein binding | 0.0486028 |
| GO:0004198 calcium-dependent cysteine-type endopeptidase activity | 0.0001181 |
| GO:0005388 calcium transmembrane transporter activity, phosphorylative mechanism | 0.0004698 |
| GO:0099604 ligand-gated calcium channel activity | 0.0083473 |

Tabla B.67: Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (CAMERA).

| Conjuntos | pval.adj |
|--|-----------|
| hsa04020 Calcium signaling pathway | 0.0000000 |
| hsa04141 Protein processing in endoplasmic reticulum | 0.0132470 |
| hsa04961 Endocrine and other factor-regulated calcium reabsorption | 0.0000000 |

Tabla B.67: Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------|
| hsa05235 PD-L1 expression and PD-1 checkpoint pathway in cancer | 0.0739022 |
| hsa05206 MicroRNAs in cancer | 0.0741508 |
| hsa05231 Choline metabolism in cancer | 0.0000341 |
| hsa05216 Thyroid cancer | 0.0897477 |
| hsa05222 Small cell lung cancer | 0.0328646 |
| hsa05205 Proteoglycans in cancer | 0.0438253 |
| hsa05200 Pathways in cancer | 0.0297704 |
| hsa05213 Endometrial cancer | 0.0285592 |
| hsa04310 Wnt signaling pathway | 0.0531075 |

Tabla B.68: Análisis de enriquecimiento de los conjuntos de genes de la base de datos KEGG (ORA).

| Conjuntos | pval.adj |
|--|-----------|
| hsa05222 Small cell lung cancer | 0.0900409 |
| hsa05219 Bladder cancer | 0.0501575 |
| hsa05215 Prostate cancer | 0.0942154 |
| hsa04961 Endocrine and other factor-regulated calcium reabsorption | 0.0395044 |
| hsa05206 MicroRNAs in cancer | 0.0002595 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA).

| Conjuntos | pval.adj |
|--|-----------|
| M15472 VECCHI GASTRIC CANCER EARLY DN | 0.0000000 |
| M8513 SMID BREAST CANCER NORMAL LIKE UP | 0.0000000 |
| M16740 GRADE COLON AND RECTAL CANCER UP | 0.0000000 |
| M18855 VECCHI GASTRIC CANCER EARLY UP | 0.0000000 |
| M3766 SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP | 0.0000000 |
| M15664 ROSTY CERVICAL CANCER PROLIFERATION CLUSTER | 0.0000000 |
| M10319 WALLACE PROSTATE CANCER RACE UP | 0.0000000 |
| M7854 WHITEFORD PEDIATRIC CANCER MARKERS | 0.0000000 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M8520 SHEDDEN LUNG CANCER POOR SURVIVAL A6 | 0.0000000 |
| M8365 RHODES UNDIFFERENTIATED CANCER | 0.0000000 |
| M4995 LINDGREN BLADDER CANCER CLUSTER 2B | 0.0000001 |
| M1345 RHODES CANCER META SIGNATURE | 0.0000000 |
| M13273 DELYS THYROID CANCER DN | 0.0000000 |
| M7517 SMID BREAST CANCER LUMINAL A UP | 0.0000000 |
| M7987 LEE LIVER CANCER SURVIVAL DN | 0.0000000 |
| M5792 LI AMPLIFIED IN LUNG CANCER | 0.0000000 |
| M10986 CHIANG LIVER CANCER SUBCLASS UNANNOTATED DN | 0.0000000 |
| M39678 WP RETINOBLASTOMA GENE IN CANCER | 0.0000000 |
| M18436 BOYAULT LIVER CANCER SUBCLASS G3 UP | 0.0000000 |
| M3268 CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP | 0.0000000 |
| M17333 FARMER BREAST CANCER CLUSTER 2 | 0.0000000 |
| M854 LINDGREN BLADDER CANCER CLUSTER 3 UP | 0.0000000 |
| M9826 VANTVEER BREAST CANCER METASTASIS DN | 0.0000000 |
| M11504 TOMLINS PROSTATE CANCER DN | 0.0000000 |
| M2043 RICKMAN HEAD AND NECK CANCER F | 0.0000000 |
| M19391 LIU PROSTATE CANCER DN | 0.0000000 |
| M7281 KONDO PROSTATE CANCER WITH H3K27ME3 | 0.0000000 |
| M17572 SMID BREAST CANCER LUMINAL B DN | 0.0000162 |
| M1311 WONG ENDMETRIUM CANCER DN | 0.0000000 |
| M336 VILLANUEVA LIVER CANCER KRT19 UP | 0.0000000 |
| M17423 LAIHO COLORECTAL CANCER SERRATED DN | 0.0000000 |
| M13072 SMID BREAST CANCER LUMINAL A DN | 0.0000000 |
| M27446 REACTOME MITOCHONDRIAL TRANSLATION | 0.0000000 |
| M19766 WU SILENCED BY METHYLATION IN BLADDER CANCER | 0.0000000 |
| M19982 WAMUNYOKOLI OVARIAN CANCER LMP UP | 0.0000000 |
| M15356 PUJANA BREAST CANCER LIT INT NETWORK | 0.0000000 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M15150 ZHANG BREAST CANCER PROGENITORS UP | 0.0000000 |
| M39740 WP METABOLIC REPROGRAMMING IN COLON CANCER | 0.0000001 |
| M12016 OUELLET OVARIAN CANCER INVASIVE VS LMP UP | 0.0000000 |
| M13831 BOYAULT LIVER CANCER SUBCLASS G123 UP | 0.0000000 |
| M2082 WANG METASTASIS OF BREAST CANCER ESR1 UP | 0.0000000 |
| M2180 PUJANA BREAST CANCER WITH BRCA1 MUTATED UP | 0.0000000 |
| M11629 MONTERO THYROID CANCER POOR SURVIVAL UP | 0.0000000 |
| M11616 ACEVEDO LIVER CANCER WITH H3K27ME3 UP | 0.0104472 |
| M39347 WP GASTRIC CANCER NETWORK 1 | 0.0000000 |
| M7995 HOSHIDA LIVER CANCER SUBCLASS S2 | 0.0000575 |
| M17471 SCHUETZ BREAST CANCER DUCTAL INVASIVE UP | 0.0001341 |
| M4572 LOCKWOOD AMPLIFIED IN LUNG CANCER | 0.0000006 |
| M9389 FINETTI BREAST CANCER KINOME RED | 0.0000000 |
| M2435 WAMUNYOKOLI OVARIAN CANCER GRADES 1 2 UP | 0.0000000 |
| M7396 SWEET LUNG CANCER KRAS DN | 0.0000000 |
| M8655 DAIRKEE CANCER PRONE RESPONSE BPA | 0.0000000 |
| M6862 NADERI BREAST CANCER PROGNOSIS UP | 0.0000000 |
| M18973 NIKOLSKY BREAST CANCER 20Q12 Q13 AMPLICON | 0.0000000 |
| M11302 SMID BREAST CANCER RELAPSE IN LUNG DN | 0.0000000 |
| M4448 BOYAULT LIVER CANCER SUBCLASS G23 UP | 0.0000000 |
| M2890 KEGG CALCIUM SIGNALING PATHWAY | 0.0000000 |
| M2427 NAKAMURA CANCER MICROENVIRONMENT DN | 0.0000000 |
| M13736 PYEON CANCER HEAD AND NECK VS CERVICAL UP | 0.0000060 |
| M1687 ZHONG SECRETOME OF LUNG CANCER AND FIBROBLAST | 0.0228940 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------|
| M707 WAMUNYOKOLI OVARIAN CANCER GRADES 1 2 DN | 0.0002490 |
| M14228 RICKMAN HEAD AND NECK CANCER A | 0.0801470 |
| M15780 GRADE COLON AND RECTAL CANCER DN | 0.0000000 |
| M15112 WONG MITOCHONDRIA GENE MODULE | 0.0000125 |
| M19837 RICKMAN HEAD AND NECK CANCER E | 0.0000000 |
| M27099 REACTOME TCF DEPENDENT SIGNALING IN RESPONSE TO WNT | 0.0089551 |
| M16542 YAMASHITA LIVER CANCER WITH EPCAM UP | 0.0000737 |
| M15015 FARMER BREAST CANCER CLUSTER 1 | 0.0000000 |
| M3710 WALLACE PROSTATE CANCER UP | 0.0000963 |
| M6754 ACEVEDO LIVER CANCER WITH H3K9ME3 UP | 0.0007650 |
| M17094 GINESTIER BREAST CANCER ZNF217 AMPLIFIED DN | 0.0000001 |
| M9577 MOOTHA MITOCHONDRIA | 0.0001762 |
| M1685 ZHONG SECRETOME OF LUNG CANCER AND MACROPHAGE | 0.0019495 |
| M416 ANDERSEN LIVER CANCER KRT19 UP | 0.0000084 |
| M1260 WATANABE RECTAL CANCER RADIOTHERAPY RESPONSIVE UP | 0.0587675 |
| M7727 RICKMAN HEAD AND NECK CANCER B | 0.0499493 |
| M10091 WAMUNYOKOLI OVARIAN CANCER LMP DN | 0.0000365 |
| M747 REACTOME METABOLISM OF POLYAMINES | 0.0000000 |
| M17299 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL UP | 0.0000000 |
| M1255 INAMURA LUNG CANCER SCC DN | 0.0000569 |
| M19148 TOMLINS PROSTATE CANCER UP | 0.0002671 |
| M10382 IIZUKA LIVER CANCER EARLY RECURRENCE | 0.0021526 |
| M39497 WP GASTRIC CANCER NETWORK 2 | 0.0000000 |
| M16374 IIZUKA LIVER CANCER PROGRESSION G1 G2 DN | 0.0252893 |
| M14670 DING LUNG CANCER EXPRESSION BY COPY NUMBER | 0.0020671 |
| M378 MINGUEZ LIVER CANCER VASCULAR INVASION UP | 0.0000082 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M39329 WP CALCIUM REGULATION IN THE CARDIAC CELL | 0.0000000 |
| M18761 SMITH LIVER CANCER | 0.0037338 |
| M7547 JIANG HYPOXIA CANCER | 0.0014851 |
| M4619 LIU PROSTATE CANCER UP | 0.0414439 |
| M12795 CHARAFE BREAST CANCER BASAL VS MESENCHYMAL UP | 0.0000181 |
| M10837 NIKOLSKY BREAST CANCER 20Q11 AMPLICON | 0.0000006 |
| M8526 CHEN LUNG CANCER SURVIVAL | 0.0345582 |
| M39628 WP INTEGRATED CANCER PATHWAY | 0.0000003 |
| M8779 LI LUNG CANCER | 0.0001359 |
| M16932 CHIANG LIVER CANCER SUBCLASS PROLIFERATION DN | 0.0000000 |
| M19988 GINESTIER BREAST CANCER 20Q13 AMPLIFICATION DN | 0.0005248 |
| M13007 GLINSKY CANCER DEATH UP | 0.0020695 |
| M38998 REACTOME ABERRANT REGULATION OF MITOTIC G1 S TRANSITION IN CANCER DUE TO RB1 DEFECTS | 0.0000000 |
| M1758 SANSOM WNT PATHWAY REQUIRE MYC | 0.0000000 |
| M924 REACTOME PLATELET CALCIUM HOMEOSTASIS | 0.0000000 |
| M867 MEINHOLD OVARIAN CANCER LOW GRADE DN | 0.0000000 |
| M672 CHIN BREAST CANCER COPY NUMBER UP | 0.0000000 |
| M15096 REACTOME MITOCHONDRIAL TRNA AMINOACYLATION | 0.0000002 |
| M15147 BOYALT LIVER CANCER SUBCLASS G12 UP | 0.0026592 |
| M9565 AGUIRRE PANCREATIC CANCER COPY NUMBER UP | 0.0087201 |
| M1253 INAMURA LUNG CANCER SCC UP | 0.0000023 |
| M15425 VANTVEER BREAST CANCER ESR1 DN | 0.0000009 |
| M2837 BARIS THYROID CANCER UP | 0.0883620 |
| M1273 SHETH LIVER CANCER VS TXNIP LOSS PAM1 | 0.0078525 |
| M11288 SCHLESINGER METHYLATED DE NOVO IN CANCER | 0.0000004 |
| M590 REACTOME MITOCHONDRIAL PROTEIN IMPORT | 0.0004258 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------|
| M12533 WALLACE PROSTATE CANCER DN | 0.0059872 |
| M38983 MEBARKI HCC PROGENITOR WNT DN | 0.0000000 |
| M11896 LINDGREN BLADDER CANCER CLUSTER 2A DN | 0.0046260 |
| M1256 INAMURA LUNG CANCER SCC SUBTYPES UP | 0.0000294 |
| M2429 SHEDDEN LUNG CANCER GOOD SURVIVAL A12 | 0.0000243 |
| M7371 LOPES METHYLATED IN COLON CANCER UP | 0.0637976 |
| M13462 LUI THYROID CANCER CLUSTER 5 | 0.0000000 |
| M38989 MEBARKI HCC PROGENITOR WNT DN BLOCKED BY FZD8CRD | 0.0000000 |
| M15807 BONOME OVARIAN CANCER POOR SURVIVAL DN | 0.0000005 |
| M6875 LANDIS BREAST CANCER PROGRESSION UP | 0.0001471 |
| M3102 SMID BREAST CANCER ERBB2 UP | 0.0130572 |
| M12895 CHARAFE BREAST CANCER BASAL VS MESENCHYMAL DN | 0.0101824 |
| M4333 LUI THYROID CANCER CLUSTER 4 | 0.0049959 |
| M17788 LINDGREN BLADDER CANCER CLUSTER 1 DN | 0.0000000 |
| M1686 ZHONG SECRETOME OF LUNG CANCER AND ENDOTHELIUM | 0.0104132 |
| M18694 LIU COMMON CANCER GENES | 0.0975198 |
| M9192 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL DN | 0.0035940 |
| M1362 SCHAEFFER PROSTATE DEVELOPMENT AND CANCER BOX5 UP | 0.0000016 |
| M18086 DAIRKEE CANCER PRONE RESPONSE BPA E2 | 0.0010623 |
| M39360 WP INTEGRATED BREAST CANCER PATHWAY | 0.0177377 |
| M39417 WP ELECTRON TRANSPORT CHAIN OXPHOS SYSTEM IN MITOCHONDRIA | 0.0000000 |
| M17372 LEE LIVER CANCER MYC TGFA DN | 0.0000403 |
| M12596 SETLUR PROSTATE CANCER TMRSS2 ERG FUSION DN | 0.0000094 |
| M15193 GRUETZMANN PANCREATIC CANCER UP | 0.0006611 |
| M9224 SAKAI CHRONIC HEPATITIS VS LIVER CANCER UP | 0.0034942 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M19164 RICKMAN HEAD AND NECK CANCER D | 0.0000001 |
| M9908 LANDIS BREAST CANCER PROGRESSION DN | 0.0000000 |
| M12824 SHEDDEN LUNG CANCER GOOD SURVIVAL A4 | 0.0000000 |
| M26973 REACTOME MITOCHONDRIAL BIOGENESIS | 0.0008117 |
| M8817 ACEVEDO LIVER CANCER WITH H3K9ME3 DN | 0.0185553 |
| M9914 YAMASHITA LIVER CANCER WITH EPCAM DN | 0.0889778 |
| M19675 FARMER BREAST CANCER CLUSTER 3 | 0.0781418 |
| M15411 BARIS THYROID CANCER DN | 0.0103911 |
| M2202 JAZAERI BREAST CANCER BRCA1 VS BRCA2 UP | 0.0060667 |
| M12461 FINAK BREAST CANCER SDPP SIGNATURE | 0.0025709 |
| M17936 NIKOLSKY BREAST CANCER 8Q23 Q24 AMPLICON | 0.0000000 |
| M19508 HATADA METHYLATED IN LUNG CANCER UP | 0.0011424 |
| M9940 HONRADO BREAST CANCER BRCA1 VS BRCA2 | 0.0017477 |
| M816 BONOME OVARIAN CANCER POOR SURVIVAL UP | 0.0000000 |
| M6046 BARRIER CANCER RELAPSE TUMOR SAMPLE UP | 0.0000022 |
| M5369 POOLA INVASIVE BREAST CANCER UP | 0.0000047 |
| M8689 CHIANG LIVER CANCER SUBCLASS CTNNB1 DN | 0.0141562 |
| M8760 MCGARVEY SILENCED BY METHYLATION IN COLON CANCER | 0.0491188 |
| M1177 KOINUMA COLON CANCER MSI DN | 0.0182911 |
| M39892 WP MITOCHONDRIAL CIII ASSEMBLY | 0.0000734 |
| M398 MINGUEZ LIVER CANCER VASCULAR INVASION DN | 0.0001458 |
| M9206 YAMASHITA LIVER CANCER STEM CELL DN | 0.0000000 |
| M9728 BARRIER COLON CANCER RECURRENCE UP | 0.0003182 |
| M39731 WP LNCRNA INVOLVEMENT IN CANONICAL WNT SIGNALING AND COLORECTAL CANCER | 0.0000078 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------------|
| M5542 SATO SILENCED EPIGENETICALLY IN PANCREATIC CANCER | 0.0007483 |
| M1065 MIDORIKAWA AMPLIFIED IN LIVER CANCER | 0.0039269 |
| M10922 WANG RECURRENT LIVER CANCER UP | 0.0000318 |
| M38985 MEBARKI HCC PROGENITOR WNT DN CTNNB1 DEPENDENT | 0.0000000 |
| M4960 SMID BREAST CANCER BASAL DN | 0.0452009 |
| M27163 REACTOME CONSTITUTIVE SIGNALING BY ABERRANT PI3K IN CANCER | 0.0048105 |
| M1293 YANG BREAST CANCER ESR1 DN | 0.0432441 |
| M40043 WP MITOCHONDRIAL FATTY ACID SYNTHESIS PATHWAY | 0.0094324 |
| M2742 LEE LIVER CANCER MYC UP | 0.0018134 |
| M15662 WATANABE ULCERATIVE COLITIS WITH CANCER DN | 0.0662342 |
| M18394 ROYLANCE BREAST CANCER 16Q COPY NUMBER UP | 0.0000753 |
| M1364 SCHAEFFER PROSTATE DEVELOPMENT AND CANCER BOX6 UP | 0.0210552 |
| M9464 WONG ENDOMETRIAL CANCER LATE | 0.0105113 |
| M4342 BOYALT LIVER CANCER SUBCLASS G6 UP | 0.0000004 |
| M4322 AUNG GASTRIC CANCER | 0.0046842 |
| M4988 JAERVINEN AMPLIFIED IN LARYNGEAL CANCER | 0.0000643 |
| M3645 DELYS THYROID CANCER UP | 0.0063841 |
| M4381 LUI THYROID CANCER PAX8 PPARG UP | 0.0645771 |
| M39015 REACTOME ABERRANT REGULATION OF MITOTIC EXIT IN CANCER DUE TO RB1 DEFECTS | 0.0000000 |
| M39781 WP MITOCHONDRIAL COMPLEX I ASSEMBLY MODEL OXPHOS SYSTEM | 0.0011968 |
| M14132 SLEBOS HEAD AND NECK CANCER WITH HPV UP | 0.0000001 |
| M235 PID TCR CALCIUM PATHWAY | 0.0095237 |
| M9538 GRADE COLON VS RECTAL CANCER UP | 0.0000000 |
| M14383 SMID BREAST CANCER RELAPSE IN BRAIN UP | 0.0006583 |
| M11507 IIZUKA LIVER CANCER PROGRESSION L1 G1 UP | 0.0000000 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------|
| M10605 BERTUCCI MEDULLARY VS DUCTAL BREAST CANCER DN | 0.0000073 |
| M4090 BOYALT LIVER CANCER SUBCLASS G3 DN | 0.0000000 |
| M39619 WP SENESENCE AND AUTOPHAGY IN CANCER | 0.0176233 |
| M1595 BARRIER CANCER RELAPSE NORMAL SAMPLE UP | 0.0000000 |
| M796 REACTOME ATF4 ACTIVATES GENES IN RESPONSE TO ENDOPLASMIC RETICULUM STRESS | 0.0000285 |
| M6362 NAKAMURA LUNG CANCER | 0.0000072 |
| M18496 LEE LIVER CANCER MYC DN | 0.0002842 |
| M2096 ABDULRAHMAN KIDNEY CANCER VHL DN | 0.0088584 |
| M2899 LINDGREN BLADDER CANCER HIGH RECURRENCE | 0.0002131 |
| M1805 WILLERT WNT SIGNALING | 0.0899743 |
| M10336 OUELLET OVARIAN CANCER INVASIVE VS LMP DN | 0.0001664 |
| M2631 FARMER BREAST CANCER APOCRINE VS BASAL | 0.0162069 |
| M15473 SAKAI CHRONIC HEPATITIS VS LIVER CANCER DN | 0.0913731 |
| M38982 MEBARKI HCC PROGENITOR WNT UP | 0.0000198 |
| M27336 REACTOME SODIUM CALCIUM EXCHANGERS | 0.0019406 |
| M2840 CHARAFE BREAST CANCER LUMINAL VS BASAL UP | 0.0356616 |
| M39881 WP MITOCHONDRIAL CIV ASSEMBLY | 0.0000003 |
| M5483 YANG BREAST CANCER ESR1 BULK DN | 0.0021764 |
| M12307 NIKOLSKY BREAST CANCER 16Q24 AMPLICON | 0.0000125 |
| M13656 NAKAMURA LUNG CANCER DIFFERENTIATION MARKERS | 0.0325822 |
| M6041 DAIRKEE CANCER PRONE RESPONSE E2 | 0.0051442 |
| M27816 REACTOME MITOCHONDRIAL CALCIUM ION TRANSPORT | 0.0089633 |
| M39795 WP PKCGAMMA CALCIUM SIGNALING PATHWAY IN ATAXIA | 0.0645620 |
| M15346 LEE LIVER CANCER E2F1 DN | 0.0000457 |
| M27100 REACTOME WNT MEDIATED ACTIVATION OF DVL | 0.0537267 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------------|
| M14690 REACTOME MITOCHONDRIAL FATTY ACID BETA OXIDATION | 0.0000000 |
| M1352 NIKOLSKY BREAST CANCER 15Q26 AMPLICON | 0.0005223 |
| M39407 WP WNTBETACATENIN SIGNALING PATHWAY IN LEUKEMIA | 0.0083028 |
| M39898 WP MITOCHONDRIAL CII ASSEMBLY | 0.0132912 |
| M6145 LEE LIVER CANCER SURVIVAL UP | 0.0000000 |
| M15125 FARMER BREAST CANCER CLUSTER 4 | 0.0000000 |
| M14353 CHIANG LIVER CANCER SUBCLASS INTERFERON DN | 0.0366152 |
| M16130 RICKMAN HEAD AND NECK CANCER C | 0.0009558 |
| M17535 CAMPS COLON CANCER COPY NUMBER DN | 0.0005371 |
| M5126 PYEON CANCER HEAD AND NECK VS CERVICAL DN | 0.0682330 |
| M7870 HEIDENBLAD AMPLIFIED IN BONE CANCER | 0.0198188 |
| M14693 VANTVEER BREAST CANCER POOR PROGNOSIS | 0.0000000 |
| M38988 MEBARKI HCC PROGENITOR WNT UP BLOCKED BY FZD8CRD | 0.0001397 |
| M26913 REACTOME PRESYNAPTIC DEPOLARIZATION AND CALCIUM CHANNEL OPENING | 0.0994565 |
| M1096 KOINUMA COLON CANCER MSI UP | 0.0001512 |
| M6767 POOLA INVASIVE BREAST CANCER DN | 0.0001886 |
| M17079 VECCHI GASTRIC CANCER ADVANCED VS EARLY UP | 0.0000000 |
| M27162 REACTOME PI3K AKT SIGNALING IN CANCER | 0.0066484 |
| M376 PUJANA BREAST CANCER WITH BRCA1 MUTATED DN | 0.0001572 |
| M38991 MEBARKI HCC PROGENITOR WNT DN CTNNB1 DEPENDENT BLOCKED BY FZD8CRD | 0.0000000 |
| M1218 BORLAK LIVER CANCER EGF UP | 0.0141526 |
| M27288 REACTOME BETA CATENIN INDEPENDENT WNT SIGNALING | 0.0613295 |
| M4420 LEE LIVER CANCER MYC E2F1 UP | 0.0009558 |
| M15884 NIKOLSKY BREAST CANCER 7P22 AMPLICON | 0.0000204 |
| M2195 LEE LIVER CANCER CIPROFIBRATE UP | 0.0031300 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M14539 BERTUCCI MEDULLARY VS DUCTAL BREAST CANCER UP | 0.0000040 |
| M1280 SHETH LIVER CANCER VS TXNIP LOSS PAM4 | 0.0000000 |
| M39563 WP NCRNAS INVOLVED IN WNT SIGNALING IN HEPATOCELLULAR CARCINOMA | 0.0020746 |
| M12210 YOKOE CANCER TESTIS ANTIGENS | 0.0000000 |
| M14702 KIM GASTRIC CANCER CHEMOSENSITIVITY | 0.0591837 |
| M2084 WANG METASTASIS OF BREAST CANCER | 0.0209549 |
| M14414 BOYALT LIVER CANCER SUBCLASS G6 DN | 0.0075225 |
| M13954 NIKOLSKY BREAST CANCER 8Q12 Q22 AMPLICON | 0.0009654 |
| M8077 SOTIRIOU BREAST CANCER GRADE 1 VS 3 DN | 0.0021655 |
| M5755 KAPOSI LIVER CANCER MET UP | 0.0146326 |
| M450 KIM LIVER CANCER POOR SURVIVAL UP | 0.0975824 |
| M4752 WATANABE RECTAL CANCER RADIOTHERAPY RESPONSIVE DN | 0.0006607 |
| M424 ANDERSEN LIVER CANCER KRT19 DN | 0.0000000 |
| M1286 HOSHIDA LIVER CANCER SUBCLASS S3 | 0.0000000 |
| M27203 REACTOME ROLE OF LAT2 NTAL LAB ON CALCIUM MOBILIZATION | 0.0098644 |
| M4671 BARRIER CANCER RELAPSE NORMAL SAMPLE DN | 0.0484428 |
| M11078 YANG BREAST CANCER ESR1 LASER DN | 0.0017724 |
| M1170 HUMMERICH SKIN CANCER PROGRESSION DN | 0.0000248 |
| M2218 BOYALT LIVER CANCER SUBCLASS G123 DN | 0.0000000 |
| M8570 DOANE BREAST CANCER ESR1 DN | 0.0124425 |
| M8629 NADERI BREAST CANCER PROGNOSIS DN | 0.0317623 |
| M5636 LEE LIVER CANCER MYC E2F1 DN | 0.0037611 |
| M16201 WONG ENDMETRIUM CANCER UP | 0.0834843 |
| M3228 KEGG SMALL CELL LUNG CANCER | 0.0545943 |
| M39378 WP BMP2WNT4FOXO1 PATHWAY IN HUMAN PRIMARY ENDOMETRIAL STROMAL CELL DIFFERENTIATION | 0.0000019 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M1424 LEE LIVER CANCER DENA DN | 0.0000002 |
| M7581 NAKAMURA CANCER MICROENVIRONMENT UP | 0.0823121 |
| M1240 JOHANSSON BRAIN CANCER EARLY VS LATE DN | 0.0000003 |
| M13709 NIKOLSKY BREAST CANCER 12Q13 Q21 AMPLICON | 0.0185100 |
| M16257 BIOCARTA MITOCHONDRIA PATHWAY | 0.0107194 |
| M39425 WP MIRNA REGULATION OF P53 PATHWAY IN PROSTATE CANCER | 0.0001330 |
| M40059 WP NAD METABOLISM IN ONCOGENEINDUCED SENESENCE AND MITOCHONDRIAL DYSFUNCTIONASSOCIATED SENESENCE | 0.0193711 |
| M2135 LUI THYROID CANCER CLUSTER 1 | 0.0991569 |
| M5311 HOSHIDA LIVER CANCER SUBCLASS S1 | 0.0350530 |
| M8941 HEDENFALK BREAST CANCER BRACX DN | 0.0228848 |
| M3238 SMID BREAST CANCER RELAPSE IN BONE UP | 0.0000111 |
| M7012 SCHUETZ BREAST CANCER DUCTAL INVASIVE DN | 0.0263194 |
| M9693 FINETTI BREAST CANCERS KINOME GRAY | 0.0117749 |
| M8475 WANG ESOPHAGUS CANCER VS NORMAL UP | 0.0014574 |
| M10953 BOYALT LIVER CANCER SUBCLASS G2 | 0.0136456 |
| M19661 WANG ESOPHAGUS CANCER PROGRESSION UP | 0.0000002 |
| M7577 ACEVEDO LIVER CANCER DN | 0.0000044 |
| M2083 WANG METASTASIS OF BREAST CANCER ESR1 DN | 0.0508752 |
| M2572 ANASTASSIOU MULTICANCER INVASIVENESS SIGNATURE | 0.0000000 |
| M7343 LUI THYROID CANCER CLUSTER 2 | 0.0117024 |
| M77 PID WNT SIGNALING PATHWAY | 0.0203734 |
| M27745 REACTOME SIGNALING BY FGFR3 FUSIONS IN CANCER | 0.0005087 |
| M39882 WP MAPK PATHWAY IN CONGENITAL THYROID CANCER | 0.0345582 |
| M29623 REACTOME LOSS OF FUNCTION OF SMAD2 3 IN CANCER | 0.0942280 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------|
| M13354 KONDO COLON CANCER HCP WITH H3K27ME1 | 0.0535027 |
| M4008 FARMER BREAST CANCER APOCRINE VS LUMINAL | 0.0232680 |
| M12892 SMID BREAST CANCER LUMINAL B UP | 0.0042870 |
| M1358 SCHAEFFER PROSTATE DEVELOPMENT AND CANCER BOX4 UP | 0.0726094 |
| M38984 MEBARKI HCC PROGENITOR WNT UP CTNNB1 DEPENDENT | 0.0016900 |
| M9951 WANG METHYLATED IN BREAST CANCER | 0.0097054 |
| M9246 ENGELMANN CANCER PROGENITORS DN | 0.0000032 |
| M9096 IIZUKA LIVER CANCER PROGRESSION G2 G3 UP | 0.0000145 |
| M12228 BOYALT LIVER CANCER SUBCLASS G12 DN | 0.0000502 |
| M16737 VANTVEER BREAST CANCER BRCA1 UP | 0.0000111 |
| M10431 GRUETZMANN PANCREATIC CANCER DN | 0.0000030 |
| M38990 MEBARKI HCC PROGENITOR WNT UP CTNNB1 DEPENDENT BLOCKED BY FZD8CRD | 0.0613109 |
| M39683 WP MIRNA REGULATION OF PROSTATE CANCER SIGNALING PATHWAYS | 0.0101179 |
| M39669 WP WNT SIGNALING | 0.0611979 |
| M821 SCHLESINGER METHYLATED IN COLON CANCER | 0.0093250 |
| M13077 WANG RECURRENT LIVER CANCER DN | 0.0025210 |
| M19877 KEGG ENDOMETRIAL CANCER | 0.0109534 |
| M11238 LINDGREN BLADDER CANCER CLUSTER 3 DN | 0.0761888 |
| M373 VILLANUEVA LIVER CANCER KRT19 DN | 0.0000000 |
| M12602 WOO LIVER CANCER RECURRENCE UP | 0.0819365 |
| M27246 REACTOME SIGNALING BY TGF BETA RECEPTOR COMPLEX IN CANCER | 0.0482780 |
| M19439 DOANE BREAST CANCER ESR1 UP | 0.0000580 |
| M5736 BARRIER COLON CANCER RECURRENCE DN | 0.0000000 |
| M15626 OUYANG PROSTATE CANCER PROGRESSION UP | 0.0066485 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------|
| M39391 WP DEREGULATION OF RAB AND RAB EFFECTOR GENES IN BLADDER CANCER | 0.0000038 |
| M2700 SMID BREAST CANCER RELAPSE IN BONE DN | 0.0216218 |
| M27821 REACTOME RUNX3 REGULATES WNT SIGNALING | 0.0086045 |
| M29535 REACTOME SIGNALING BY ERBB2 IN CANCER | 0.0453163 |
| M18658 WIKMAN ASBESTOS LUNG CANCER UP | 0.0241252 |
| M27719 REACTOME MITOCHONDRIAL FATTY ACID BETA OXIDATION OF SATURATED FATTY ACIDS | 0.0000000 |
| M682 FARMER BREAST CANCER CLUSTER 5 | 0.0001660 |
| M13123 LEE LIVER CANCER HEPATOBLAST | 0.0979369 |
| M10748 YANG BREAST CANCER ESRI BULK UP | 0.0048701 |
| M39365 WP MITOCHONDRIAL LCFATTY ACID BETAOXIDATION | 0.0000000 |
| M8432 NIKOLSKY BREAST CANCER 12Q24 AMPLICON | 0.0680459 |
| M11640 LEE LIVER CANCER CIPROFIBRATE DN | 0.0603310 |
| M18299 YANG BREAST CANCER ESRI UP | 0.0139540 |
| M5827 GINESTIER BREAST CANCER ZNF217 AMPLIFIED UP | 0.0105612 |
| M19612 LEE LIVER CANCER ACOX1 UP | 0.0208243 |
| M13910 CLIMENT BREAST CANCER COPY NUMBER UP | 0.0697927 |
| M6972 MONTERO THYROID CANCER POOR SURVIVAL DN | 0.0001344 |
| M26946 REACTOME ACTIVATION OF PUMA AND TRANSLOCATION TO MITOCHONDRIA | 0.0730149 |
| M18536 FALVELLA SMOKERS WITH LUNG CANCER | 0.0423278 |
| M27202 REACTOME SIGNALING BY NOTCH1 HD DOMAIN MUTANTS IN CANCER | 0.0991777 |
| M19818 KEGG NON SMALL CELL LUNG CANCER | 0.0215113 |
| M40066 WP SARSCOV2 MITOCHONDRIAL INTERACTIONS | 0.0117453 |
| M12889 LINDGREN BLADDER CANCER CLUSTER 2A UP | 0.0702388 |
| M13658 HOSHIDA LIVER CANCER LATE RECURRENCE DN | 0.0002027 |
| M19779 FARMER BREAST CANCER CLUSTER 6 | 0.0001666 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------|
| M2108 VECCHI GASTRIC CANCER ADVANCED VS EARLY DN | 0.0000000 |
| M15456 MCCABE HOXC6 TARGETS CANCER UP | 0.0098344 |
| M38986 MEBARKI HCC PROGENITOR WNT UP CTNNB1 INDEPENDENT | 0.0000078 |
| M244 SMID BREAST CANCER RELAPSE IN BRAIN DN | 0.0000400 |
| M9719 CAVARD LIVER CANCER MALIGNANT VS BENIGN | 0.0050750 |
| M3003 MEINHOLD OVARIAN CANCER LOW GRADE UP | 0.0002539 |
| M5451 HOSHIDA LIVER CANCER SURVIVAL DN | 0.0026422 |
| M15351 SUZUKI AMPLIFIED IN ORAL CANCER | 0.0673143 |
| M834 CHIANG LIVER CANCER SUBCLASS POLYSOMY7 UP | 0.0239217 |
| M10541 YANG BREAST CANCER ESRI LASER UP | 0.0000073 |
| M11771 DOANE BREAST CANCER CLASSES DN | 0.0303789 |
| M2118 SMID BREAST CANCER RELAPSE IN PLEURA DN | 0.0973278 |
| M39901 WP ENDOPLASMIC RETICULUM STRESS RESPONSE IN CORONAVIRUS INFECTION | 0.0067729 |
| M534 KIM LIVER CANCER POOR SURVIVAL DN | 0.0000388 |
| M17427 DOANE BREAST CANCER CLASSES UP | 0.0000000 |
| M19610 CHIANG LIVER CANCER SUBCLASS UNANNOTATED UP | 0.0000007 |
| M4707 IIZUKA LIVER CANCER PROGRESSION G2 G3 DN | 0.0713856 |
| M2862 FINETTI BREAST CANCERS KINOME BLUE | 0.0350796 |
| M13766 NIKOLSKY BREAST CANCER 5P15 AMPLICON | 0.0508293 |
| M17044 WANG HCP PROSTATE CANCER | 0.0926271 |
| M1788 NIKOLSKY BREAST CANCER 16P13 AMPLICON | 0.0191109 |
| M38987 MEBARKI HCC PROGENITOR WNT DN CTNNB1 INDEPENDENT | 0.0165236 |
| M9911 WOO LIVER CANCER RECURRENCE DN | 0.0000000 |
| M1239 JOHANSSON BRAIN CANCER EARLY VS LATE UP | 0.0086624 |
| M16524 LEE LIVER CANCER DENA UP | 0.0240405 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------------|
| M7068 BONOME OVARIAN CANCER SURVIVAL OPTIMAL DEBULKING | 0.0250514 |
| M5840 SMID BREAST CANCER RELAPSE IN LUNG UP | 0.0014309 |
| M39738 WP NONSMALL CELL LUNG CANCER | 0.0311667 |
| M12212 WANG ESOPHAGUS CANCER VS NORMAL DN | 0.0000062 |
| M3464 VANTVEER BREAST CANCER BRCA1 DN | 0.0237355 |
| M1883 BOYVAULT LIVER CANCER SUBCLASS G1 DN | 0.0037157 |
| M16496 CHIANG LIVER CANCER SUBCLASS CTNNB1 UP | 0.0937794 |
| M16586 LAIHO COLORECTAL CANCER SERRATED UP | 0.0032674 |
| M39623 WP ENDOMETRIAL CANCER | 0.0010332 |
| M500 VANTVEER BREAST CANCER METASTASIS UP | 0.0479403 |
| M393 VANTVEER BREAST CANCER ESRI UP | 0.0000002 |
| M39831 WP VITAMIN DSENSITIVE CALCIUM SIGNALING IN DEPRESSION | 0.0551975 |
| M697 KONDO COLON CANCER HCP WITH H3K27ME3 | 0.0252173 |
| M1371 SCHAEFFER PROSTATE DEVELOPMENT AND CANCER BOX3 | 0.0000741 |
| M1344 FUJIWARA PARK2 IN LIVER CANCER DN | 0.0046982 |
| M1343 FUJIWARA PARK2 IN LIVER CANCER UP | 0.0345383 |
| M5098 CLIMENT BREAST CANCER COPY NUMBER DN | 0.0023238 |
| M2184 KAPOSI LIVER CANCER MET DN | 0.0000268 |
| M39621 WP TCA CYCLE NUTRIENT UTILIZATION AND INVASIVENESS OF OVARIAN CANCER | 0.0037403 |
| M27068 REACTOME PRE NOTCH PROCESSING IN THE ENDOPLASMIC RETICULUM | 0.0066061 |
| M26903 REACTOME ACTIVATION OF NOXA AND TRANSLOCATION TO MITOCHONDRIA | 0.0024455 |
| M14019 DING LUNG CANCER MUTATED RECURRENTLY | 0.0576514 |
| M26905 REACTOME RELEASE OF APOPTOTIC FACTORS FROM THE MITOCHONDRIA | 0.0565366 |
| M27720 REACTOME MITOCHONDRIAL FATTY ACID BETA OXIDATION OF UNSATURATED FATTY ACIDS | 0.0000129 |

Tabla B.69: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (CAMERA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------|
| M16009 ACEVEDO METHYLATED IN LIVER CANCER DN | 0.0086338 |

Tabla B.70: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA).

| Conjuntos | pval.adj |
|--|-----------|
| M16740 GRADE COLON AND RECTAL CANCER UP | 0.0000000 |
| M18855 VECCHI GASTRIC CANCER EARLY UP | 0.0000000 |
| M15664 ROSTY CERVICAL CANCER PROLIFERATION CLUSTER | 0.0000000 |
| M8520 SHEDDEN LUNG CANCER POOR SURVIVAL A6 | 0.0000000 |
| M3766 SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP | 0.0000000 |
| M7854 WHITEFORD PEDIATRIC CANCER MARKERS | 0.0000000 |
| M5792 LI AMPLIFIED IN LUNG CANCER | 0.0012055 |
| M854 LINDGREN BLADDER CANCER CLUSTER 3 UP | 0.0000000 |
| M15150 ZHANG BREAST CANCER PROGENITORS UP | 0.0000000 |
| M10986 CHIANG LIVER CANCER SUBCLASS UNANNOTATED DN | 0.0000000 |
| M8365 RHODES UNDIFFERENTIATED CANCER | 0.0000000 |
| M1345 RHODES CANCER META SIGNATURE | 0.0000000 |
| M3268 CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP | 0.0000000 |
| M7987 LEE LIVER CANCER SURVIVAL DN | 0.0000000 |
| M18436 BOYALT LIVER CANCER SUBCLASS G3 UP | 0.0000000 |
| M17423 LAIHO COLORECTAL CANCER SERRATED DN | 0.0119930 |
| M39678 WP RETINOBLASTOMA GENE IN CANCER | 0.0000000 |
| M13831 BOYALT LIVER CANCER SUBCLASS G123 UP | 0.0000911 |
| M18973 NIKOLSKY BREAST CANCER 20Q12 Q13 AMPLICON | 0.0042064 |
| M19982 WAMUNYOKOLI OVARIAN CANCER LMP UP | 0.0000150 |
| M17788 LINDGREN BLADDER CANCER CLUSTER 1 DN | 0.0000000 |

Tabla B.70: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M15356 PUJANA BREAST CANCER LIT INT NETWORK | 0.0000004 |
| M9826 VANTVEER BREAST CANCER METASTASIS DN | 0.0000000 |
| M17299 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL UP | 0.0070560 |
| M12016 OUELLET OVARIAN CANCER INVASIVE VS LMP UP | 0.0000053 |
| M17094 GINESTIER BREAST CANCER ZNF217 AMPLIFIED DN | 0.0809056 |
| M336 VILLANUEVA LIVER CANCER KRT19 UP | 0.0000000 |
| M17333 FARMER BREAST CANCER CLUSTER 2 | 0.0000000 |
| M1687 ZHONG SECRETOME OF LUNG CANCER AND FIBROBLAST | 0.0117571 |
| M2427 NAKAMURA CANCER MICROENVIRONMENT DN | 0.0000114 |
| M7995 HOSHIDA LIVER CANCER SUBCLASS S2 | 0.0011621 |
| M15193 GRUETZMANN PANCREATIC CANCER UP | 0.0008054 |
| M2435 WAMUNYOKOLI OVARIAN CANCER GRADES 1 2 UP | 0.0143771 |
| M39740 WP METABOLIC REPROGRAMMING IN COLON CANCER | 0.0001267 |
| M15112 WONG MITOCHONDRIA GENE MODULE | 0.0018832 |
| M1685 ZHONG SECRETOME OF LUNG CANCER AND MACROPHAGE | 0.0990650 |
| M9577 MOOTHA MITOCHONDRIA | 0.0010106 |
| M3710 WALLACE PROSTATE CANCER UP | 0.0000638 |
| M1260 WATANABE RECTAL CANCER RADIOTHERAPY RESPONSIVE UP | 0.0740031 |
| M6862 NADERI BREAST CANCER PROGNOSIS UP | 0.0000002 |
| M12795 CHARAFE BREAST CANCER BASAL VS MESENCHYMAL UP | 0.0375507 |
| M13736 PYEON CANCER HEAD AND NECK VS CERVICAL UP | 0.0048261 |
| M19148 TOMLINS PROSTATE CANCER UP | 0.0860589 |
| M39628 WP INTEGRATED CANCER PATHWAY | 0.0001608 |
| M16374 IIZUKA LIVER CANCER PROGRESSION G1 G2 DN | 0.0131734 |
| M16542 YAMASHITA LIVER CANCER WITH EPCAM UP | 0.0791312 |

Tabla B.70: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M4448 BOYALT LIVER CANCER SUBCLASS G23 UP | 0.0000390 |
| M13072 SMID BREAST CANCER LUMINAL A DN | 0.0000001 |
| M18086 DAIRKEE CANCER PRONE RESPONSE BPA E2 | 0.0043351 |
| M19097 SWEET LUNG CANCER KRAS UP | 0.0013496 |
| M590 REACTOME MITOCHONDRIAL PROTEIN IMPORT | 0.0586836 |
| M8655 DAIRKEE CANCER PRONE RESPONSE BPA | 0.0604582 |
| M747 REACTOME METABOLISM OF POLYAMINES | 0.0904575 |
| M14146 BOYALT LIVER CANCER SUBCLASS G1 UP | 0.0496174 |
| M1686 ZHONG SECRETOME OF LUNG CANCER AND ENDOTHELIUM | 0.0904575 |
| M8779 LI LUNG CANCER | 0.0117799 |
| M39347 WP GASTRIC CANCER NETWORK 1 | 0.0000410 |
| M5369 POOLA INVASIVE BREAST CANCER UP | 0.0060207 |
| M2180 PUJANA BREAST CANCER WITH BRCA1 MUTATED UP | 0.0000000 |
| M416 ANDERSEN LIVER CANCER KRT19 UP | 0.0198854 |
| M2082 WANG METASTASIS OF BREAST CANCER ESR1 UP | 0.0602808 |
| M15425 VANTVEER BREAST CANCER ESR1 DN | 0.0000066 |
| M11629 MONTERO THYROID CANCER POOR SURVIVAL UP | 0.0000187 |
| M1273 SHETH LIVER CANCER VS TXNIP LOSS PAM1 | 0.0000593 |
| M867 MEINHOLD OVARIAN CANCER LOW GRADE DN | 0.0083882 |
| M39497 WP GASTRIC CANCER NETWORK 2 | 0.0013208 |
| M9389 FINETTI BREAST CANCER KINOME RED | 0.0000000 |
| M39360 WP INTEGRATED BREAST CANCER PATHWAY | 0.0015414 |
| M26973 REACTOME MITOCHONDRIAL BIOGENESIS | 0.0514610 |
| M836 OUELLET CULTURED OVARIAN CANCER INVASIVE VS LMP UP | 0.0001555 |
| M8411 LINDGREN BLADDER CANCER WITH LOH IN CHR9Q | 0.0929375 |

Tabla B.70: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M10837 NIKOLSKY BREAST CANCER 20Q11 AMPLICON | 0.0332468 |
| M4619 LIU PROSTATE CANCER UP | 0.0466993 |
| M378 MINGUEZ LIVER CANCER VASCULAR INVASION UP | 0.0816451 |
| M5652 FARMER BREAST CANCER BASAL VS LULMINAL | 0.0001234 |
| M9728 BARRIER COLON CANCER RECURRENCE UP | 0.0051969 |
| M7678 CHIANG LIVER CANCER SUBCLASS POLYSOMY7 DN | 0.0412455 |
| M8817 ACEVEDO LIVER CANCER WITH H3K9ME3 DN | 0.0115481 |
| M1758 SANSOM WNT PATHWAY REQUIRE MYC | 0.0001234 |
| M9940 HONRADO BREAST CANCER BRCA1 VS BRCA2 | 0.0043411 |
| M2631 FARMER BREAST CANCER APOCRINE VS BASAL | 0.0000415 |
| M3645 DELYS THYROID CANCER UP | 0.0004259 |
| M39731 WP LNCRNA INVOLVEMENT IN CANONICAL WNT SIGNALING AND COLORECTAL CANCER | 0.0186399 |
| M7547 JIANG HYPOXIA CANCER | 0.0455740 |
| M18761 SMITH LIVER CANCER | 0.0546494 |
| M8526 CHEN LUNG CANCER SURVIVAL | 0.0151244 |
| M38998 REACTOME ABERRANT REGULATION OF MITOTIC G1 S TRANSITION IN CANCER DUE TO RB1 DEFECTS | 0.0037212 |
| M4290 LUI THYROID CANCER CLUSTER 3 | 0.0468355 |
| M14383 SMID BREAST CANCER RELAPSE IN BRAIN UP | 0.0008351 |
| M9150 GRESHOCK CANCER COPY NUMBER UP | 0.0288948 |
| M2202 JAZAERI BREAST CANCER BRCA1 VS BRCA2 UP | 0.0636209 |
| M796 REACTOME ATF4 ACTIVATES GENES IN RESPONSE TO ENDOPLASMIC RETICULUM STRESS | 0.0614403 |
| M19076 GINESTIER BREAST CANCER 20Q13 AMPLIFICATION UP | 0.0964102 |
| M27145 REACTOME TRANSCRIPTIONAL ACTIVATION OF MITOCHONDRIAL BIOGENESIS | 0.0031647 |

Tabla B.70: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------------|
| M9874 OUYANG PROSTATE CANCER MARKERS | 0.0012744 |
| M17936 NIKOLSKY BREAST CANCER 8Q23 Q24 AMPLICON | 0.0933550 |
| M4925 STEARMAN LUNG CANCER EARLY VS LATE UP | 0.0090587 |
| M8124 SMID BREAST CANCER BASAL UP | 0.0811818 |
| M1293 YANG BREAST CANCER ES1 DN | 0.0070422 |
| M1218 BORLAK LIVER CANCER EGF UP | 0.0108398 |
| M672 CHIN BREAST CANCER COPY NUMBER UP | 0.0038019 |
| M3228 KEGG SMALL CELL LUNG CANCER | 0.0090249 |
| M39563 WP NCRNAS INVOLVED IN WNT SIGNALING IN HEPATOCELLULAR CARCINOMA | 0.0586836 |
| M4381 LUI THYROID CANCER PAX8 PPARG UP | 0.0723893 |
| M39407 WP WNTBETACATENIN SIGNALING PATHWAY IN LEUKEMIA | 0.0955687 |
| M3432 LEE LIVER CANCER MYC TGFA UP | 0.0213828 |
| M6875 LANDIS BREAST CANCER PROGRESSION UP | 0.0048005 |
| M376 PUJANA BREAST CANCER WITH BRCA1 MUTATED DN | 0.0263231 |
| M6046 BARRIER CANCER RELAPSE TUMOR SAMPLE UP | 0.0378543 |
| M13954 NIKOLSKY BREAST CANCER 8Q12 Q22 AMPLICON | 0.0072096 |
| M39417 WP ELECTRON TRANSPORT CHAIN OXPHOS SYSTEM IN MITOCHONDRIA | 0.0198018 |
| M1805 WILLERT WNT SIGNALING | 0.0756380 |
| M19675 FARMER BREAST CANCER CLUSTER 3 | 0.0035453 |
| M6195 HWANG PROSTATE CANCER MARKERS | 0.0007826 |
| M13709 NIKOLSKY BREAST CANCER 12Q13 Q21 AMPLICON | 0.0076128 |
| M39619 WP SENESCENCE AND AUTOPHAGY IN CANCER | 0.0350714 |
| M16257 BIOCARTA MITOCHONDRIA PATHWAY | 0.0975254 |
| M17044 WANG HCP PROSTATE CANCER | 0.0012089 |
| M1177 KOINUMA COLON CANCER MSI DN | 0.0697393 |
| M1253 INAMURA LUNG CANCER SCC UP | 0.0972543 |
| M2742 LEE LIVER CANCER MYC UP | 0.0410723 |

Tabla B.70: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------|
| M39425 WP MIRNA REGULATION OF P53 PATHWAY IN PROSTATE CANCER | 0.0317172 |
| M39387 WP WNT SIGNALING PATHWAY AND PLURIPOTENCY | 0.0624626 |
| M16737 VANTVEER BREAST CANCER BRCA1 UP | 0.0300734 |
| M18394 ROYLANCE BREAST CANCER 16Q COPY NUMBER UP | 0.0148994 |
| M9538 GRADE COLON VS RECTAL CANCER UP | 0.0446288 |
| M11078 YANG BREAST CANCER ESR1 LASER DN | 0.0000773 |
| M14132 SLEBOS HEAD AND NECK CANCER WITH HPV UP | 0.0050964 |
| M1239 JOHANSSON BRAIN CANCER EARLY VS LATE UP | 0.0238817 |
| M16141 CHIANG LIVER CANCER SUBCLASS INTERFERON UP | 0.0566231 |
| M2700 SMID BREAST CANCER RELAPSE IN BONE DN | 0.0000665 |
| M534 KIM LIVER CANCER POOR SURVIVAL DN | 0.0780851 |
| M39834 WP SMALL CELL LUNG CANCER | 0.0309442 |
| M9642 GALLUZZI PERMEABILIZE MITOCHONDRIA | 0.0214354 |
| M2084 WANG METASTASIS OF BREAST CANCER | 0.0514610 |
| M16955 ACEVEDO LIVER CANCER WITH H3K27ME3 DN | 0.0127185 |
| M10922 WANG RECURRENT LIVER CANCER UP | 0.0650525 |
| M16956 YAMASHITA LIVER CANCER STEM CELL UP | 0.0000553 |
| M12868 KEGG PATHWAYS IN CANCER | 0.0048312 |
| M1595 BARRIER CANCER RELAPSE NORMAL SAMPLE UP | 0.0016912 |
| M9951 WANG METHYLATED IN BREAST CANCER | 0.0078494 |
| M4008 FARMER BREAST CANCER APOCRINE VS LUMINAL | 0.0025759 |
| M8897 AGUIRRE PANCREATIC CANCER COPY NUMBER DN | 0.0292113 |
| M2572 ANASTASSIOU MULTICANCER INVASIVENESS SIGNATURE | 0.0000147 |
| M7577 ACEVEDO LIVER CANCER DN | 0.0015532 |

Tabla B.70: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M4752 WATANABE RECTAL CANCER RADIOTHERAPY RESPONSIVE DN | 0.0007404 |
| M19661 WANG ESOPHAGUS CANCER PROGRESSION UP | 0.0179443 |
| M8659 IIZUKA LIVER CANCER PROGRESSION L1 G1 DN | 0.0540524 |
| M39544 WP MITOCHONDRIAL GENE EXPRESSION | 0.0979293 |
| M5755 KAPOSI LIVER CANCER MET UP | 0.0432479 |
| M1353 SCHAEFFER PROSTATE DEVELOPMENT AND CANCER BOX1 UP | 0.0837468 |
| M5840 SMID BREAST CANCER RELAPSE IN LUNG UP | 0.0816451 |
| M4342 BOYALT LIVER CANCER SUBCLASS G6 UP | 0.0680360 |
| M14693 VANTVEER BREAST CANCER POOR PROGNOSIS | 0.0000171 |
| M500 VANTVEER BREAST CANCER METASTASIS UP | 0.0560593 |
| M39523 WP BLADDER CANCER | 0.0432773 |
| M13191 KEGG PROSTATE CANCER | 0.0719182 |
| M6939 HOSHIDA LIVER CANCER SURVIVAL UP | 0.0054224 |
| M14690 REACTOME MITOCHONDRIAL FATTY ACID BETA OXIDATION | 0.0290993 |
| M10861 LI PROSTATE CANCER EPIGENETIC | 0.0515130 |
| M1167 HUMMERICH SKIN CANCER PROGRESSION UP | 0.0012145 |
| M15125 FARMER BREAST CANCER CLUSTER 4 | 0.0046949 |
| M13123 LEE LIVER CANCER HEPATOBLAST | 0.0801241 |
| M12212 WANG ESOPHAGUS CANCER VS NORMAL DN | 0.0062951 |
| M14353 CHIANG LIVER CANCER SUBCLASS INTERFERON DN | 0.0088942 |
| M2212 TAKADA GASTRIC CANCER COPY NUMBER UP | 0.0394063 |
| M19096 KEGG BLADDER CANCER | 0.0248856 |
| M11238 LINDGREN BLADDER CANCER CLUSTER 3 DN | 0.0264157 |
| M12210 YOKOE CANCER TESTIS ANTIGENS | 0.0032832 |
| M38984 MEBARKI HCC PROGENITOR WNT UP CTNNB1 DEPENDENT | 0.0986189 |
| M1275 SHETH LIVER CANCER VS TXNIP LOSS PAM2 | 0.0007749 |

Tabla B.70: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | pval.adj |
|---|-----------------|
| M7517 SMID BREAST CANCER LUMINAL A UP | 0.0001819 |
| M11302 SMID BREAST CANCER RELAPSE IN LUNG DN | 0.0133852 |
| M10748 YANG BREAST CANCER ESRI BULK UP | 0.0516508 |
| M39015 REACTOME ABERRANT REGULATION OF MITOTIC EXIT IN CANCER DUE TO RB1 DEFECTS | 0.0175236 |
| M6767 POOLA INVASIVE BREAST CANCER DN | 0.0988275 |
| M4034 YEGNASUBRAMANIAN PROSTATE CANCER | 0.0421578 |
| M13273 DELYS THYROID CANCER DN | 0.0167545 |
| M10541 YANG BREAST CANCER ESRI LASER UP | 0.0594917 |
| M12602 WOO LIVER CANCER RECURRENCE UP | 0.0008401 |
| M16496 CHIANG LIVER CANCER SUBCLASS CTNNB1 UP | 0.0393208 |
| M393 VANTVEER BREAST CANCER ESRI UP | 0.0571696 |
| M17427 DOANE BREAST CANCER CLASSES UP | 0.0242391 |
| M27719 REACTOME MITOCHONDRIAL FATTY ACID BETA OXIDATION OF SATURATED FATTY ACIDS | 0.0366316 |
| M11213 SATO SILENCED BY DEACETYLATION IN PANCREATIC CANCER | 0.0438897 |
| M6145 LEE LIVER CANCER SURVIVAL UP | 0.0744350 |
| M8976 WATANABE COLON CANCER MSI VS MSS DN | 0.0897458 |
| M1280 SHETH LIVER CANCER VS TXNIP LOSS PAM4 | 0.0002076 |
| M15351 SUZUKI AMPLIFIED IN ORAL CANCER | 0.0065424 |
| M9908 LANDIS BREAST CANCER PROGRESSION DN | 0.0012137 |
| M11405 TAKADA GASTRIC CANCER COPY NUMBER DN | 0.0580066 |
| M10394 FARMER BREAST CANCER CLUSTER 7 | 0.0930303 |
| M7012 SCHUETZ BREAST CANCER DUCTAL INVASIVE DN | 0.0527988 |
| M38982 MEBARKI HCC PROGENITOR WNT UP | 0.0443648 |
| M11771 DOANE BREAST CANCER CLASSES DN | 0.0633886 |
| M7068 BONOME OVARIAN CANCER SURVIVAL OPTIMAL DEBULKING | 0.0546527 |
| M1352 NIKOLSKY BREAST CANCER 15Q26 AMPLICON | 0.0530918 |

Tabla B.70: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MCDB (ORA)
(continuación)

| Conjuntos | pval.adj |
|--|-----------------|
| M13658 HOSHIDA LIVER CANCER LATE RECURRENCE DN | 0.0821174 |
| M2043 RICKMAN HEAD AND NECK CANCER F | 0.0842539 |
| M2108 VECCHI GASTRIC CANCER ADVANCED VS EARLY DN | 0.0004418 |
| M6972 MONTERO THYROID CANCER POOR SURVIVAL DN | 0.0740031 |
| M14181 YAMASHITA METHYLATED IN PROSTATE CANCER | 0.0238495 |
| M38988 MEBARKI HCC PROGENITOR WNT UP BLOCKED BY FZD8CRD | 0.0432479 |
| M9206 YAMASHITA LIVER CANCER STEM CELL DN | 0.0531856 |
| M39365 WP MITOCHONDRIAL LCFATTY ACID BETAOXIDATION | 0.0003589 |
| M15780 GRADE COLON AND RECTAL CANCER DN | 0.0061789 |
| M1286 HOSHIDA LIVER CANCER SUBCLASS S3 | 0.0038414 |
| M16932 CHIANG LIVER CANCER SUBCLASS PROLIFERATION DN | 0.0979293 |
| M424 ANDERSEN LIVER CANCER KRT19 DN | 0.0007041 |
| M7396 SWEET LUNG CANCER KRAS DN | 0.0969064 |
| M15472 VECCHI GASTRIC CANCER EARLY DN | 0.0000000 |
| M14507 CHARAFE BREAST CANCER LUMINAL VS BASAL DN | 0.0002942 |
| M1362 SCHAEFFER PROSTATE DEVELOPMENT AND CANCER BOX5 UP | 0.0213185 |
| M2899 LINDGREN BLADDER CANCER HIGH RECURRENCE | 0.0530060 |
| M9192 CHARAFE BREAST CANCER LUMINAL VS MESENCHYMAL DN | 0.0434822 |
| M19391 LIU PROSTATE CANCER DN | 0.0323479 |
| M11504 TOMLINS PROSTATE CANCER DN | 0.0279722 |
| M7081 FARMER BREAST CANCER CLUSTER 8 | 0.0897537 |

Tabla B.71: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MH (CAMERA).

| Conjuntos | pval.adj |
|--|-----------------|
| M5895 HALLMARK WNT BETA CATENIN SIGNALING | 0.0031551 |

Tabla B.72: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (CAMERA).

| Conjuntos | pval.adj |
|---|-----------|
| M8477 GSE26351 WNT VS BMP PATHWAY STIM HEMATOPOIETIC PROGENITORS UP | 0.0329825 |
| M8482 GSE26351 UNSTIM VS WNT PATHWAY STIM HEMATOPOIETIC PROGENITORS UP | 0.0398691 |

Tabla B.73: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MIM (ORA).

| Conjuntos | pval.adj |
|---|-----------|
| M8477 GSE26351 WNT VS BMP PATHWAY STIM HEMATOPOIETIC PROGENITORS UP | 0.0092232 |
| M8482 GSE26351 UNSTIM VS WNT PATHWAY STIM HEMATOPOIETIC PROGENITORS UP | 0.0543756 |

Tabla B.74: Análisis de enriquecimiento de los conjuntos de genes de la base de datos MOG (CAMERA).

| Conjuntos | pval.adj |
|--------------------|-----------|
| M2689 WNT UP.V1 DN | 0.0177093 |

Apéndice C

Análisis funcional del efecto del DFMO sobre la homeostasis del Ca^{2+} intracelular

A continuación, se muestran la modelización, mediante modelos lineales mixtos, de cada una de las variables respuesta siguientes: *Área Bajo la Curva (AUC) de la señal correspondiente a los depósitos (AUCDepositos)*, *Incremento Máximo (Max) de la señal correspondiente a los depósitos (MaxDepositos)*, *AUC de la señal correspondiente a SOCE (AUCSOCE)* y *Máximo de la señal correspondiente a SOCE (MaxSOCE)*. Además, se presenta un caso práctico que consta de un breve resumen del análisis llevado a cabo para modelizar la variable respuesta *AUCDepositos* mediante modelos lineales mixtos, pues dicho procedimiento y planteamiento seguido se puede trasladar al análisis del resto de variables respuesta de estos experimentos de imagen de Ca^{2+} . Es más, dicho planteamiento y procedimiento, puesto que es similar, también se puede trasladar a los experimentos de *Western Blot*. Así, salvo en el caso práctico, idéntico para los modelos de las otras tres variables respuesta evaluadas, sólo se mostrarán las figuras y tablas de estos otros tres modelos.

C.1. Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva (AUC) de la señal correspondiente al tamaño de los depósitos*

A continuación, se muestra detalladamente el proceso de modelización de la variable respuesta *Área bajo la curva de la señal de los depósitos de Ca^{2+} intracelular (AUCDepositos)* mediante modelos lineales mixtos. En primer lugar, se ajustará el modelo completo o sobreóptimo y se hará un diagnóstico del modelo de este con el fin de encontrar posibles problemas o de plantear la posibilidad

de realizar alguna transformación de alguna de las variables. Después, se realiza una selección del modelo mediante diferentes criterios y, del modelo seleccionado, se realiza nuevamente un diagnóstico, se eliminan los posibles *outliers* y se rehace el diagnóstico con el fin de cerciorarse que se cumplan los supuestos del modelo. Finalmente, se contratan las hipótesis de interés tanto sobre los coeficientes del modelo, de los efectos aleatorios como de las comparaciones dos a dos de las condiciones experimentales. Finalmente, en caso de que el modelo tenga ciertas limitaciones, debido a que la violación de los supuestos sea muy grande, se aplicarán técnicas de remuestreo o *Boostraping* paramétrico.

Como librerías básicas, se han empleado *lme4* (Bates et al. (2015)), *nlme* (Pinheiro et al. (2021)), *cAIC4* (Saeften et al. (2018)), *car* (Fox and Weisberg (2019)) y *multcomp* (Hothorn, Bretz, and Westfall (2008)).

C.1.0.1. evaluación del modelo sobreóptimo

Para analizar el efecto del DFMO sobre el tamaño de los depósitos cómo de diferente es este efecto entre las líneas celulares de interés se ha modelizado la variable respuesta *AUCDepositos* para cada una de las células registradas. Puesto que cada día se han realizado dos réplicas de cada condición experimental y a que en cada una de las réplicas se registran varias células, estos datos muestran una estructura clúster o agrupada y, además, jerarquizada. En particular, la estructura jerarquizada se debe a que cada día hay dos réplicas de cada condición experimental, por lo que todas ellas están agrupadas dentro del mismo día y, además, cada célula está agrupada dentro de un cubreobjetos. En consecuencia, cada célula está agrupada dentro de un cubreobjetos que, a su vez, se agrupa dentro de un día. En cuanto a estos factores aleatorios o de agrupación, que se denominarán *Día* y *Cristal*, cabe destacar que los factores fijos *Línea Celular* y *Tratamiento con DFMO* varían dentro de las muestras agrupadas en el mismo *Día*, por lo que el factor aleatorio día puede tener asociados los efectos aleatorios *intercepto aleatoria* y/o *pendiente aleatoria* para cada uno de los factores fijos ($1 + \text{Tratamiento} * \text{Línea}|\text{Día}$). Por el contrario, dentro de cada cristal las células tienen el mismo tratamiento y son de la misma línea celular ($1|\text{Día}:\text{Cristal}$). En consecuencia, hay que recurrir a modelos lineales mixtos.

En primer lugar, se ajusta el modelo completo o sobreóptimo, $\text{AUCDeposito} = \text{linea} * \text{tratamiento} + \text{dia}/\text{Cristal}$, y se hace un diagnóstico a través del estudio tanto de los residuos de Pearson (figura C.1) como de los efectos aleatorios estimados mediante el *mejor predictor lineal insesgado empírico (EBLUP)* (C.2). En primer lugar, mediante el estudio de los residuos de Pearson, los cuales siguen una $N(0, 1)$, lo que supone que son *outliers* aquellas observaciones con un valor absoluto del residuo asociado superior a 1.96, se observan algunos problemas (figura C.1).

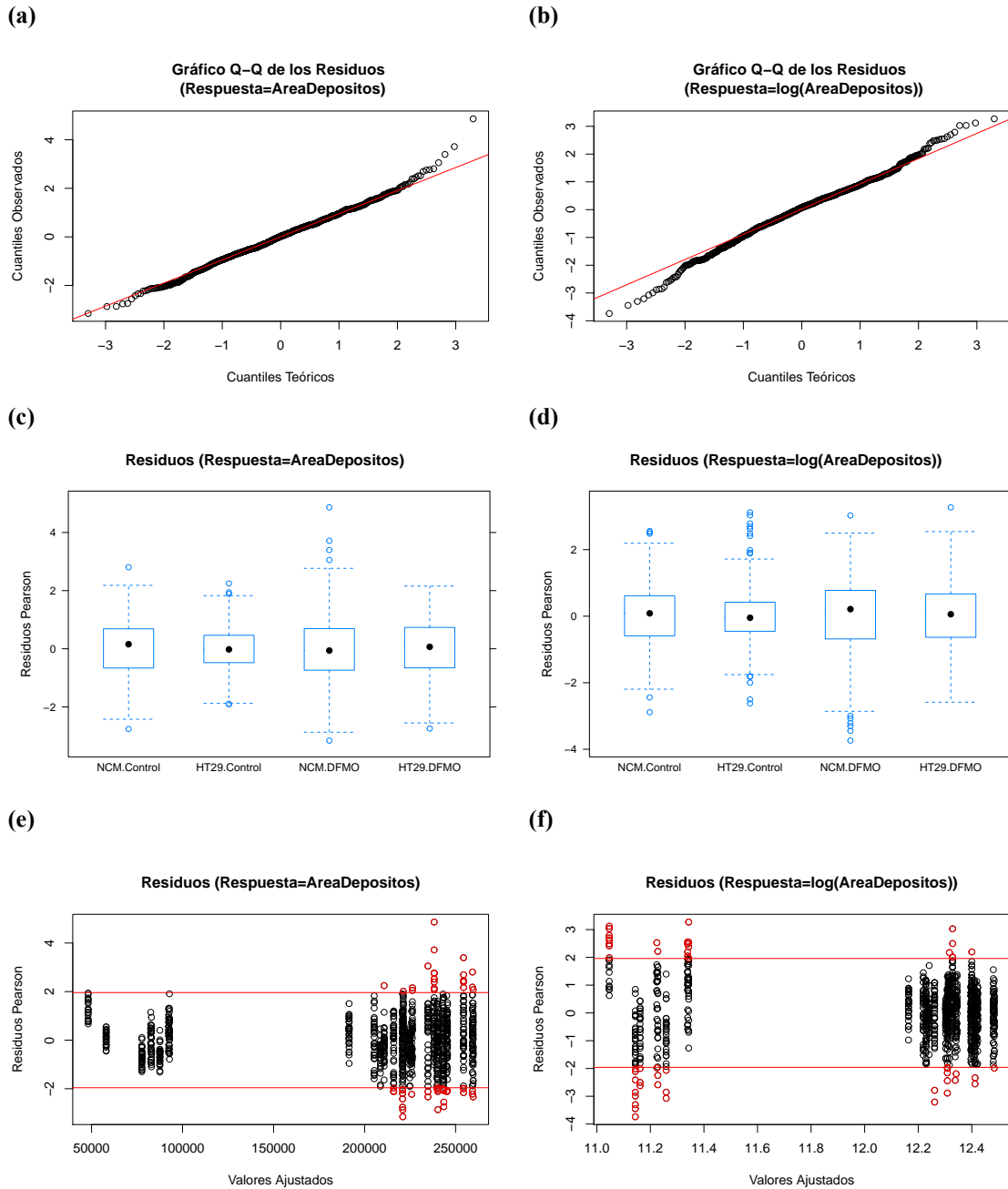


Figura C.1: Diagnóstico de los residuos de los modelos lineales mixtos para la variable respuesta AUCDepositos y $\log(\text{AUCDepositos})$. Para los modelos lineales mixtos cuyas variables respuesta AUCDepositos (gráficos de la izquierda) y $\log(\text{AUCDepositos})$ (gráficos de la derecha) se representan los residuos de Pearson mediante los siguientes métodos gráficos: **a, b) Gráficos Q-Q global**, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; **c, d) Residuos de Pearson vs. condiciones experimentales**, y **e, f) Residuos de Pearson vs. Valores Ajustados**, donde una falta de aleatoriedad indicaría la violación de los supuestos de independencia, una tendencia creciente o decreciente la falta de homocedasticidad y valores absolutos de los residuos superiores a 1.96 (líneas horizontales rojas) la presencia de *outliers* (señalados en rojo).

En concreto, los residuos muestran una distribución que se aleja de la normalidad (figura C.1a), y una clara falta de homocedasticidad (figuras C.1c y C.1e) que, además, aumenta con el valor predicho con un patrón que recuerda a un megáfono, por lo que podría pensarse aplicar una transformación logarítmica de la variable respuesta. Así, el modelo sobreóptimo para la variable respuesta transformada, $\log(\text{AUCDepositos})$, mejora considerablemente la homocedasticidad de los residuos (figuras C.1d y C.1f), si bien es cierto que su distribución se sigue alejando de la normalidad (figura C.1b), posiblemente debido a la presencia de *outliers*, como puede observarse en las figuras C.1d y C.1f en las que se ven residuos cuyos valores absolutos son superiores a 1.96.

Posteriormente, se evalúa la distribución de las estimaciones de los efectos aleatorios y, obviando el factor aleatorio *día* debido al escaso número de “niveles” debido a la dificultad para evaluar los supuestos con una muestra tan pequeña -sólo cuatro-, parece cumplirse la normalidad del factor aleatorio *día/Cristal* tanto en el modelo con la variable transformada como sin transformar (C.2).

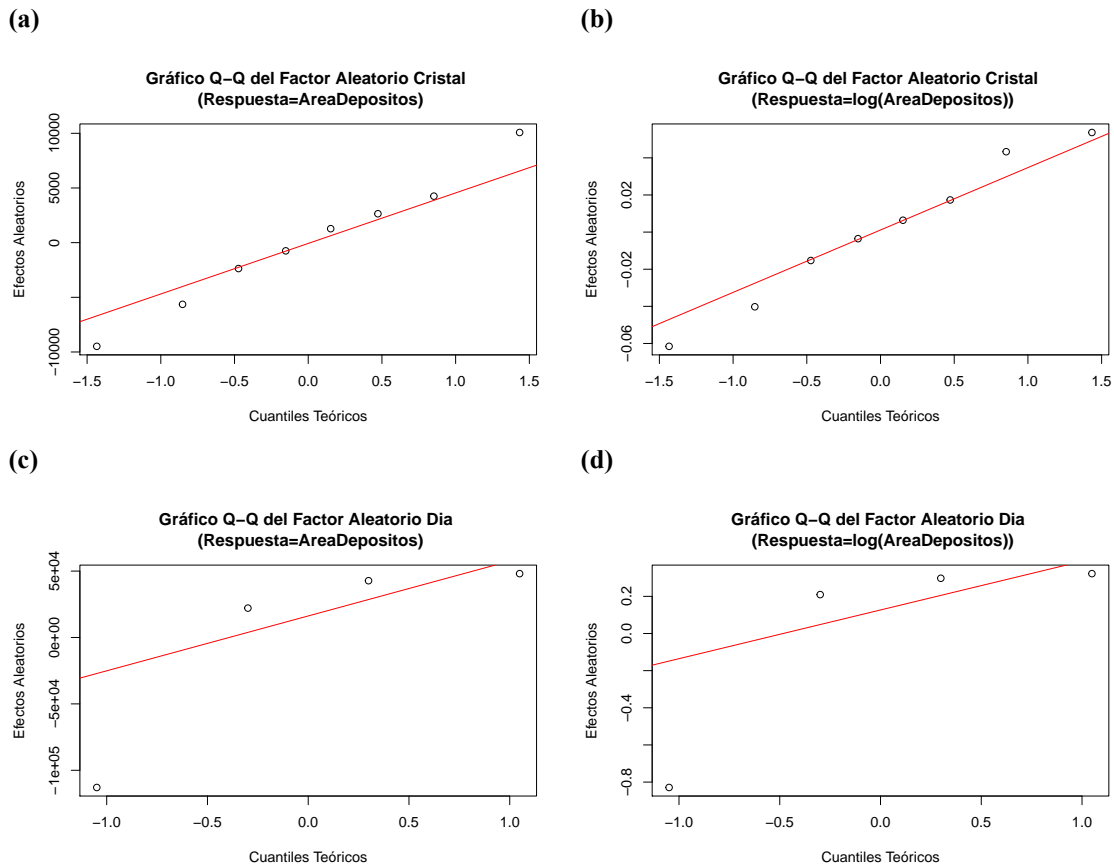


Figura C.2: Diagnóstico de los efectos aleatorios de los modelos lineales mixtos para la variable respuesta AUCDepositos y $\log(\text{AUCDepositos})$. Para las variables respuesta AUCDepositos y $\log(\text{AUCDepositos})$, respectivamente, se representan los efectos aleatorios *día/Cristal* (a, b) y *día* (c, d).

Tras seleccionar el modelo con la variable respuesta transformada, $\log(\text{AUCDepositos})$, se procede a la selección de variables; es decir, a seleccionar aquel modelo que mejor se ajuste a los datos. No obstante, puesto que la parte de fija, en su totalidad, son los factores de interés y, por ende, no han de ser eliminados para poder realizar los contrastes de hipótesis pertinentes, únicamente se lleva a cabo la selección de la parte aleatoria. Para ello, a partir del modelo sobreóptimo, se prueban diferentes estructuras aleatorias y se selecciona aquella que proporcione el menor valor cAIC y que, además, no genere estimaciones problemáticas o errores en el cálculo tanto de cAIC como de las propias estimaciones. En la tabla C.1 se muestra el valor de cAIC para cada una de las estructuras aleatorias evaluadas, donde la estructura aleatoria se expresa según el convenio usado en el paquete lme4 de R:

Tabla C.1: Selección de modelos para la variable respuesta $\log(\text{AUCDepositos})$ con cAIC.

| modelo | Parte Aleatoria (Anotación según lme4) | cAIC |
|---------------|---|-------------|
| C_DTiL | (1 Dia:Cristal) + (1 + Tratamiento * Linea Dia) | error |
| C_DTL | (1 Dia:Cristal) + (1 + Tratamiento + Linea Dia) | -802.72570 |
| C_DT | ((1 Dia:Cristal) + (1 + Tratamiento Dia) | -787.1512 |
| C_DL | (1 Dia:Cristal) + (1 + Linea Dia) | -850.82160 |
| C_D | (1 Dia:Cristal) + (1 Dia) | -787.1512 |
| C | (1 Cristal) | error |
| DTiL | (1 + Tratamiento * Linea Dia) | error |
| DTL | (1 + Tratamiento + Linea Dia) | -841.25861 |
| DT | (1 + Tratamiento Dia) | -840.07916 |
| DL | (1 + Linea Dia) | -811.98116 |
| D | (1 Dia) | -747.417858 |
| Fijo | | 1209.0522 |

Como puede comprobarse, el valor más pequeño de cAIC es para el modelo C_DL. No obstante, algunos de los modelos con menor cAIC estiman unas matrices de varianzas-covarianzas para la parte aleatoria peligrosas en el sentido de que hay efectos aleatorios con varianzas igual a cero (o su límite inferior es cero) y correlaciones entre efectos ± 1 (o al menos uno de sus límites inferior y/o superior). En consecuencia, estos modelos se descartan (Barr et al. (2013), Matuschek et al. (2017)). Así, de los modelos evaluados (en orden creciente de cAIC), en la tabla C.2 se muestran las estimaciones de la varianza de los diferentes efectos aleatorios y sus correlaciones (y sus intervalos de confianza) para aquellos modelos con un cAIC inferior al modelo seleccionado en los que se han identificado algunos de estos problemas.

Tabla C.2: Parte aleatoria de los modelos problemáticos para $\log(\text{AUCDepositos})$. Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza. Con una x se indican las estimaciones problemáticas.

```
#Modelo C_DL
Formula: logAUCDepositos ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 + Linea | Dia)
Data: datosr
REML criterion at convergence: -799.072
Random effects:
Groups      Name          Std.Dev. Corr
Dia:Cristal (Intercept) 0.04916
Dia         (Intercept) 0.58835
           LineaHT29   0.09876 -0.66
Residual                                0.15554

           2.5 %    97.5 % Problemática
.sig01      0.02493630 0.12548479
.sig02      0.28795509 1.25164201
.sig03     -1.00000000 0.43135288      x
.sig04      0.04528200 0.21353927
.sigma      0.14870638 0.16255870
```

```
#Modelo DTL
Formula: logAreaDepositos ~ Linea * Tratamiento + (1 + Tratamiento + Linea | Dia)
Data: datosr
REML criterion at convergence: -796.8524
Random effects:
Groups      Name          Std.Dev. Corr
Dia         (Intercept) 0.54292
           TratamientoDFMO 0.07562 1.00
           LineaHT29      0.09645 -0.63 -0.53
Residual                                0.15653
```

Tabla C.3: Parte aleatoria de los modelos problemáticos para log(AUCDepositos). Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza. Con una x se indican las estimaciones problemáticas. (continuación)

```
#Modelo DT
Formula: logAUCDepositos ~ Linea * Tratamiento + (1 + Tratamiento | Dia)
Data: datosr
REML criterion at convergence: -740.6865
Random effects:
Groups   Name                Std.Dev. Corr
Dia      (Intercept)         0.51154
         TratamientoDFMO 0.08122  1.00
Residual                    0.16180

          2.5 %    97.5 %  Problemática
.sig01    0.2516228  1.08704191
.sig02   -1.0000000  1.00000000      x
.sig03    0.0367120  0.15845727
.sigma    0.1546404  0.16898325
```

```
#Modelo DL
Formula: logAUCDepositos ~ Linea * Tratamiento + (1 + Linea | Dia)
Data: datosr
REML criterion at convergence: -766.6911
Random effects:
Groups   Name                Std.Dev. Corr
Dia      (Intercept)         0.5887
         LineaHT29       0.1009  -0.66
Residual                    0.1590

          2.5 %    97.5 %  Problemática
.sig01    0.28975391  1.25052644
.sig02   -1.00000000  0.27567809      x
.sig03    0.04632770  0.21826604
.sigma    0.15200660  0.16613524
```

Tabla C.4: Parte aleatoria de los modelos problemáticos para log(AUCDepositos). Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza. Con una x se indican las estimaciones problemáticas. (continuación)

```
#Modelo C_DTL
Formula: logAUCDepositos ~ Linea * Tratamiento + (1 | Dia:Cristal) +
                                             + (1 + Tratamiento + Linea | Dia)

Data: datosr
REML criterion at convergence: -827.1375
Random effects:
Groups      Name                Std.Dev. Corr
Dia:Cristal (Intercept)      0.04578
Dia         (Intercept)      0.54505
           TratamientoDFMO  0.07196  0.99
           LineaHT29        0.09449 -0.63 -0.52
Residual                                0.15333

                2.5 %      97.5 %  Problemática
.sig01          0.02314390  0.10885792
.sig02          0.21180246  0.99991983
.sig03          0.73163565  1.00000000      x
.sig04         -1.00000000 -0.48148875      x
.sig05          0.03281527  0.14759046
.sig06         -0.88676804  1.00000000      x
.sig07          0.04373417  0.17965744
.sigma          0.14660469  0.16026153

#Modelo C_DT
Formula: logAreaDepositos ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 + Tratamiento | Dia)
Data: datosr
REML criterion at convergence: -771.6759
Random effects:
Groups      Name                Std.Dev. Corr
Dia:Cristal (Intercept)      0.04825
Dia         (Intercept)      0.51396
           TratamientoDFMO  0.07747  1.00
Residual                                0.15842

                2.5 %      97.5 %  Problemática
.sig01          NA         NA      x
.sig02          NA         NA      x
.sig03          NA         NA      x
.sig04          NA         NA      x
.sigma          NA         NA      x
```

En consecuencia, tras descartar los modelos problemáticos, el modelo que presenta un menor valor de cAIC y no presenta ninguno de los problemas considerados es el *modelo C_D* (tabla C.5).

Tabla C.5: Parte aleatoria del modelo seleccionado para AUCDepositos. Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza.

```
#Modelo C_D
Linear mixed model fit by REML ['lmerMod']
Formula: logAUCDepositos ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 | Dia)
Data: datosr
REML criterion at convergence: -740.7864
Random effects:
Groups      Name          Std.Dev.
Dia:Cristal (Intercept) 0.05179
Dia         (Intercept) 0.55793
Residual                    0.16093
Number of obs: 981, groups: Dia:Cristal, 8; Dia, 4
```

Posteriormente, mediante la AIC, se evalúa la parte fija del modelo lineal mixto. Como resultado, en la tabla C.6 se muestra la selección de la parte fija que, como puede verse, resultan ser tanto los dos efectos principales y su interacción.

Tabla C.6: Selección de modelos para la variable respuesta $\log(\text{AUCDepositos})$ con AIC.

| Parte Fija | AIC |
|--|---------|
| Linea + Tratamiento + Linea:Tratamiento | -748.14 |
| Linea + Tratamiento | -613.17 |
| Linea | -595.47 |
| Tratamiento | -588.93 |
| nulo | -571.06 |

Tras evaluar el modelo sobreóptimo con el fin de encontrar tanto la parte fija como la parte aleatoria más adecuada, se obtuvo que el modelo más adecuado para modelizar AUCDepositos es (ecuación C.1):

$$\log(\text{AUCDepositos})_{ij\text{dcr}} = \beta_0 + \beta_1 \text{Linea}_i + \beta_2 \text{Tratamiento}_j + \beta_3 \text{Tejido}_i \text{Tratamiento}_j + b_{0,d} + b_{0,c} + \epsilon_{ij\text{dcr}} \quad (\text{C.1})$$

donde $\log(\text{AUCDepositos})_{ij\text{dcr}}$ es el valor de la variable respuesta de AUCDepositos medido en la réplica r de la muestra perteneciente a la línea celular i (NCM460 o HT29), tratamiento j (Control o DFMO) medidas en el día d y en el cristal c . En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel NCM460.Control); β_1 es el cambio de expresión en la línea HT29 respecto del de referencia, β_2 el cambio en $\log(\text{AUCDepositos})$ en las células tratadas con DFMO respecto de las de referencia (NCM460.Control); β_3 el efecto de la interacción entre los factores *Linea* y *Tratamiento* que indica

si el cambio entre células tratadas con DFMO y sin tratar es diferente en NCM460 que en HT29 (en caso de ser significativo el término de interacción); $b_{0,d}$ es el intercepto aleatorio debido al día que indica la desviación respecto de β_0 que hay en el día d , y $b_{0,c}$ es el intercepto aleatorio debido a cristal o cubreobjetos c que indica la desviación respecto de β_0 que se observa en el cristal c .

A continuación, se realiza el diagnóstico del modelo para evaluar el cumplimiento de los supuestos. Como se puede observar en la figura C.1, hay presencia de *outliers*, por lo que estos se eliminan, se ajusta el modelo a los nuevos datos y se rehace el diagnóstico (figura C.3 y tabla C.7). Por un lado, mediante test de contrastes de hipótesis, se observa una ligera violación de los supuestos de normalidad para los residuos y de la homocedasticidad para los residuos y el efecto aleatorio *Cristal*, pero no es preocupante (tabla C.7. Por otro lado, mediante el uso de métodos gráficos (figura C.3), podría aceptarse el cumplimiento de los supuestos del modelo. En concreto, en relación a los residuos de Pearson (figuras C.3c, d), no se observa una gran heterocedasticidad, sí se observa aleatoriedad y, a pesar de observar una ligera desviación de la normalidad (distribución de colas ligeras) en el gráfico Q-Q (figura C.3a) y una ligera asimetría en el histograma (figura C.3b), no es preocupante.

En cuanto a los efectos aleatorios, se observa que no hay evidencias para pensar que no se distribuyen normalmente (figuras C.3e, f). No obstante, hay que tener precaución con estos gráficos y sólo asumir la desviación de la normalidad en casos claros y la violación del supuesto de normalidad sea obvia, pues los test F (t) usados posteriormente para contrastar las diferentes hipótesis de interés son robustos frente a ello y más aun para un número de observaciones tan elevado (más de 1000 células), por lo que los resultados de los diferentes contrastes son fiables y las inferencias correctas. No obstante, aunque en los métodos gráficos los supuestos de normalidad y homocedasticidad no parecen violarse de forma preocupante, los tests sugieren que la violación es considerable. Por ello, tras realizar las estimaciones y contrastes de hipótesis pertinentes con los modelos lineales mixtos, se emplearán las técnicas de estimación *bootstrapping* paramétrica para evaluar si el modelo es o no adecuado en función de la semejanza entre ambos y, en caso de que no lo sean, se emplearán las estimaciones *bootstrapping*.

Tabla C.7: Supuestos del modelo seleccionado.

| | Normalidad | | | Homocedasticidad | |
|-------------------------------|------------|---------|----------|------------------|----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 1.01e-05 | 0.00348 | 0.000496 | 0.000451 | 0 |
| Día:Cristal intercepto | 7.29e-01 | 0.76900 | 0.957000 | | |
| Día intercepto | 1.01e-02 | 0.62600 | | | |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test de Breush-Pagan.

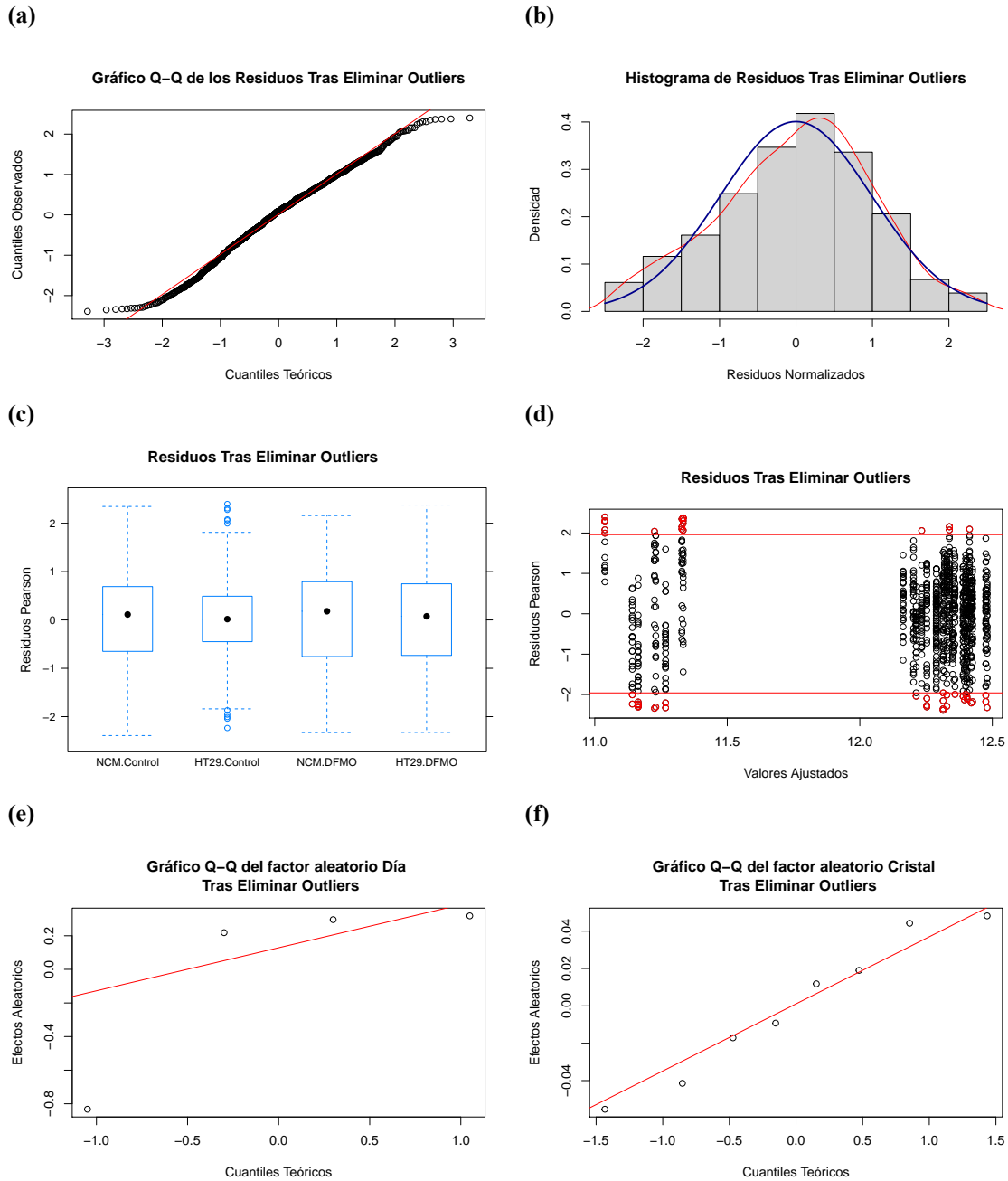


Figura C.3: Diagnóstico del modelo lineal mixto final (completo) para la variable respuesta $\log(\text{AUC}_{\text{Depositos}})$ tras eliminar los *outliers*. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como *outliers*; e) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Día, y f) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Cristal.

Otro gráfico de diagnóstico de suma importancia es el que muestra las predicciones de la variable respuesta para cada una de las células evaluadas (figura {fig:AreaDeppredicciones}). Es más, puesto que existe una estructura jerárquica en los datos (cada célula pertenece a un cristal y cada cristal se ha realizado un día en concreto), también es posible evaluar las predicciones de la variable respuesta para cada grupo de jerarquía; es decir, para las células de un cristal o para las células de un cristal y de un día concreto y, como puede observarse en la figura {fig:AreaDeppredicciones}, las predicciones son bastante buenas, lo que es indicativo de que el modelo es aceptable.

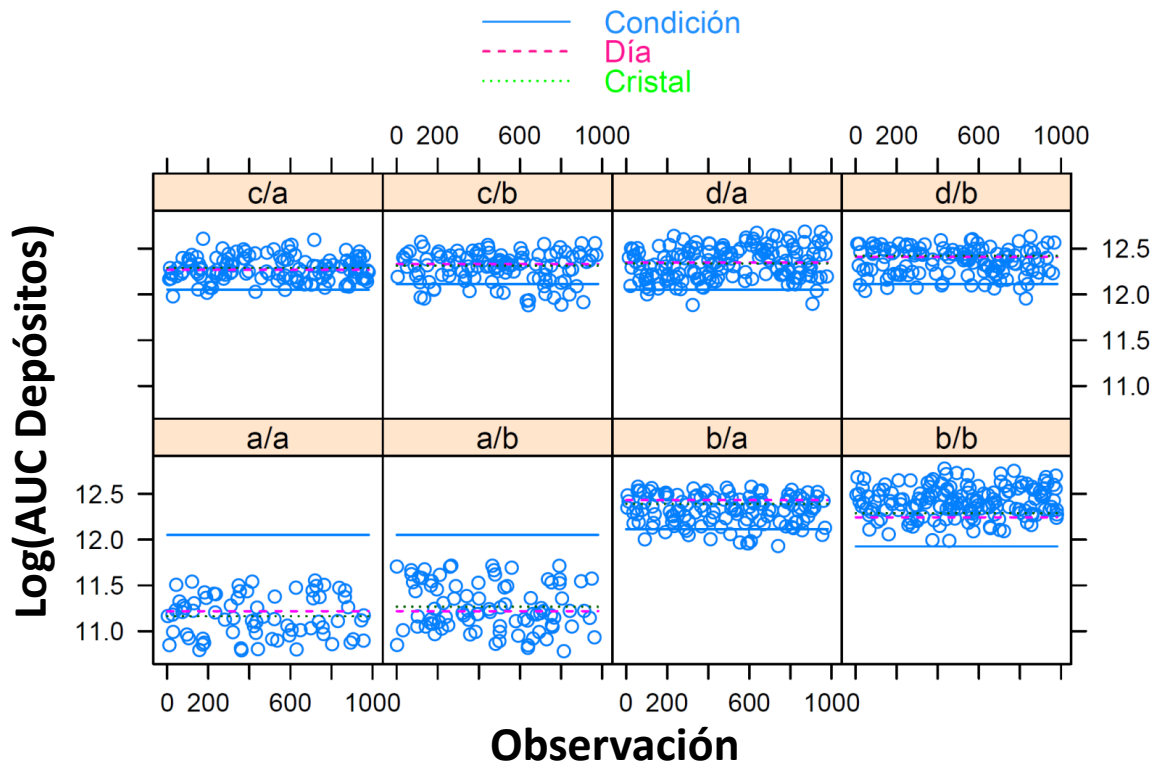


Figura C.4: Predicciones realizadas con el modelo lineal mixto final (completo) para la variable respuesta $\log(\text{AUC Depositos})$ tras eliminar los *outliers*. Se muestran las predicciones realizadas no solo para condición experimental (azul), sino también para el día realizado (rojo) y el cristal (verde).

Por lo tanto, los parámetros estimados de la parte fija y de la parte aleatoria se muestran en las tablas C.8 y C.9, respectivamente.

Tabla C.8: Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto seleccionado para $\log(\text{AUC Depósitos})$.

| | Estimación | SEM | t | p-valor | R^2_{Xu00} |
|-------------------------------------|------------|-----------|------------|----------|---------------------|
| β_0 | 12.1130749 | 0.2797475 | 43.300031 | 0 | 0.8751914 |
| $\beta_{\text{HT29}}(\beta_1)$ | -0.1870926 | 0.0149219 | -12.538133 | 0 | 0.8751914 |
| $\beta_{\text{DFMO}}(\beta_2)$ | -0.0615153 | 0.0138772 | -4.432847 | 1.04E-05 | 0.8751914 |
| $\beta_{\text{HT29:DFMO}}(\beta_3)$ | 0.2534924 | 0.0209268 | 12.113304 | 0 | 0.8751914 |

Tabla C.9: Estimaciones de la varianza de la parte aleatoria del modelo mixto seleccionado para $\log(\text{AUC Depósitos})$.

| Componente Aleatorio | LI | Estimación SEM | LS |
|--|--------|----------------|--------|
| Día intercepto (b_{0d}) | 0.2499 | 0.5579 | 1.2456 |
| Día:Cristal intercepto (b_{0c}) | 0.0242 | 0.0518 | 0.1108 |
| Residual (ϵ) | 0.1539 | 0.1609 | 0.1682 |

Como puede observarse, la desviación estándar estimada de los efectos aleatorios *Día* y *Cristal* es, respectivamente, 0.5579 y 0.0518, mientras que la desviación estándar residual, que es la desviación estándar estimada del modelo, es 0.1609. Por lo tanto, puesto que la desviación estándar estimada para el factor aleatorio *Día* es 3 veces superior a la de la desviación estándar residual, y que la del factor *Cristal* es un tercio de la residual, es evidente la necesidad de especificar ambos componentes aleatorios en la formulación del modelo (sobre todo el *Día*), o sea, de especificar tanto la variabilidad intra-cristal como intra-día en el modelo. Es decir, que hacer los experimentos en día diferentes y el hecho de que cada medida se haga sobre cristales diferentes introduce una gran variabilidad en los datos que es necesaria controlar para poder obtener unas inferencias más adecuadas y evitar enmascarar diferencias entre condiciones experimentales.

C.1.0.2. Estimaciones y comparaciones dos a dos del modelo lineal mixto Clásico

Aunque de carácter teórico, a continuación se presenta, paso a paso, a partir de los valores proporcionados por el modelo lineal (mixto o no), el cálculo de las estimaciones de cada una de las condiciones experimentales. En primer lugar, hay que tener presente cómo se han estimado las respuestas que, en este caso ha sido $\log(\text{AUC Ratio de Fluorescencia } 340/380)$ y para otros también el anterior o $\log(\text{Máximo Ratio de Fluorescencia } 340/380)$ de la señal de interés para cada una de las condiciones experimentales evaluadas. Por un lado, el modelo genera una serie de coeficientes que, salvo el intercepto, no se corresponden con el valor absoluto de la señal para ninguna de las condiciones y que, en este caso, son (tabla C.8): $\beta_{\text{Intercepto}}$, que estima la respuesta para la condición tomada como referencia, *NCM.Control*; β_{DFMO} , que es el efecto del cambio en la variable al usar *DFMO* en la línea celular *NCM* respecto de la condición de referencia *NCM.Control*; β_{HT29} , que es el efecto del cambio en la variable respuesta que existe entre las *HT29 control* y la condición

de referencia *NCM.Control*, y, por último, $\beta_{\text{HT29:DFMO}}$, que es el cambio en la respuesta entre la condición *HT29.DFMO* y la de referencia *NCM.Control* (tabla C.8).

Por otro lado, al haber sido estimados a la vez, estos coeficientes no son independientes entre sí y, por ende, existe una correlación entre ellos (tabla C.10), por lo que la varianza de la suma de dos coeficientes no es la suma de las varianzas de cada uno de ellos, sino que también interviene su covarianza (tabla C.10).

Tabla C.10: Covarianzas entre los coeficientes del modelo mixto para $\log(\text{AUC Depósitos})$.

| | β_0 | β_{HT29} | β_{DFMO} | $\beta_{\text{HT29:DFMO}}$ |
|----------------------------|------------|-----------------------|-----------------------|----------------------------|
| β_0 | 0.0782587 | -0.0001003 | -0.0001010 | 0.0001008 |
| β_{HT29} | -0.0001003 | 0.0002227 | 0.0001000 | -0.0002238 |
| β_{DFMO} | -0.0001010 | 0.0001000 | 0.0001926 | -0.0001924 |
| $\beta_{\text{HT29:DFMO}}$ | 0.0001008 | -0.0002238 | -0.0001924 | 0.0004379 |

Por ejemplo, para obtener la suma $\beta_0 + \beta_{\text{DFMO}}$, que estima $\log(\text{AUC Ratio de Fluorescencia } 340/380)$ ($\log(\text{Señal})$ para abreviar) para la condición experimental *NCM.DFMO*, no es $\text{Var}(\beta_0 + \beta_{\text{DFMO}}) = \text{Var}(\beta_0) + \text{Var}(\beta_{\text{DFMO}})$, sino que también hay que tener en cuenta la covarianza, $\text{Var}(\beta_0 + \beta_{\text{DFMO}}) = \text{Var}(\beta_0) + \text{Var}(\beta_{\text{DFMO}}) + 2\text{Cov}(\beta_0, \beta_{\text{DFMO}})$. Es decir, que la estimación de la respuesta, ($\log(\text{Señal})$), para las líneas celulares NCM460 tratadas con DFMO, $\log(\text{Señal}_{\text{NCM.DFMO}} = 12.1131 - 0.06151 = 12.05159$, $\text{Var}_{\text{NCM.DFMO}} = 0.078258 + 0.0001926 + 2 \cdot (-0.0001010) = 0.0782486$ y, por lo tanto, $\log(\text{Señal}_{\text{NCM.DFMO}} = 12.05159 \pm Z_{0.025} \sqrt{0.0782486}$. Siguiendo el mismo procedimiento, se obtienen las estimaciones para el resto de las estimaciones de las ($\log(\text{Señal})$) para condición experimental, tanto puntuales como sus intervalos de confianza (tabla C.11):

Tabla C.11: Estimaciones de la variable respuesta $\log(\text{AUC Depósitos})$ con el modelo lineal mixto seleccionado.

| Condiciones | $\log(\text{Señal})$ | Límite Superior | Límite Inferior | $\log(\text{Señal-SEM})$ | $\log(\text{Señal+SEM})$ |
|---------------------|----------------------|-----------------|-----------------|--------------------------|--------------------------|
| NCM.Control | 12.11307 | 12.66138 | 11.56477 | 11.83333 | 12.39282 |
| HT29.Control | 11.92598 | 12.47437 | 11.37760 | 11.64620 | 12.20577 |
| NCM.DFMO | 12.05156 | 12.59983 | 11.50329 | 11.77183 | 12.33129 |
| HT29.DFMO | 12.11796 | 12.66634 | 11.56958 | 11.83818 | 12.39774 |

A partir de la representación gráfica de los valores anteriores, puede parecer que las diferentes condiciones experimentales no son diferentes entre sí cuando, en realidad, sí lo son (salvo *NCM.Control* y *HT29.DFMO*). No obstante, es posible entender cómo sí lo son a partir de las restas dos a dos de las estimaciones ($\log(\text{Señal})$) de las diferentes condiciones. Así, al igual que la varianza de la suma de dos variables aleatorias es $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y)$, para la resta sucede algo similar, $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2 \cdot \text{Cov}(X, Y)$. Por ello, la diferencia entre $\log(\text{Señal}_{\text{NCM.DFMO}})$ y $\log(\text{Señal}_{\text{NCM.Control}})$ es $12.05159 - 12.1131 = -0.06151$ y su varianza se puede calcular como se muestra a continuación (ecuaciones C.2 a C.5).

En primer lugar, como se ha dicho, se tiene que la varianza de la resta de dos variables aleatorias es (ecuación C.2)¹:

$$\begin{aligned}
 \text{Var}(X - Y) &= E\{[X - Y - E(X - Y)]^2\} \\
 &= E\{[(X - E(X)) - (Y - E(Y))]^2\} \\
 &= E\{[(X - E(X))^2 + (Y - E(Y))^2 - 2(X - E(X))(Y - E(Y))]\} \\
 &= E[(X - E(X))^2] + E[(Y - E(Y))^2] - 2E[(X - E(X))(Y - E(Y))] \\
 &= \text{Var}(X) + \text{Var}(Y) - 2 \cdot \text{Cov}(X, Y) = \\
 &= \text{Var}(\beta_0) + [\text{Var}(\beta_0) + \text{Var}(\beta_{\text{DFMO}}) + 2 \cdot \text{Cov}(\beta_0, \beta_{\text{DFMO}})] + 2 \cdot \text{Cov}(X, Y)
 \end{aligned} \tag{C.2}$$

donde, para simplificar la notación, se ha sustituido $\log(\text{Señal}_{\text{NCM.DFMO}})$ por X y $\log(\text{Señal}_{\text{NCM.Control}})$ por Y , y los valores β_i son los coeficientes del modelo mostrados en la tabla C.8. Además, se tiene que $\text{Cov}(X, Y)$ es (ecuación C.3):

$$\begin{aligned}
 \text{Cov}(X, Y) &= E\{[X - E(X)] \cdot [Y - E(Y)]\} \\
 &= E\{[\beta_0 - E(\beta_0)] \cdot [\beta_{\text{DFMO}} + \beta_0 - E(\beta_{\text{DFMO}}) - E(\beta_0)]\} = \\
 &= E\{[\beta_0 - E(\beta_0)] \cdot [\beta_{\text{DFMO}} - E(\beta_{\text{DFMO}}) + \beta_0 - E(\beta_0)]\} = \\
 &= E\{[\beta_0 - E(\beta_0)] \cdot [\beta_0 - E(\beta_0)] + [\beta_{\text{DFMO}} - E(\beta_{\text{DFMO}})] \cdot [\beta_{\text{DFMO}} - E(\beta_{\text{DFMO}})]\} \\
 &= E\{[\beta_0 - E(\beta_0)] \cdot [\beta_0 - E(\beta_0)]\} + E\{[\beta_{\text{DFMO}} - E(\beta_{\text{DFMO}})] \cdot [\beta_{\text{DFMO}} - E(\beta_{\text{DFMO}})]\} \\
 &= \text{Var}(\beta_0) + \text{Cov}(\beta_0, \beta_{\text{DFMO}})
 \end{aligned} \tag{C.3}$$

y combinando las ecuaciones C.2 y C.3 se tiene que (ecuación C.4):

$$\begin{aligned}
 \text{Var}(X - Y) &= \text{Var}(\beta_0) + [\text{Var}(\beta_0) + \text{Var}(\beta_{\text{DFMO}}) + 2 \cdot \text{Cov}(\beta_0, \beta_{\text{DFMO}})] + 2 \cdot \text{Cov}(X, Y) \\
 &= \cancel{\text{Var}(\beta_0)} + \cancel{\text{Var}(\beta_0)} + \text{Var}(\beta_{\text{DFMO}}) + \\
 &\quad 2 \cdot \cancel{\text{Cov}(\beta_0, \beta_{\text{DFMO}})} + 2 \cdot [\cancel{\text{Var}(\beta_0)} + \cancel{\text{Cov}(\beta_0, \beta_{\text{DFMO}})}] = \\
 &= \text{Var}(\beta_{\text{DFMO}})
 \end{aligned} \tag{C.4}$$

es decir, que $\text{Var}(\log(\text{Señal}_{\text{NCM.DFMO}}) - \log(\text{Señal}_{\text{NCM.Control}})) = \text{Var}(\beta_{\text{DFMO}})$ y, por ende, el contraste de hipótesis $H_0 : X - Y = 0$ (ecuación C.5):

$$Z = \frac{\log(\text{Señal}_{\text{NCM.DFMO}}) - \log(\text{Señal}_{\text{NCM.Control}})}{\sqrt{\text{Var}(\beta_{\text{DFMO}})}} = \frac{12.05159 - 12.1131}{\sqrt{0.0001926}} = -4.432 \tag{C.5}$$

y, puesto que $Z \sim (0, 1)$, el p-valor asociado al contraste de hipótesis es $3.1 \cdot 10^{-6}$; es decir, que la diferencia entre las condiciones NCM.DFMO y NCM.Control es significativa para un nivel de significación de $\alpha = 0.05$. Siguiendo este mismo procedimiento es posible obtener los contrastes de las diferencias dos a dos entre las diferentes condiciones experimentales (tabla C.12):

¹De forma general, siguiendo un procedimiento similar, se puede demostrar que la varianza de la suma de n variables aleatorias es $\text{Var}(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j \text{Cov}(X_i, X_j)$.

Tabla C.12: Diferencias dos a dos entre las estimaciones de las diferentes condiciones experimentales, $\log(\text{AUC Depósitos})$, con el modelo lineal mixto seleccionado.

| | Diferencias | SEM | t | pvalor | asterisco |
|----------------------------|-------------|-----------|-------------|-----------|-----------|
| HT29.Control - NCM.Control | -0.1870926 | 0.0149219 | -12.5381331 | 0.0000000 | * |
| NCM.DFMO - NCM.Control | -0.0615153 | 0.0138772 | -4.4328465 | 0.0000642 | * |
| HT29.DFMO - NCM.Control | 0.0048845 | 0.0148618 | 0.3286625 | 0.9877294 | |
| NCM.DFMO - HT29.Control | 0.1255774 | 0.0146729 | 8.5584640 | 0.0000000 | * |
| HT29.DFMO - HT29.Control | 0.1919772 | 0.0156773 | 12.2455802 | 0.0000000 | * |
| HT29.DFMO - NCM.DFMO | 0.0663998 | 0.0145974 | 4.5487557 | 0.0000249 | * |

donde, como puede apreciarse, todas las comparaciones dos a dos en las que se ha usado el método de Tukey para corregir el p-valor, resultan ser significativas (rechazan H_0) para un nivel de significación $\alpha = 0.05$, a excepción de *HT29.DFMO - NCM.Control*.

C.1.0.3. Estimación y comparaciones dos a dos del modelo lineal mixto mediante *Boostrapping* paramétrico

En el apartado anterior, se ha empleado un enfoque clásico para la modelización de la variable respuesta. Sin embargo, hay ocasiones en las que es necesario emplear otro tipo de enfoque más robustos debido a la violación de alguno de los supuestos de los modelos. En concreto, en este caso no son preocupantes las violaciones de los supuestos, puesto que no es clara la violación de ninguno de ellos. Por lo tanto, no sería necesario recurrir a este tipo de métodos. No obstante, para asegurar la robustez de estos métodos ante pequeñas violaciones de los supuestos y asegurar la fiabilidad de las estimaciones y contrastes de hipótesis realizados, también se ha empleado métodos de *boostrapping* paramétrico para estimar el modelo lineal mixto seleccionado y realizar los contrastes de hipótesis.

A grandes rasgos, lo que se hace en este tipo de métodos es, en primer lugar, estimar el modelo anterior (clásico), a cuyas estimaciones se llamarán t^0 y, posteriormente, bajo el supuesto de que la muestra es una muestra representativa de la población, se han llevado a cabo remuestreos de un tamaño muestral igual al original; es decir, que sobre la muestra original de N elementos, se han tomado N elementos con reemplazo (un elemento puede ser escogido más de una vez y otros no ser escogidos) y se ha estimado el modelo lineal mixto clásico sobre esta nueva muestra. Así, este procedimiento se repite una cantidad elevada de veces que, en este caso, ha sido 1000. Obviamente, realizar este tipo de estimaciones requiere, por lo general, de un elevado número de observaciones originales.

Tras estimar 1000 veces los coeficientes con mil muestras diferentes procedentes de la muestra original, se ha generado la hipótesis nula de cada uno de los parámetros a estimar (coeficientes del modelo, diferencias dos a dos...). Por lo tanto, es posible estimar si los parámetros originales (t^0) son o no significativos. Para ello, por ejemplo, se toman los $\hat{\beta}_0^{(i)}$ donde el superíndice (*i*) indica el remuestreo $i \in \{1, 2, \dots, 1000\}$, y se obtienen los estadísticos para cada remuestreo (por ejemplo, el

estadístico *t* de *Student*) $t^{(i)}$ para obtener la distribución de la hipótesis nula sin necesidad de suponer ninguna. A continuación, se calcula el estadístico correspondiente a $\hat{\beta}_0$ para el modelo original (el clásico, sin remuestreo), t^0 , y se evalúa cómo de lejos está de la distribución nula (figura C.5). En cuanto al p valor asociado, se obtiene como (ecuación C.6):

$$p - \text{valor} = 2 \cdot \min \{ \# \{ t^i \leq t^0 \}, \# \{ t^i \geq t^0 \} \} \quad (\text{C.6})$$

que, posteriormente, será corregido mediante algún método de comparaciones múltiples como Bonferroni. Cabe destacar que los coeficientes mostrados son los coeficientes del modelo reparametrizado donde en vez de usar los factores por separado, se ha construido un nuevo factor que es la combinación de los niveles de ambos y que se denominará *Combinación*; es decir, que ahora el modelo tiene un solo factor cuyos niveles son: *NCM.Control* (referencia, β_0), *HT29.Control* ($\beta_{\text{HT29.Control}}$), *NCM.DFMO* ($\beta_{\text{NCM.DFMO}}$) y *HT29.DFMO* ($\beta_{\text{HT29.DFMO}}$). No obstante, a excepción del coeficiente $\beta_{\text{HT29.DFMO}}$, el resto de coeficientes tienen el mismo significado que los del modelo completo que usa los factores por separado y su interacción ($\beta_{\text{HT29.Control}} = \beta_{\text{HT29}}$ y $\beta_{\text{NCM.DFMO}} = \beta_{\text{DFMO}}$). Así, en la figura C.5 se muestra cómo se han obtenido los p valores de los diferentes contrastes de hipótesis de interés y en las tablas C.14 y C.13 los resultados numéricos correspondientes a dichos contrastes.

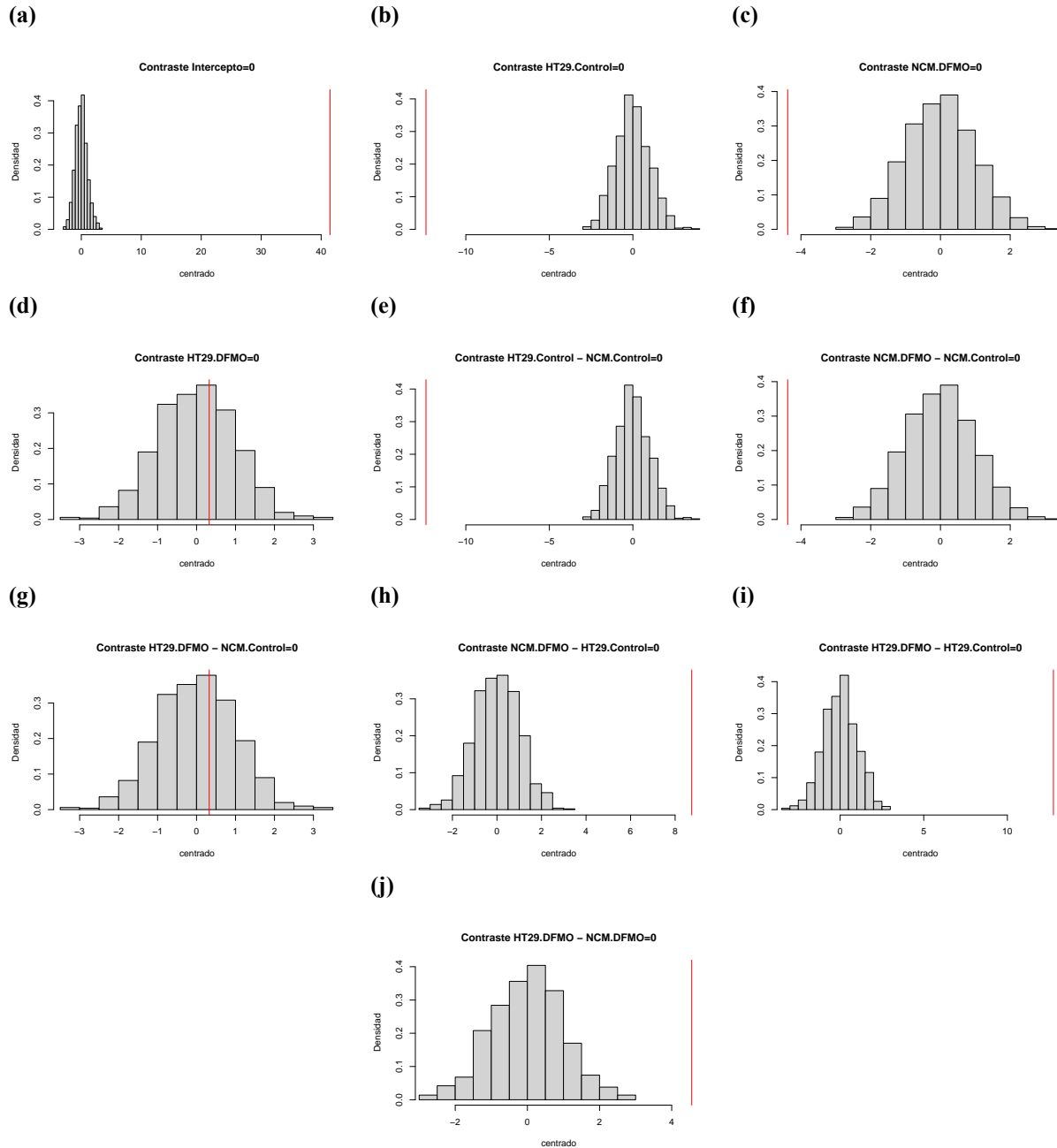


Figura C.5: Contrastes de hipótesis mediante *Bootstrapping*. Se muestran las distribuciones nulas, con ayuda de un histograma, de cada uno de los estadísticos correspondientes a los contrastes de hipótesis de interés, y el estadístico original se representa en cada uno de ellos mediante una línea vertical roja para evidenciar cómo de cerca o lejos está de dicha distribución. Los contrastes de hipótesis evaluados son: **a)** $B_0 = 0$; **b)** $B_{\text{HT29.Control}} = 0$; **c)** $B_{\text{NCM.DFMO}} = 0$; **d)** $B_{\text{HT29.DFMO}} = 0$; **e)** $\text{HT29.Control} - \text{NCM.Control} = 0$; **f)** $\text{NCM.DFMO} - \text{NCM.Control} = 0$; **g)** $\text{HT29.DFMO} - \text{NCM.Control} = 0$; **h)** $\text{NCM.DFMO} - \text{HT29.Control} = 0$; **i)** $\text{HT29.DFMO} - \text{HT29.Control} = 0$; **j)** $\text{HT29.DFMO} - \text{NCM.DFMO} = 0$.

Como puede observarse en las tablas C.13 y C.14, las estimaciones son muy similares a las del modelo clásico, salvo la interacción entre los factores fijos *Línea* y *Tratamiento*, lo que puede ser debido a la ligera violación de la normalidad de los residuos. Además, las estimaciones de las SEM son ligeramente superiores, pero prácticamente idénticas, lo que puede deberse a la ligera violación de los supuestos, si bien es cierto que no es preocupante. Además, puesto que se tiene un tamaño muestras muy grande (más de 1000 células) el *Teorema Central de Límite* asegura que la distribución de cualquier variable aleatoria, aunque no sea normal, su media sí lo será. En cuanto a las comparaciones dos a dos, los resultados obtenidos también son prácticamente un calco de las obtenidas mediante los métodos clásicos de estimación de los modelos lineales mixtos.

Tabla C.13: Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto *Bootstrap* para $\log(\text{AUC Depósitos})$.

| | Estimación | SEM | t | p-valor |
|-------------------------------|------------|----------|-----------|---------|
| β_0 | 12.12085 | 0.29204 | 41.50379 | 0 |
| $\beta_{\text{HT29.Control}}$ | -0.18723 | 0.015106 | -12.39561 | 0 |
| $\beta_{\text{NCM.DFMO}}$ | -0.06161 | 0.014044 | -4.3871 | 0 |
| $\beta_{\text{HT29.DFMO}}$ | 0.00483 | 0.015019 | 0.32139 | 0.784 |

Tabla C.14: Diferencias dos a dos entre las estimaciones obtenidas para las diferentes condiciones experimentales mediante el del modelo lineal mixto *Bootstrap* para $\log(\text{AUC Depósitos})$.

| | Diferencias | SEM | pvalor |
|-----------------------------------|-------------|-----------|----------|
| HT29.Control - NCM.Control | -0.1872265 | 0.0151055 | 0 * |
| NCM.DFMO - NCM.Control | -0.0616140 | 0.0140444 | 0.0001 * |
| HT29.DFMO - NCM.Control | 0.0048270 | 0.0150191 | 1 |
| NCM.DFMO - HT29.Control | 0.1256125 | 0.0143603 | 0 * |
| HT29.DFMO - HT29.Control | 0.1920534 | 0.0150540 | 0 * |
| HT29.DFMO - NCM.DFMO | 0.0664410 | 0.0145784 | 0 * |

C.2. Modelado de la variable respuesta *Incremento Máximo* de la señal correspondiente a los depósitos

Con el mismo procedimiento llevado a cabo en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos”, se han empleado modelos lineales mixtos estimados tanto mediante métodos clásicos como *bootstrapping* paramétrico para analizar el efecto del DFMO sobre el tamaño de los depósitos de Ca^{2+} intracelular y cómo de diferente es este efecto entre las líneas celulares de interés. En concreto, se ha modelizado el incremento máximo para la señal de los depósitos de cada una de las células registradas (MaxDepositos). Cabe destacar que los posibles factores aleatorios también son los mismos que en el caso de *AUCDepósitos* (apéndice “C.1”). En la tabla C.15 se muestra el valor de cAIC para cada una de las estructuras aleatorias evaluadas, donde la estructura aleatoria se expresa según el convenio usado en el paquete lme4 de R:

Tabla C.15: Selección de modelos para la variable respuesta $\log(\text{MaxDepositos})$ con cAIC.

| modelo | Parte Aleatoria (Anotación según lme4) | cAIC |
|--------|---|-------------|
| C_DTiL | (1 Dia:Cristal) + (1 + Tratamiento * Linea Dia) | error |
| C_DTL | (1 Dia:Cristal) + (1 + Tratamiento + Linea Dia) | -599.48554 |
| C_DT | ((1 Dia:Cristal) + (1 + Tratamiento Dia) | -472.8148 |
| C_DL | (1 Dia:Cristal) + (1 + Linea Dia) | -573.73846 |
| C_D | (1 Dia:Cristal) + (1 Dia) | -440.67746 |
| C | (1 Cristal) | 54.16462 |
| DTiL | (1 + Tratamiento * Linea Dia) | error |
| DTL | (1 + Tratamiento + Linea Dia) | -533.83328 |
| DT | (1 + Tratamiento Dia) | -405.8868 |
| DL | (1 + Linea Dia) | -520.26790 |
| D | (1 Dia) | -385.637851 |
| Fijo | | 80.0886 |

Como puede comprobarse, el valor más pequeño de cAIC es para el modelo cuasi sobreóptimo modelo DTL. No obstante, algunos de los modelos con menor cAIC estiman unas matrices de varianzas-covarianzas para la parte aleatoria peligrosas en el sentido de que hay efectos aleatorios con varianzas igual a cero (o su límite inferior es cero) y correlaciones entre efectos ± 1 (o al menos uno de sus límites inferior y/o superior). En consecuencia, estos modelos se descartan (Barr et al. (2013), Matuschek et al. (2017)). Así, de los modelos evaluados (en orden creciente de cAIC), en las tablas C.16-C.18 se muestran las estimaciones de la varianza de los diferentes efectos aleatorios y sus correlaciones (y sus intervalos de confianza) para aquellos modelos con un cAIC inferior al modelo seleccionado en los que se han identificado algunos de estos problemas.

Tabla C.16: Parte aleatoria de los modelos problemáticos para log(MaxDepositos). Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza. Con una x se indican las estimaciones problemáticas.

```
#Modelo C_DTL
Linear mixed model fit by REML ['lmerMod']
Formula: logMaxDepositos ~ Linea * Tratamiento + (1 | Dia:Cristal) +
          + (1 + Tratamiento + Linea | Dia)

Data: datosr
REML criterion at convergence: -554.5987
Random effects:
Groups      Name          Std.Dev. Corr
Dia:Cristal (Intercept)  0.07229
Dia         (Intercept)  0.23399
           TratamientoDFMO 0.08301  0.75
           LineaHT29       0.16905 -0.97 -0.64
Residual                    0.17653

          2.5 %      97.5 %  Problemática
.sig01      0.03826991  0.13883175
.sig02      0.10380044  0.50821325
.sig03     -0.20406952  0.93713379      x
.sig04     -1.00000000 -0.64302730      x
.sig05      0.03374252  0.18285146
.sig06     -0.95588415  0.34944998      x
.sig07      0.07313910  0.36255965
.sigma      0.16885863  0.18459530

#Modelo C_DL
Formula: logMaxDepositos ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 + Linea | Dia)
Data: datosr
REML criterion at convergence: -529.2044
Random effects:
Groups      Name          Std.Dev. Corr
Dia:Cristal (Intercept)  0.06775
Dia         (Intercept)  0.27429
           LineaHT29     0.17479 -0.95
Residual                    0.17929

          2.5 %      97.5 %  Problemática
.sig01      0.03570520  0.157080629
.sig02      0.12472205  0.590372207
.sig03     -1.00000000 -0.576995932      x
.sig04      0.08335489  0.374554818
.sigma      0.17142739  0.187368680
```

Tabla C.17: Parte aleatoria de los modelos problemáticos para $\log(\text{MaxSOCE})$. Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza. Con una x se indican las estimaciones problemáticas. (continuación)

```
#Modelo DTL
Formula: logMaxDepositos ~ Linea * Tratamiento + (1 + Tratamiento + Linea | Dia)
Data: datosr
REML criterion at convergence: -501.0831
Random effects:
Groups   Name                Std.Dev. Corr
Dia      (Intercept)         0.24656
          TratamientoDFMO 0.06954  0.86
          LineaHT29      0.17439 -0.94 -0.69
Residual                               0.18275

          2.5 %      97.5 % Problemática
.sig01      0.12000327  0.52679877
.sig02     -0.37419523  0.99213974      x
.sig03     -1.00000000 -0.55230026      x
.sig04      0.02482184  0.15577686
.sig05     -0.94623975  0.22091669      x
.sig06      0.06222216  0.37412622
.sigma      0.17481978  0.19107330

#Modelo DL
Formula: logMaxDepositos ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 + Linea | Dia)
Data: datosr
REML criterion at convergence: -529.2044
Random effects:
Groups   Name                Std.Dev. Corr
Dia:Cristal (Intercept) 0.06775
Dia      (Intercept) 0.27429
          LineaHT29    0.17479 -0.95
Residual                               0.17929

          2.5 %      97.5 % Problemática
.sig01      0.03570520  0.157080629
.sig02      0.12472205  0.590372207
.sig03     -1.00000000 -0.576995932      x
.sig04      0.08335489  0.374554818
.sigma      0.17142739  0.187368680
```

Tabla C.18: Parte aleatoria de los modelos problemáticos para $\log(\text{MaxSOCE})$. Se muestra la salida de R de la función `lmer()` usada para estimar el modelo lineal mixto y la función `confint(level = 0.95)` para los intervalos de confianza. Con una x se indican las estimaciones problemáticas. (*continuación*)

```
#Modelo C_DT
Formula: logMaxDepositos ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 +      Tratamiento | Dia)
Data: datosr
REML criterion at convergence: -431.535
Random effects:
Groups      Name                Std.Dev. Corr
Dia:Cristal (Intercept)      0.07756
Dia         (Intercept)      0.14654
              TratamientoDFMO 0.09653  0.78
Residual                    0.18869

              2.5 %      97.5 %  Problemática
.sig01          0.04115667  0.17893471
.sig02          0.02474243  0.33081812
.sig03         -0.22821871  1.00000000      x
.sig04          0.04068092  0.21143167
.sigma          0.18041762  0.19719884
```

En consecuencia, tras descartar los modelos problemáticos, el modelo que presenta un menor valor de cAIC y no presenta ninguno de los problemas considerados es el *modelo C_D* (tabla C.19).

Tabla C.19: Parte aleatoria del modelo seleccionado para MaxDepositos. Se muestra la salida de R de la función `lmer()` usada para estimar el modelo lineal mixto y la función `confint(level = 0.95)` para los intervalos de confianza.

```
#Modelo C_D
Linear mixed model fit by REML ['lmerMod']
Formula: logMaxDepositos ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 |      Dia)
Data: datosr
REML criterion at convergence: -401.7457
Random effects:
Groups      Name                Std.Dev.
Dia:Cristal (Intercept) 0.07245
Dia         (Intercept) 0.19249
Residual                    0.19215
Number of obs: 984, groups: Dia:Cristal, 8; Dia, 4
```

Posteriormente, mediante la AIC, se evalúa la parte fija que, al tratarse de un modelo de efectos fijos, es la única parte que contiene. Como resultado, en la tabla C.20 se muestra la selección de la parte fija que, como puede verse, resultan ser tanto los dos efectos principales y su interacción.

Tabla C.20: Selección de modelos para la variable respuesta $\log(\text{MaxDepositos})$ con AIC.

| Parte Fija | AIC |
|--|---------|
| Línea + Tratamiento + Línea:Tratamiento | -410.10 |
| Línea + Tratamiento | -191.27 |
| Línea | -107.11 |
| Tratamiento | -113.12 |
| nulo | -39.03 |

Tras evaluar el modelo sobreóptimo con el fin de encontrar tanto la parte fija como la parte aleatoria más adecuada, se obtuvo que el modelo más adecuado para modelizar MaxDepositos es (ecuación C.7):

$$\log(\text{MaxDepositos})_{ijdc} = \beta_0 + \beta_1 \text{Línea}_i + \beta_2 \text{Tratamiento}_j + \beta_3 \text{Tejido}_i \text{Tratamiento}_j + \mathbf{b}_{0,d} + \mathbf{b}_{0,c} + \epsilon_{ijdc} \quad (\text{C.7})$$

donde $\log(\text{MaxDepositos})_{ijdc}$ es el valor de la variable respuesta medido en la réplica r de la muestra perteneciente a la línea celular i (NCM460 o HT29), tratamiento j (Control o DFMO) medidas en el día d y en el cristal c . En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel NCM460.Control); β_1 es el cambio de expresión en la línea HT29 respecto del de referencia, β_2 el cambio en $\log(\text{MaxDepositos})$ en las células tratadas con DFMO respecto de las de referencia (NCM460.Control); β_3 el efecto de la interacción entre los factores *Línea* y *Tratamiento* que indica si el cambio entre células tratadas con DFMO y sin tratar es diferente en NCM460 que en HT29 (en caso de ser significativo el término de interacción); $\mathbf{b}_{0,d}$ es el intercepto aleatorio debido al día que indica la desviación respecto de β_0 que hay en el día d , y $\mathbf{b}_{0,c}$ es el intercepto aleatorio debido a cristal o cubreobjetos c que indica la desviación respecto de β_0 que se observa en el cristal c .

En cuanto al diagnóstico del modelo o la evaluación del cumplimiento de los supuestos, en la figura C.6 se muestran los métodos gráficos y en la tabla C.21 los métodos basados en contrastes de hipótesis. Como puede observarse, aunque en los métodos gráficos los supuestos de normalidad y homocedasticidad no parecen violarse de forma preocupante, los tests sugieren que la violación es considerable. Por ello, tras realizar las estimaciones y contrastes de hipótesis pertinentes con los modelos lineales mixtos, se emplearán las técnicas de estimación *bootstrapping* para evaluar si el modelo es o no adecuado en función de la semejanza entre ambos y, en caso de que no lo sean, se emplearán las estimaciones *bootstrapping*.

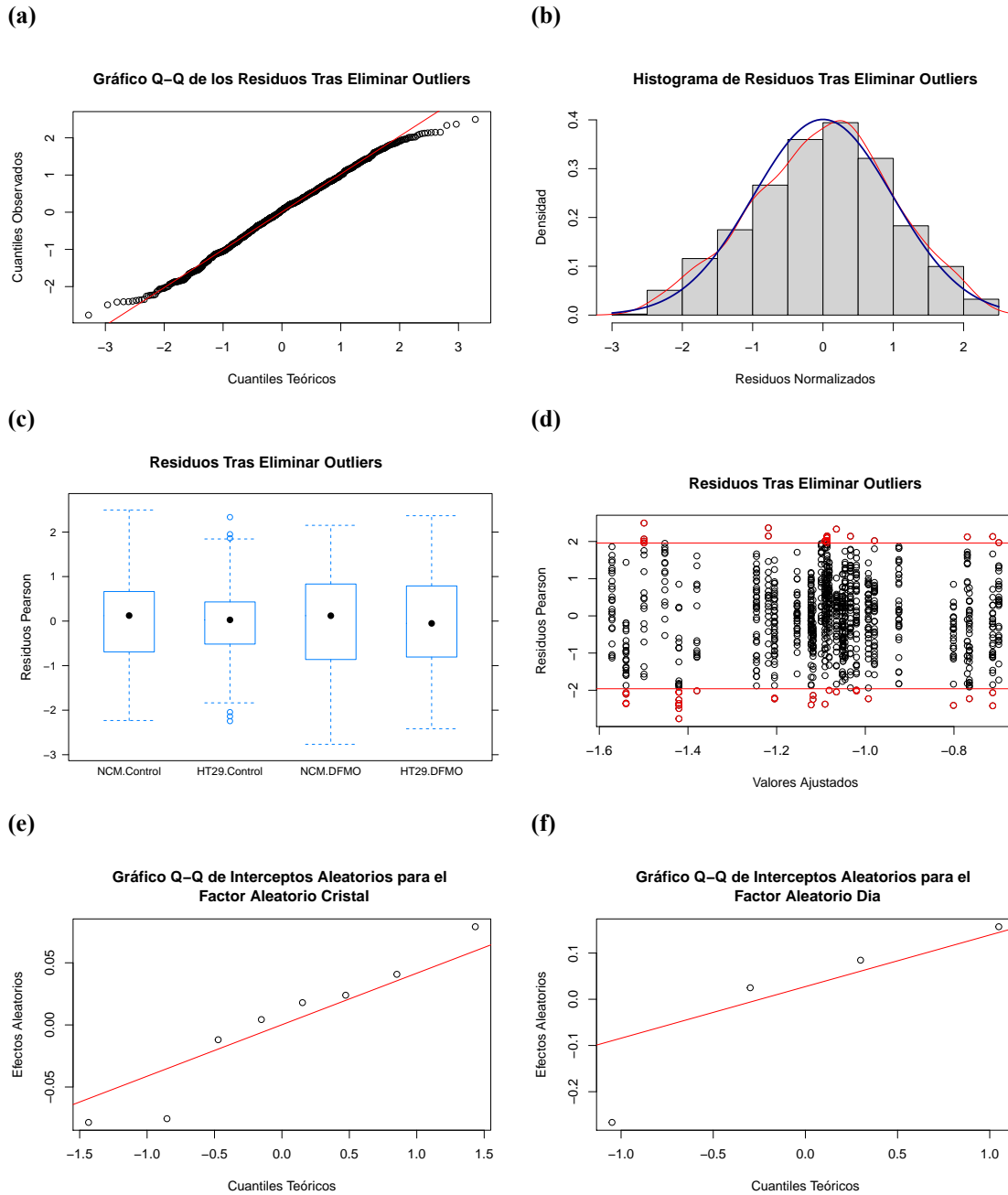


Figura C.6: Diagnóstico del modelo lineal de efectos fijos para la variable respuesta $\log(\text{MaxDepositos})$ tras eliminar los *outliers*. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como *outliers*; e) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Día, y f) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Cristal.

Tabla C.21: Supuestos del modelo seleccionado (p valores).

| | Normalidad | | | Homocedasticidad | |
|-------------------------------|------------|--------|-------|------------------|----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 0.00127 | 0.0108 | 0.042 | 0.000229 | 0 |
| Día:Cristal intercepto | 0.51300 | 0.8130 | 0.736 | | |
| Día intercepto | 0.32000 | 0.7310 | | | |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test. de Breush-Pagan.

En la tabla C.22 se muestran las estimaciones de los coeficientes del modelo y en la tabla C.23 su matriz de varianzas y covarianzas.

Tabla C.22: Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para log (MaxDepositos).

| | Estimación | SEM | t | gdl | p-valor |
|----------------------------|------------|-----------|------------|----------|-----------|
| β_0 | -1.1542145 | 0.1003209 | -11.505226 | 3.045013 | 0.0013147 |
| β_{HT29} | -0.0722167 | 0.0177397 | -4.070920 | 3.045013 | 0.0259992 |
| β_{DFMO} | -0.0405685 | 0.0165056 | -2.457857 | 3.045013 | 0.0897823 |
| $\beta_{\text{HT29:DFMO}}$ | 0.3932087 | 0.0249913 | 15.733809 | 3.045013 | 0.0005134 |

Abreviaturas:

gdl: grados de libertad

Tabla C.23: Covarianzas entre los coeficientes del modelo final para log (MaxDepositos).

| | β_0 | β_{HT29} | β_{DFMO} | $\beta_{\text{HT29:DFMO}}$ |
|----------------------------|------------|-----------------------|-----------------------|----------------------------|
| β_0 | 0.0100643 | -0.0001427 | -0.0001421 | 0.0001421 |
| β_{HT29} | -0.0001427 | 0.0003147 | 0.0001421 | -0.0003163 |
| β_{DFMO} | -0.0001421 | 0.0001421 | 0.0002724 | -0.0002732 |
| $\beta_{\text{HT29:DFMO}}$ | 0.0001421 | -0.0003163 | -0.0002732 | 0.0006246 |

En cuanto a la estructura de la parte aleatoria, la varianza de los efectos aleatorios se muestra en la tabla en la tabla C.24.

Tabla C.24: Estimaciones de la varianza de la parte aleatoria del modelo mixto final para log (MaxDepositos).

| Componente Aleatorio | LI | Estimación SEM | LS |
|---------------------------|-----------|----------------|-----------|
| Cristal Intercepto | 0.0381829 | 0.0724462 | 0.1827635 |
| Día Intercepto | 0.0601368 | 0.1924946 | 0.4207105 |
| Sigma | 0.1836472 | 0.1921491 | 0.2006900 |

La tabla ANOVA del modelo seleccionado se muestra en la tabla C.25. Como puede observarse, tanto según AIC como la tabla ANOVA indican que ambos efectos principales son singificativos o relevantes en el modelo, al igual que el término de interacción entre ambos.

Tabla C.25: Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para log (MaxDepositos).

| | Sum Sq | gdl | p-valor |
|--------------------------|----------|-----|---------|
| Linea | 104.2988 | 1 | 0 |
| Tratamiento | 112.9221 | 1 | 0 |
| Linea:Tratamiento | 247.5527 | 1 | 0 |

Abreviaturas:

gdl: grados de libertad

Las estimaciones de MaxDepositos se muestran en la tabla C.26.

Tabla C.26: Estimaciones de MaxDepositos con el modelo seleccionado.

| Condiciones | log(Señal) | Límite Superior | Límite Inferior | log(Señal-SEM) | log(Señal+SEM) |
|---------------------|------------|-----------------|-----------------|----------------|----------------|
| NCM.Control | 0.3153051 | 0.3838185 | 0.2590217 | 0.2852083 | 0.3485779 |
| NCM.DFMO | 0.3027697 | 0.3685167 | 0.2487526 | 0.2738855 | 0.3346999 |
| HT29.Control | 0.2933376 | 0.3571799 | 0.2409064 | 0.2652989 | 0.3243396 |
| HT29.DFMO | 0.4173663 | 0.5082132 | 0.3427589 | 0.3774682 | 0.4614816 |

Las comparaciones múltiples se muestran en la tabla C.27.

Tabla C.27: Diferencias dos a dos entre las condiciones experimentales de sus valores log (MaxDepositos) estimados con el modelo seleccionado.

| | Diferencias | SEM | t | pvalor | |
|-----------------------------------|-------------|-----------|-----------|-----------|---|
| HT29.Control - NCM.Control | -0.0722167 | 0.0177397 | -4.070920 | 0.0002882 | * |
| NCM.DFMO - NCM.Control | -0.0405685 | 0.0165056 | -2.457857 | 0.0668665 | * |
| HT29.DFMO - NCM.Control | 0.2804234 | 0.0178014 | 15.752914 | 0.0000000 | * |
| NCM.DFMO - HT29.Control | 0.0316482 | 0.0174046 | 1.818387 | 0.2641149 | |
| HT29.DFMO - HT29.Control | 0.3526401 | 0.0187253 | 18.832296 | 0.0000000 | * |
| HT29.DFMO - NCM.DFMO | 0.3209919 | 0.0175101 | 18.331776 | 0.0000000 | * |

No obstante, puesto que el supuesto de homocedasticidad se viola en los modelos empleados, también se recurre a la estimación del modelo lineal mixto seleccionado mediante *bootstrapping* paramétrico. En la figura C.7 se muestran los gráficos de los contrastes de hipótesis, en la tabla C.28 las estimaciones de los coeficientes y los resultados del contraste de hipótesis, en la tabla C.29 la matriz de varianzas y covarianzas de los coeficientes estimados y en la tabla C.30 las diferencias

dos a dos que, como puede comprobarse, son similares a las obtenidas mediante el método clásico de estimación de los modelos lineales mixtos. Seguramente, se debe a la robustez del modelo a pequeñas desviaciones de los supuestos.

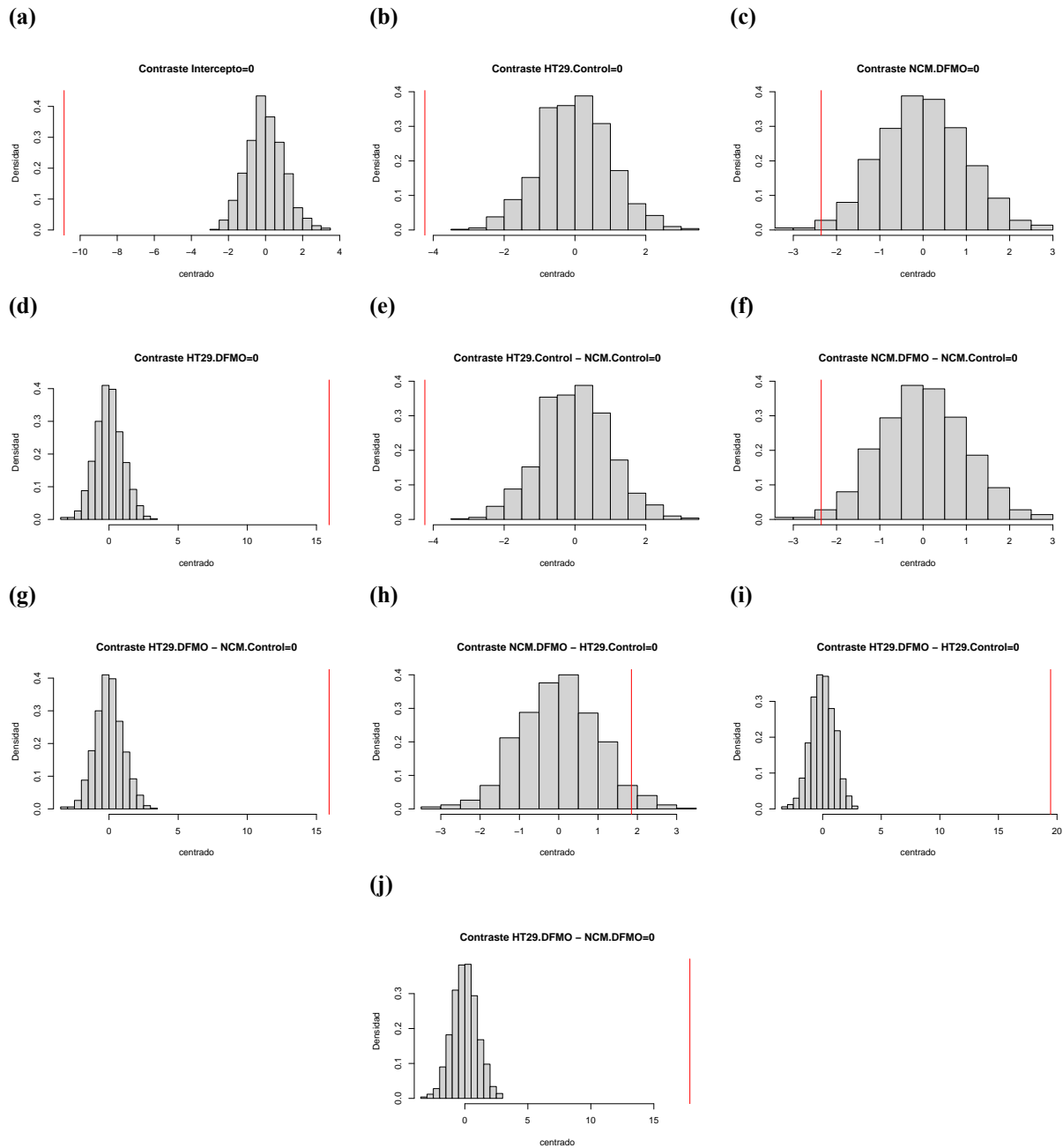


Figura C.7: Contrastes de hipótesis mediante *Boostrapping*. Se muestran las distribuciones nulas, con ayuda de un histograma, de cada uno de los estadísticos correspondientes a los contrastes de hipótesis de interés, y el estadístico original se representa en cada uno de ellos mediante una línea vertical roja para evidenciar cómo de cerca o lejos está de dicha distribución. Los contrastes de hipótesis evaluados son: **a)** $B_0 = 0$; **b)** $B_{\text{HT29.Control}} = 0$; **c)** $B_{\text{NCM.DFMO}} = 0$; **d)** $B_{\text{HT29.DFMO}} = 0$; **e)** $\text{HT29.Control} - \text{NCM.Control} = 0$; **f)** $\text{NCM.DFMO} - \text{NCM.Control} = 0$; **g)** $\text{HT29.DFMO} - \text{NCM.Control} = 0$; **h)** $\text{NCM.DFMO} - \text{HT29.Control} = 0$; **i)** $\text{HT29.DFMO} - \text{HT29.Control} = 0$; **j)** $\text{HT29.DFMO} - \text{NCM.DFMO} = 0$.

Tabla C.28: Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto *Bootstrap*.

| | Estimación | SEM | t | p-valor |
|-------------------------------|------------|-----------|------------|---------|
| β_0 | -1.1525820 | 0.1062359 | -10.849268 | 0 |
| $\beta_{\text{HT29.Control}}$ | -0.0725151 | 0.0170365 | -4.256445 | 0 |
| $\beta_{\text{NCM.DFMO}}$ | -0.0405220 | 0.0172276 | -2.352162 | 0.018 |
| $\beta_{\text{HT29.DFMO}}$ | 0.2804405 | 0.0176182 | 15.917677 | 0 |

Tabla C.29: Covarianzas entre los coeficientes del modelo mixto *Bootstrap*.

| | β_0 | β_{HT29} | β_{DFMO} | $\beta_{\text{HT29:DFMO}}$ |
|----------------------------|------------|-----------------------|-----------------------|----------------------------|
| β_0 | 0.0112861 | -0.0002617 | -0.0002282 | -0.0002539 |
| β_{HT29} | -0.0002617 | 0.0002902 | 0.0001471 | 0.0001360 |
| β_{DFMO} | -0.0002282 | 0.0001471 | 0.0002968 | 0.0001421 |
| $\beta_{\text{HT29:DFMO}}$ | -0.0002539 | 0.0001360 | 0.0001421 | 0.0003104 |

Tabla C.30: Diferencias dos a dos entre las condiciones experimentales del modelo lineal mixto *Bootstrap*.

| | Diferencias | SEM | pvalor | |
|-----------------------------------|-------------|-----------|--------|---|
| HT29.Control - NCM.Control | -0.0725151 | 0.0170365 | 0 | * |
| NCM.DFMO - NCM.Control | -0.0405220 | 0.0172276 | 0.108 | |
| HT29.DFMO - NCM.Control | 0.2804405 | 0.0176182 | 0 | * |
| NCM.DFMO - HT29.Control | 0.0319931 | 0.0171142 | 0.360 | |
| HT29.DFMO - HT29.Control | 0.3529557 | 0.0181270 | 0 | * |
| HT29.DFMO - NCM.DFMO | 0.3209626 | 0.0179721 | 0 | * |

C.3. Modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente a SOCE

Con el mismo procedimiento llevado a cabo en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos”, se han empleado modelos lineales mixtos estimados tanto mediante métodos clásicos como *bootstrapping* paramétrico para analizar el efecto del DFMO sobre SOCE y cómo de diferente es este efecto entre las líneas celulares de interés. En concreto, se ha modelizado el AUC de la señal de SOCE de cada una de las células registradas (AUCSOCE). Cabe destacar que los posibles factores aleatorios también son los mismos que en el caso de *AUCDepósitos* (apéndice “C.1”). En la tabla C.31 se muestra el valor de cAIC para cada una de las estructuras aleatorias evaluadas, donde la estructura aleatoria se expresa según el convenio usado en el paquete lme4 de R:

Tabla C.31: Selección de modelos para la variable respuesta $\log(\text{AUCSOCE})$ con cAIC.

| modelo | Parte Aleatoria (Anotación según lme4) | cAIC |
|---------------|---|----------|
| C_DTiL | (1 Dia:Cristal) + (1 + Tratamiento * Linea Dia) | error |
| C_DTL | (1 Dia:Cristal) + (1 + Tratamiento + Linea Dia) | 14653.34 |
| C_DT | ((1 Dia:Cristal) + (1 + Tratamiento Dia) | 14696.72 |
| C_DL | (1 Dia:Cristal) + (1 + Linea Dia) | 14662.36 |
| C_D | (1 Dia:Cristal) + (1 Dia) | 14710.51 |
| C | (1 Cristal) | error |
| DTiL | (1 + Tratamiento * Linea Dia) | error |
| DTL | (1 + Tratamiento + Linea Dia) | 14705.54 |
| DT | (1 + Tratamiento Dia) | 14742.65 |
| DL | (1 + Linea Dia) | 14719.23 |
| D | (1 Dia) | 14760.55 |
| Fijo | | 15801.75 |

Como puede comprobarse, el valor más pequeño de cAIC es para el modelo cuasi sobreóptimo modelo DTL. No obstante, algunos de los modelos con menor cAIC estiman unas matrices de varianzas-covarianzas para la parte aleatoria peligrosas en el sentido de que hay efectos aleatorios con varianzas igual a cero (o su límite inferior es cero) y correlaciones entre efectos ± 1 (o al menos uno de sus límites inferior y/o superior). En consecuencia, estos modelos se descartan (Barr et al. (2013), Matuschek et al. (2017)). Así, de los modelos evaluados (en orden creciente de cAIC), en las tablas C.32-C.33 se muestran las estimaciones de la varianza de los diferentes efectos aleatorios y sus correlaciones (y sus intervalos de confianza) para aquellos modelos con un cAIC inferior al modelo seleccionado en los que se han identificado algunos de estos problemas.

Tabla C.32: Parte aleatoria de los modelos problemáticos para log(AUCSOCE). Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza. Con una x se indican las estimaciones problemáticas.

```
#Modelo C_DTL
Linear mixed model fit by REML ['lmerMod']
Formula: logAreaSOCE ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 + Tratamiento + Linea | Dia)
Data: datosr
REML criterion at convergence: 14635.12
Random effects:
Groups      Name          Std.Dev. Corr
Dia:Cristal (Intercept) 153.7
Dia         (Intercept) 913.9
           TratamientoDFMO 127.9   -0.03
           LineaHT29      241.1   -0.83  0.37
Residual                    445.9

           2.5 %      97.5 %  Problemática
.sig01      80.9508581 388.43648691
.sig02     436.4425038 1956.07417286
.sig03     -0.7833471   0.74035642
.sig04     -1.0000000  -0.09271349      x
.sig05      38.4898451 294.38510578
.sig06     -0.7020757   1.00000000      x
.sig07      75.7521508 532.23080255
.sigma     426.3589291 466.31737366

#Modelo C_DL
Linear mixed model fit by REML ['lmerMod']
Formula: logAreaSOCE ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 + Linea | Dia)
Data: datosr
REML criterion at convergence: 14645.5
Random effects:
Groups      Name          Std.Dev. Corr
Dia:Cristal (Intercept) 159.0
Dia         (Intercept) 919.2
           LineaHT29    248.6   -0.80
Residual                    449.1

           2.5 %      97.5 %  Problemática
.sig01      83.8113945 399.89512314
.sig02     440.2734444 1965.49366301
.sig03     -0.9794918   0.07484729      x
.sig04     111.2127484 540.60278323
.sig05     429.3111528 469.47675900
```

Tabla C.33: Parte aleatoria de los modelos problemáticos para log(AUCSOCE). Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza. Con una x se indican las estimaciones problemáticas. (continuación)

```
#Modelo C_DT
Linear mixed model fit by REML ['lmerMod']
Formula: logAreaSOCE ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 + Tratamiento | Dia)
Data: datosr
REML criterion at convergence: 14677.81
Random effects:
Groups      Name                Std.Dev. Corr
Dia:Cristal (Intercept)    146.6
Dia         (Intercept)    815.6
           TratamientoDFMO 145.9    0.15
Residual                    457.1

                2.5 %      97.5 % Problemática
.sig01          76.3765760 371.9849044
.sig02          389.0055858 1746.3083019
.sig03          -0.9554856   0.9747381      x
.sig04          49.1805712  330.5163565
.sigma          436.9289352  477.8002725

#Modelo DTL
Linear mixed model fit by REML ['lmerMod']
Formula: logAreaSOCE ~ Linea * Tratamiento + (1 + Tratamiento + Linea | Dia)
Data: datosr
REML criterion at convergence: 14680.74
Random effects:
Groups      Name                Std.Dev. Corr
Dia         (Intercept)    910.7
           TratamientoDFMO 149.5    0.11
           LineaHT29      231.2   -0.84  0.16
Residual                    459.0

                2.5 %      97.5 % Problemática
.sig01          446.2523696 1938.62206475
.sig02          -0.7013244   0.85733386
.sig03          -1.0000000  -0.05625074      x
.sig04          52.5603372  338.60754150
.sig05          -0.7626554   1.00000000      x
.sig06          101.3897843  511.38238385
.sigma          438.9654108  480.02107800
```


En consecuencia, tras descartar los modelos problemáticos, el modelo que presenta un menor valor de cAIC y no presenta ninguno de los problemas considerados es el *modelo C_D* (tabla C.34).

Tabla C.34: Parte aleatoria del modelo seleccionado para AUCSOCE. Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza.

```
#Modelo C_D
Linear mixed model fit by REML ['lmerMod']
Formula: logAreaSOCE ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 | Dia)
Data: datosr
REML criterion at convergence: 14689.6
Random effects:
Groups      Name          Std.Dev.
Dia:Cristal (Intercept) 153.3
Dia         (Intercept) 833.3
Residual                                461.2
```

Posteriormente, mediante la AIC, se evalúa la parte fija que, al tratarse de un modelo de efectos fijos, es la única parte que contiene. Como resultado, en la tabla C.35 se muestra la selección de la parte fija que, como puede verse, resultan ser tanto los dos efectos principales y su interacción.

Tabla C.35: Selección de modelos para la variable respuesta log(AUCSOCE) con AIC.

| Parte Fija | AIC |
|--|-------|
| Linea + Tratamiento + Linea:Tratamiento | 14745 |
| Linea + Tratamiento | 14746 |
| Linea | 14747 |
| Tratamiento | 15212 |
| nulo | 15213 |

Como puede observarse, según el valor de AIC el modelo debería considerar tanto los factores *Linea* y *Tratamiento* como la interacción entre ambos. Tras evaluar el modelo sobreóptimo con el fin de encontrar tanto la parte fija como la parte aleatoria más adecuada, se obtuvo que el modelo más adecuado para modelizar AUCSOCE es (ecuación C.8):

$$T(\text{AUCSOCE})_{ijdc} = \beta_0 + \beta_1 \text{Linea}_i + \beta_2 \text{Tratamiento}_j + \beta_3 \text{Tejido}_i \text{Tratamiento}_j + b_{0,d} + b_{0,c} + \epsilon_{ijdc} \quad (\text{C.8})$$

donde $T(\text{AUCSOCE})_{ijdc}$ es el valor de la variable respuesta transformada según la familia de Box-Cox propuesto por Hawkins and Weisberg (2017), medido en la réplica r de la muestra perteneciente a la línea celular i (NCM460 o HT29), tratamiento j (Control o DFMO) medidas en el día d y en el cristal c . En cuanto a los parámetros: β_0 es el intercepto o término independiente del

modelo (referencia que, en este caso es el nivel NCM460.Control); β_1 es el cambio de expresión en la línea HT29 respecto del de referencia, β_2 el cambio en T (AUCSOCE) en las células tratadas con DFMO respecto de las de referencia (NCM460.Control); β_3 el efecto de la interacción entre los factores *Línea* y *Tratamiento* que indica si el cambio entre células tratadas con DFMO y sin tratar es diferente en NCM460 que en HT29 (en caso de ser significativo el término de interacción); $b_{0,d}$ es el intercepto aleatorio debido al día que indica la desviación respecto de β_0 que hay en el día d , y $b_{0,c}$ es el intercepto aleatorio debido a cristal o cubreobjetos c que indica la desviación respecto de β_0 que se observa en el cristal c .

En cuanto al diagnóstico del modelo o la evaluación del cumplimiento de los supuestos, en la tabla C.36 se muestran los métodos basados en contrastes de hipótesis y en la figura C.8 los métodos gráficos.

Tabla C.36: Supuestos del modelo seleccionado (p valores).

| | Normalidad | | | Homocedasticidad | |
|-------------------------------|------------|----------|-------|------------------|----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 1.98e-05 | 0.94e-04 | 0.484 | 0.6340 | 0.88e-04 |
| Día:Cristal intercepto | 0.806 | 0.870000 | 0.466 | | |
| Día intercepto | 0.139 | 0.699000 | | | |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test. de Breush-Pagan.

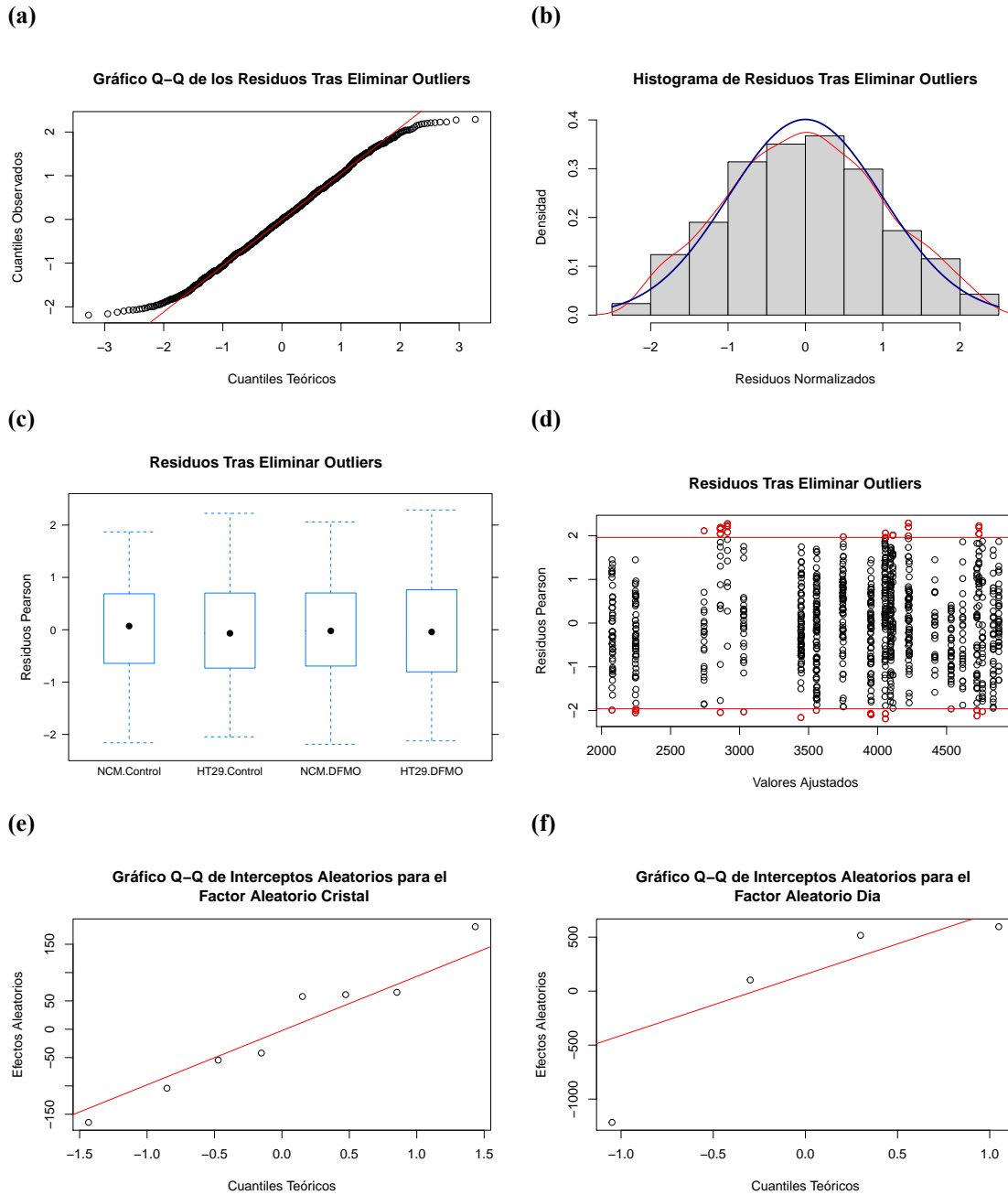


Figura C.8: Diagnóstico del modelo lineal de efectos fijos para la variable respuesta T (AUCSOCE) tras eliminar los outliers. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como outliers; e) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Día, y f) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Cristal.

En la tabla C.37 se muestran las estimaciones de los coeficientes del modelo y en la tabla C.38 su matriz de varianzas y covarianzas.

Tabla C.37: Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para T (AUCSOCE).

| | Estimación | SEM | t | gdl | p-valor |
|----------------------------|-------------|-----------|------------|----------|-----------|
| β_0 | 3395.583909 | 427.24712 | 7.9475876 | 3.008217 | 0.0041140 |
| β_{HT29} | 784.953275 | 40.24713 | 19.5033365 | 3.008217 | 0.0002895 |
| β_{DFMO} | 3.745256 | 36.80323 | 0.1017643 | 3.008217 | 0.9253478 |
| $\beta_{\text{HT29:DFMO}}$ | -121.717634 | 56.88345 | -2.1397724 | 3.008217 | 0.1216328 |

Abreviaturas:

gdl: grados de libertad

Tabla C.38: Covarianzas entre los coeficientes del modelo final para T (AUCSOCE).

| | β_0 | β_{HT29} | β_{DFMO} | $\beta_{\text{HT29:DFMO}}$ |
|----------------------------|-------------|-----------------------|-----------------------|----------------------------|
| β_0 | 182540.1030 | -707.9590 | -710.2965 | 715.5854 |
| β_{HT29} | -707.9590 | 1619.8313 | 705.6459 | -1625.2690 |
| β_{DFMO} | -710.2965 | 705.6459 | 1354.4779 | -1355.2296 |
| $\beta_{\text{HT29:DFMO}}$ | 715.5854 | -1625.2690 | -1355.2296 | 3235.7269 |

En cuanto a la estructura de la parte aleatoria, la varianza de los efectos aleatorios se muestra en la tabla en la tabla C.39.

Tabla C.39: Estimaciones de la varianza de la parte aleatoria del modelo mixto final para T (AUCSOCE).

| Componente Aleatorio | LI | Estimación SEM | LS |
|---------------------------|-----------|----------------|-----------|
| Cristal Intercepto | 80.05865 | 151.3094 | 380.8958 |
| Día Intercepto | 405.73070 | 846.0875 | 1808.8130 |
| Sigma | 404.76977 | 424.0069 | 443.3354 |

La tabla ANOVA del modelo seleccionado con interacción se muestra en la tabla C.40. Aunque el factor *Tratamiento* no es significativo, por lo menos para un nivel de significación de 0.05, sí lo es al 0.1, aunque sí lo es al 0.05 el tanto el factor *Línea* como la interacción entre ambos factores. Esto, además, coincide con la selección del modelo mediante AIC.

Tabla C.40: Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para T (AUCSOCE).

| | Sum Sq | gdl | p-valor |
|--------------------------|------------|-----|-----------|
| Linea | 652.052756 | 1 | 0 |
| Tratamiento | 2.835389 | 1 | 0.0922085 |
| Linea:Tratamiento | 4.578626 | 1 | 0.0323732 |

Abreviaturas:

gdl: grados de libertad

\begin{table}[H]

Las estimaciones de AUCSOCE, expresado como T(AUCSOCE), se muestran en la tabla C.41.

Tabla C.41: Estimaciones de AUCSOCE con el modelo seleccionado.

| Condiciones | log(Señal) | Límite Superior | Límite Inferior | log(Señal-SEM) | log(Señal+SEM) |
|---------------------|------------|-----------------|-----------------|----------------|----------------|
| NCM.Control | 75956.38 | 104506.3 | 50414.84 | 62521.78 | 90171.54 |
| HT29.Control | 102636.98 | 133706.3 | 74248.91 | 87798.72 | 118171.69 |
| NCM.DFMO | 76077.68 | 104634.8 | 50526.03 | 62638.35 | 90296.80 |
| HT29.DFMO | 98470.82 | 129184.5 | 70482.32 | 83830.00 | 113819.41 |

Puesto que el el factor de interacción es significativo y que según AIC ha de estar en el modelo, se estiman las diferencias dos a dos del modelo con interacción (tabla C.42).

Tabla C.42: Diferencias dos a dos entre las condiciones experimentales de sus valores T (AUCSOCE) estimados con el modelo seleccionado.

| | Diferencias | SEM | t | pvalor | |
|-----------------------------------|-------------|----------|-------------|-----------|---|
| HT29.Control - NCM.Control | 784.953275 | 40.24713 | 19.5033365 | 0.0000000 | * |
| NCM.DFMO - NCM.Control | 3.745256 | 36.80323 | 0.1017643 | 0.9996216 | |
| HT29.DFMO - NCM.Control | 666.980896 | 40.74716 | 16.3687716 | 0.0000000 | * |
| NCM.DFMO - HT29.Control | -781.208019 | 39.53501 | -19.7599020 | 0.0000000 | * |
| HT29.DFMO - HT29.Control | -117.972378 | 43.35603 | -2.7210141 | 0.0332317 | * |
| HT29.DFMO - NCM.DFMO | 663.235641 | 40.06270 | 16.5549398 | 0.0000000 | * |

No obstante, puesto que el supuesto de homocedasticidad se viola en los modelos empleados, también se recurre a la estimación del modelo lineal mixto seleccionado mediante *bootstrapping* paramétrico. En la figura C.9 se muestran los gráficos de los contrastes de hipótesis, en la tabla C.43 las estimaciones de los coeficientes y los resultados del contraste de hipótesis, en la tabla C.44 la matriz de varianzas y covarianzas de los coeficientes estimados y en la tabla C.45 las diferencias dos a dos que, como puede comprobarse, son similares a las obtenidas mediante el método clásico

de estimación de los modelos lineales mixtos. Seguramente, se debe a la robustez del modelo a pequeñas desviaciones de los supuestos.

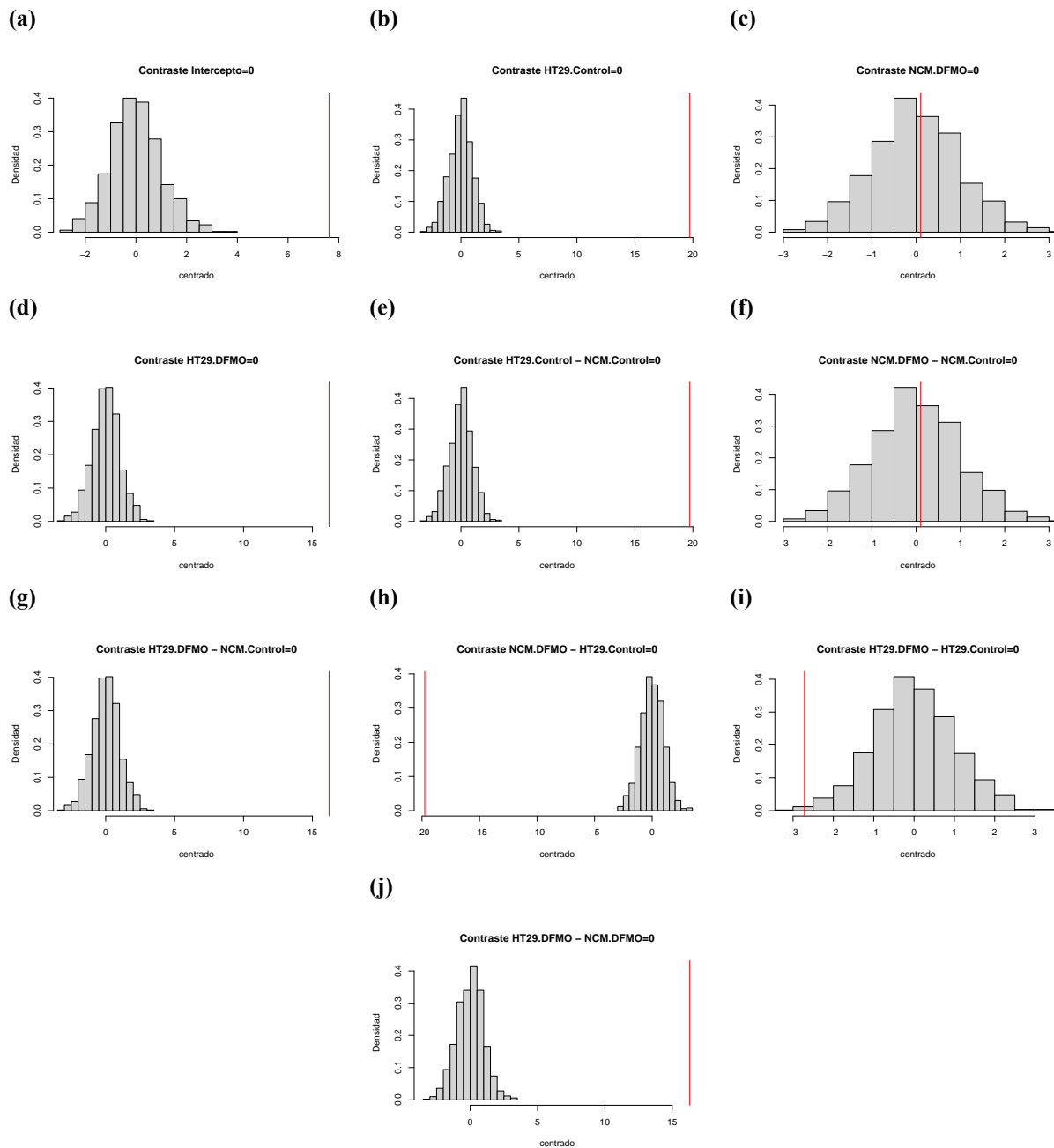


Figura C.9: Contrastes de hipótesis mediante *Bootstrapping*. Se muestran las distribuciones nulas, con ayuda de un histograma, de cada uno de los estadísticos correspondientes a los contrastes de hipótesis de interés, y el estadístico original se representa en cada uno de ellos mediante una línea vertical roja para evidenciar cómo de cerca o lejos está de dicha distribución. Los contrastes de hipótesis evaluados son: **a)** $B_0 = 0$; **b)** $B_{\text{HT29.Control}} = 0$; **c)** $B_{\text{NCM.DFMO}} = 0$; **d)** $B_{\text{HT29.DFMO}} = 0$; **e)** $\text{HT29.Control} - \text{NCM.Control} = 0$; **f)** $\text{NCM.DFMO} - \text{NCM.Control} = 0$; **g)** $\text{HT29.DFMO} - \text{NCM.Control} = 0$; **h)** $\text{NCM.DFMO} - \text{HT29.Control} = 0$; **i)** $\text{HT29.DFMO} - \text{HT29.Control} = 0$; **j)** $\text{HT29.DFMO} - \text{NCM.DFMO} = 0$.

Tabla C.43: Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto *Bootstrap*.

| | Estimación | SEM | Z | p-valor |
|-------------------------------|-------------|-----------|------------|--------------|
| β_0 | 3406.333431 | 445.66663 | 7.6432319 | 4.218847e-14 |
| $\beta_{\text{HT29.Control}}$ | 785.636472 | 39.78149 | 19.7487957 | 0 |
| $\beta_{\text{NCM.DFMO}}$ | 1.531929 | 37.06611 | 0.0413297 | 1 |
| $\beta_{\text{HT29.DFMO}}$ | 664.819581 | 41.15790 | 16.1529030 | 0 |

Tabla C.44: Covarianzas entre los coeficientes del modelo mixto *Bootstrap*.

| | β_0 | β_{HT29} | β_{DFMO} | $\beta_{\text{HT29:DFMO}}$ |
|----------------------------|-------------|-----------------------|-----------------------|----------------------------|
| β_0 | 198618.7486 | -944.5101 | -462.1353 | -234.1019 |
| β_{HT29} | -944.5101 | 1582.5668 | 696.6708 | 695.7959 |
| β_{DFMO} | -462.1353 | 696.6708 | 1373.8962 | 706.4702 |
| $\beta_{\text{HT29:DFMO}}$ | -234.1019 | 695.7959 | 706.4702 | 1693.9728 |

Tabla C.45: Diferencias dos a dos entre las condiciones experimentales del modelo lineal mixto *Bootstrap*.

| | Diferencias | SEM | pvalor | |
|-----------------------------------|-------------|----------|-----------|---|
| HT29.Control - NCM.Control | 785.636472 | 39.78149 | 0 | * |
| NCM.DFMO - NCM.Control | 1.531929 | 37.06611 | 0.4835165 | |
| HT29.DFMO - NCM.Control | 664.819581 | 41.15790 | 0 | * |
| NCM.DFMO - HT29.Control | -784.104543 | 39.53633 | 0 | * |
| HT29.DFMO - HT29.Control | -120.816891 | 43.41598 | 0.0026948 | * |
| HT29.DFMO - NCM.DFMO | 663.287651 | 40.68081 | 0 | * |

C.4. Modelado de la variable respuesta *Incremento Máximo* de la señal correspondiente a SOCE

Con el mismo procedimiento llevado a cabo en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos”, se han empleado modelos lineales mixtos estimados tanto mediante métodos clásicos como *bootstrapping* paramétrico para analizar el efecto del DFMO sobre SOCE y cómo de diferente es este efecto entre las líneas celulares de interés. En concreto, se ha modelizado el incremento máximo de la señal de SOCE de cada una de las células registradas (MaxSOCE). Cabe destacar que los posibles factores aleatorios también son los mismos que en el caso de *AUCDepósitos* (apéndice “C.1”). En la tabla C.46 se muestra el valor de cAIC para cada una de las estructuras aleatorias evaluadas, donde la estructura aleatoria se expresa según el convenio usado en el paquete lme4 de R:

Tabla C.46: Selección de modelos para la variable respuesta $\log(\text{MaxSOCE})$ con cAIC.

| modelo | Parte Aleatoria (Anotación según lme4) | cAIC |
|--------|---|-------------|
| C_DTiL | (1 Dia:Cristal) + (1 + Tratamiento * Linea Dia) | error |
| C_DTL | (1 Dia:Cristal) + (1 + Tratamiento + Linea Dia) | -390.40399 |
| C_DT | ((1 Dia:Cristal) + (1 + Tratamiento Dia) | -228.67447 |
| C_DL | (1 Dia:Cristal) + (1 + Linea Dia) | -355.74870 |
| C_D | (1 Dia:Cristal) + (1 Dia) | -237.78571 |
| C | (1 Cristal) | 309.387963 |
| DTiL | (1 + Tratamiento * Linea Dia) | error |
| DTL | (1 + Tratamiento + Linea Dia) | -376.9554 |
| DT | (1 + Tratamiento Dia) | -219.34984 |
| DL | (1 + Linea Dia) | -337.64046 |
| D | (1 Dia) | -224.770023 |
| Fijo | | 305.882 |

Como puede comprobarse, el valor más pequeño de cAIC es para el modelo cuasi sobreóptimo modelo DTL. No obstante, algunos de los modelos con menor cAIC estiman unas matrices de varianzas-covarianzas para la parte aleatoria peligrosas en el sentido de que hay efectos aleatorios con varianzas igual a cero (o su límite inferior es cero) y correlaciones entre efectos ± 1 (o al menos uno de sus límites inferior y/o superior). En consecuencia, estos modelos se descartan (Barr et al. (2013), Matuschek et al. (2017)). Así, de los modelos evaluados (en orden creciente de cAIC), en las tablas C.47-C.48 se muestran las estimaciones de la varianza de los diferentes efectos aleatorios y sus correlaciones (y sus intervalos de confianza) para aquellos modelos con un cAIC inferior al modelo seleccionado en los que se han identificado algunos de estos problemas.

Tabla C.47: Parte aleatoria de los modelos problemáticos para log(MaxSOCE). Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza. Con una x se indican las estimaciones problemáticas.

```
#Modelo C_DTL
Linear mixed model fit by REML ['lmerMod']
Formula: logMaxSOCE ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 + Tratamiento + Linea | Dia)
Data: datosr
REML criterion at convergence: -349.0612
Random effects:
Groups      Name                Std.Dev. Corr
Dia:Cristal (Intercept)  0.04364
Dia         (Intercept)  0.30107
           TratamientoDFMO 0.09601 -0.07
           LineaHT29       0.16956 -0.96  0.24
Residual                    0.19631

           2.5 %      97.5 %  Problemática
.sig01      0.01755046  0.11471485
.sig02      0.14393704  0.64420932
.sig03     -0.79760479  0.70789534
.sig04     -1.00000000 -0.60374310      x
.sig05      0.04158784  0.21035167
.sig06     -0.56057854  0.97768817      x
.sig07      0.08101954  0.36836822
.sigma      0.18777314  0.20527143

#Modelo DTL
Formula: logMaxSOCE ~ Linea * Tratamiento + (1 + Tratamiento + Linea | Dia)
Data: datosr
REML criterion at convergence: -339.0693
Random effects:
Groups      Name                Std.Dev. Corr
Dia         (Intercept)  0.2966
           TratamientoDFMO 0.1018  0.01
           LineaHT29       0.1677  -0.95  0.15
Residual                    0.1981

           2.5 %      97.5 %  Problemática
.sig01      0.14464931  0.6320788
.sig02     -0.73098082  0.7437408
.sig03     -1.00000000 -0.6141523      x
.sig04      0.04463729  0.2222875
.sig05     -1.13797332  0.9016600      x
.sig06      0.07939119  0.3602390
.sigma      0.18949079  0.2071081
```

Tabla C.48: Parte aleatoria de los modelos problemáticos para log(MaxSOCE). Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza. Con una x se indican las estimaciones problemáticas. (continuación)

```
#Modelo C_DL
Formula: logMaxSOCE ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 + Linea | Dia)
Data: datosr
REML criterion at convergence: -314.7501
Random effects:
Groups      Name          Std.Dev. Corr
Dia:Cristal (Intercept) 0.04894
Dia         (Intercept) 0.30631
           LineaHT29  0.17559 -0.91
Residual                                0.20028

           2.5 %      97.5 %  Problemática
.sig01      0.02148436  0.12897665
.sig02      0.14678897  0.65482056
.sig03     -1.00000000 -0.37811030      x
.sig04      0.08334000  0.37666491
.sigma      0.19149715  0.20931069

#Modelo DL
Formula: logMaxSOCE ~ Linea * Tratamiento + (1 + Linea | Dia)
Data: datosr
REML criterion at convergence: -301.2957
Random effects:
Groups      Name          Std.Dev. Corr
Dia         (Intercept) 0.3073
           LineaHT29  0.1739 -0.91
Residual                                0.2026

           2.5 %      97.5 %  Problemática
.sig01      0.15048030  0.65375995
.sig02     -1.00000000 -0.39906524      x
.sig03      0.08243320  0.37332378
.sigma      0.19369257  0.21166721
```

En consecuencia, tras descartar los modelos problemáticos, el modelo que presenta un menor valor de cAIC y no presenta ninguno de los problemas considerados es el *modelo C_D* (tabla C.49).

Tabla C.49: Parte aleatoria del modelo seleccionado para MaxSOCE. Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza.

```
#Modelo C_D
Linear mixed model fit by REML ['lmerMod']
Formula: logMaxSOCE ~ Linea * Tratamiento + (1 | Dia:Cristal) + (1 | Dia)
Data: datosr
REML criterion at convergence: -203.4229
Random effects:
Groups      Name          Std.Dev.
Dia:Cristal (Intercept) 0.04608
Dia         (Intercept) 0.23411
Residual                    0.21300
Number of obs: 984, groups: Dia:Cristal, 8; Dia, 4
```

Posteriormente, mediante la AIC, se evalúa la parte fija que, al tratarse de un modelo de efectos fijos, es la única parte que contiene. Como resultado, en la tabla C.50 se muestra la selección de la parte fija que, como puede verse, resultan ser tanto los dos efectos principales y su interacción.

Tabla C.50: Selección de modelos para la variable respuesta log(MaxSOCE) con AIC.

| Parte Fija | AIC |
|---|---------|
| Linea + Tratamiento + Linea:Tratamiento | -210.81 |
| Linea + Tratamiento | -212.79 |
| Linea | -212.80 |
| Tratamiento | 421.35 |
| nulo | 422.58 |

Como puede observarse, según el valor de AIC el modelo no debería tener ni el término de interacción ni el del factor *Tratamiento*. No obstante, debido al elevado número de observaciones, se ha creído conveniente dejarlo puesto que la pérdida de dos grado de libertad, en este caso, no se considera problemático. Tras evaluar el modelo sobreóptimo con el fin de encontrar tanto la parte fija como la parte aleatoria más adecuada, se obtuvo que el modelo más adecuado para modelizar MaxSOCE es (ecuación C.9):

$$\log(\text{MaxSOCE})_{ijdr} = \beta_0 + \beta_1 \text{Linea}_i + \beta_2 \text{Tratamiento}_j + \beta_3 \text{Tejido}_i \text{Tratamiento}_j + b_{0,d} + b_{0,c} + \epsilon_{ijdr} \quad (\text{C.9})$$

donde $\log(\text{MaxSOCE})_{ijdr}$ es el valor de la variable respuesta medido en la réplica r de la muestra perteneciente a la línea celular i (NCM460 o HT29), tratamiento j (Control o DFMO) medidas en el día d y en el cristal c . En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel NCM460.Control); β_1 es el

cambio de expresión en la línea HT29 respecto del de referencia, β_2 el cambio en $\log(\text{MaxSOCE})$ en las células tratadas con DFMO respecto de las de referencia (NCM460.Control); β_3 el efecto de la interacción entre los factores *Línea* y *Tratamiento* que indica si el cambio entre células tratadas con DFMO y sin tratar es diferente en NCM460 que en HT29 (en caso de ser significativo el término de interacción); $b_{0,d}$ es el intercepto aleatorio debido al día que indica la desviación respecto de β_0 que hay en el día d , y $b_{0,c}$ es el intercepto aleatorio debido a cristal o cubreobjetos c que indica la desviación respecto de β_0 que se observa en el cristal c .

En cuanto al diagnóstico del modelo o la evaluación del cumplimiento de los supuestos, en la tabla C.51 se muestran los métodos basados en contrastes de hipótesis y en la figura C.10 los métodos gráficos. Como puede observarse, aunque en los métodos gráficos los supuestos de normalidad y homocedasticidad no parecen violarse de forma preocupante, los tests sugieren que la violación es considerable. Por ello, tras realizar las estimaciones y contrastes de hipótesis pertinentes con los modelos lineales mixtos, se emplearán las técnicas de estimación *bootstrapping* para evaluar si el modelo es o no adecuado en función de la semejanza entre ambos y, en caso de que no lo sean, se emplearán las estimaciones *bootstrapping*.

Tabla C.51: Supuestos del modelo seleccionado (p valores).

| | Normalidad | | | Homocedasticidad | |
|-------------------------------|------------|--------|--------|------------------|----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 0.00727 | 0.0171 | 0.5070 | 0.0338 | 1e-07 |
| Día:Cristal intercepto | 0.18700 | 0.7110 | 0.0952 | | |
| Día intercepto | 0.18700 | 0.7070 | | | |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test. de Breush-Pagan.

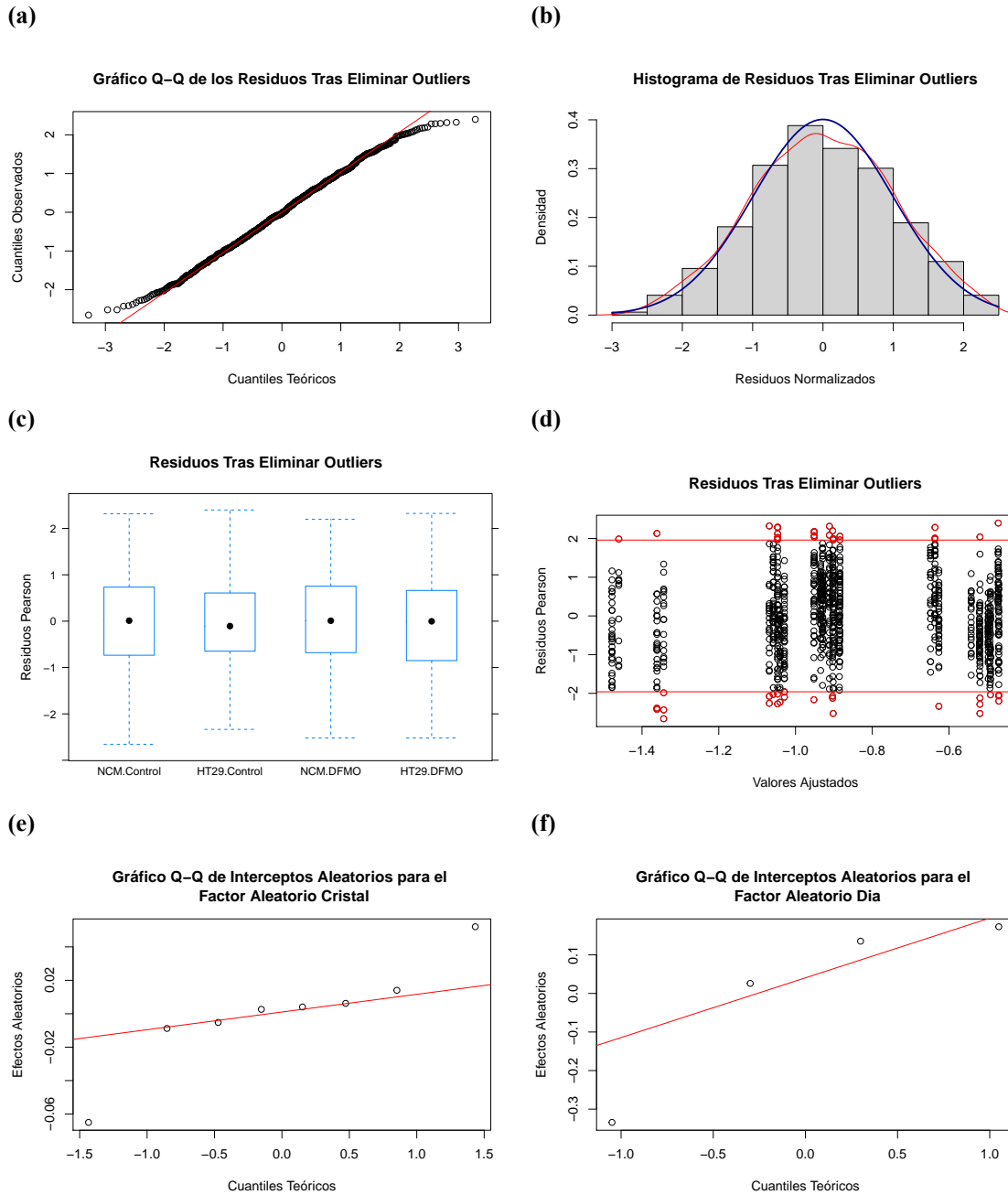


Figura C.10: Diagnóstico del modelo lineal de efectos fijos para la variable respuesta $\log(\text{MaxSOCE})$ tras eliminar los *outliers*. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como *outliers*; e) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Día, y f) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Cristal.

En la tabla C.52 se muestran las estimaciones de los coeficientes del modelo y en la tabla C.53 su matriz de varianzas y covarianzas.

Tabla C.52: Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para $\log(\text{MaxSOCE})$.

| | Estimación | SEM | t | gdl | p-valor |
|----------------------------|------------|-----------|------------|----------|-----------|
| β_0 | -1.0611736 | 0.1189419 | -8.9217789 | 3.139897 | 0.0024773 |
| β_{HT29} | 0.4139178 | 0.0197426 | 20.9657645 | 3.139897 | 0.0001763 |
| β_{DFMO} | -0.0175972 | 0.0182907 | -0.9620884 | 3.139897 | 0.4040995 |
| $\beta_{\text{HT29:DFMO}}$ | -0.0040243 | 0.0277198 | -0.1451785 | 3.139897 | 0.8933924 |

Abreviaturas:

gdl: grados de libertad

Tabla C.53: Covarianzas entre los coeficientes del modelo final para $\log(\text{MaxSOCE})$.

| | β_0 | β_{HT29} | β_{DFMO} | $\beta_{\text{HT29:DFMO}}$ |
|----------------------------|------------|-----------------------|-----------------------|----------------------------|
| β_0 | 0.0141485 | -0.0001776 | -0.0001783 | 0.0001789 |
| β_{HT29} | -0.0001776 | 0.0003898 | 0.0001771 | -0.0003921 |
| β_{DFMO} | -0.0001783 | 0.0001771 | 0.0003345 | -0.0003346 |
| $\beta_{\text{HT29:DFMO}}$ | 0.0001789 | -0.0003921 | -0.0003346 | 0.0007684 |

En cuanto a la estructura de la parte aleatoria, la varianza de los efectos aleatorios se muestra en la tabla en la tabla C.54.

Tabla C.54: Estimaciones de la varianza de la parte aleatoria del modelo mixto final para $\log(\text{MaxSOCE})$.

| Componente Aleatorio | LI | Estimación SEM | LS |
|---------------------------|-----------|----------------|-----------|
| Cristal Intercepto | 0.0181572 | 0.0460813 | 0.1240629 |
| Día Intercepto | 0.1109537 | 0.2341123 | 0.5017179 |
| Sigma | 0.2035701 | 0.2129954 | 0.2224663 |

La tabla ANOVA del modelo seleccionado con interacción se muestra en la tabla C.55, y sin interacción en la tabla C.56. Evidentemente, el factor *Tratamiento* no es significativo y, además, según AIC tampoco ha de estar en el modelo. En cuanto a la interacción, en ningún caso resulta ser necesaria en el modelo; es decir, no es significativa.

Tabla C.55: Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para log (MaxSOCE).

| | Sum Sq | gdl | p-valor |
|--------------------------|-------------|-----|-----------|
| Linea | 894.1829478 | 1 | 0.0000000 |
| Tratamiento | 1.9831242 | 1 | 0.1590617 |
| Linea:Tratamiento | 0.0210768 | 1 | 0.8845700 |

Abreviaturas:

gdl: grados de libertad

Tabla C.56: Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado sin interacción para log (MaxSOCE).

| | Sum Sq | gdl | p-valor |
|--------------------|------------|-----|-----------|
| Linea | 895.078992 | 1 | 0.0000000 |
| Tratamiento | 1.985057 | 1 | 0.1588586 |

Abreviaturas:

gdl: grados de libertad

Las estimaciones de MaxSOCE se muestran en la tabla C.57.

Tabla C.57: Estimaciones de MaxSOCE con el modelo seleccionado.

| Condiciones | log(Señal) | Límite Superior | Límite Inferior | log(Señal-SEM) | log(Señal+SEM) |
|---------------------|------------|-----------------|-----------------|----------------|----------------|
| NCM.Control | 0.3460494 | 0.4369057 | 0.2740870 | 0.3072417 | 0.3897589 |
| HT29.Control | 0.5234803 | 0.6611099 | 0.4145024 | 0.4647070 | 0.5896868 |
| NCM.DFMO | 0.3400132 | 0.4292070 | 0.2693548 | 0.3019103 | 0.3829249 |
| HT29.DFMO | 0.5122834 | 0.6469954 | 0.4056200 | 0.4547579 | 0.5770857 |

Las comparaciones múltiples se muestran en la tabla C.58.

Tabla C.58: Diferencias dos a dos entre las condiciones experimentales de sus valores log (MaxSOCE) estimados con el modelo seleccionado.

| | Diferencias | SEM | t | pvalor | |
|-----------------------------------|-------------|-----------|-------------|-----------|---|
| HT29.Control - NCM.Control | 0.4139179 | 0.0197426 | 20.9657691 | 0.0000000 | * |
| NCM.DFMO - NCM.Control | -0.0175971 | 0.0182907 | -0.9620797 | 0.7706236 | |
| HT29.DFMO - NCM.Control | 0.3922965 | 0.0198361 | 19.7768854 | 0.0000000 | * |
| NCM.DFMO - HT29.Control | -0.4315149 | 0.0192379 | -22.4304365 | 0.0000000 | * |
| HT29.DFMO - HT29.Control | -0.0216214 | 0.0208240 | -1.0382922 | 0.7264696 | |
| HT29.DFMO - NCM.DFMO | 0.4098935 | 0.0193390 | 21.1951338 | 0.0000000 | * |

De igual modo que considerando el modelo sin interacción, en el que al no existir dicho término y ser el factor *Línea* significativo, las comparaciones dos a dos con el modelo que sí tiene interacción también revela que las únicas diferencias existen entre las líneas celulares.

No obstante, puesto que el supuesto de homocedasticidad se viola en los modelos empleados, también se recurre a la estimación del modelo lineal mixto seleccionado mediante *bootstrapping* paramétrico. En la figura C.11 se muestran los gráficos de los contrastes de hipótesis, en la tabla C.59 las estimaciones de los coeficientes y los resultados del contraste de hipótesis, en la tabla C.29 la matriz de varianzas y covarianzas de los coeficientes estimados y en la tabla C.61 las diferencias dos a dos que, como puede comprobarse, son similares a las obtenidas mediante el método clásico de estimación de los modelos lineales mixtos. Seguramente, se debe a la robustez del modelo a pequeñas desviaciones de los supuestos.

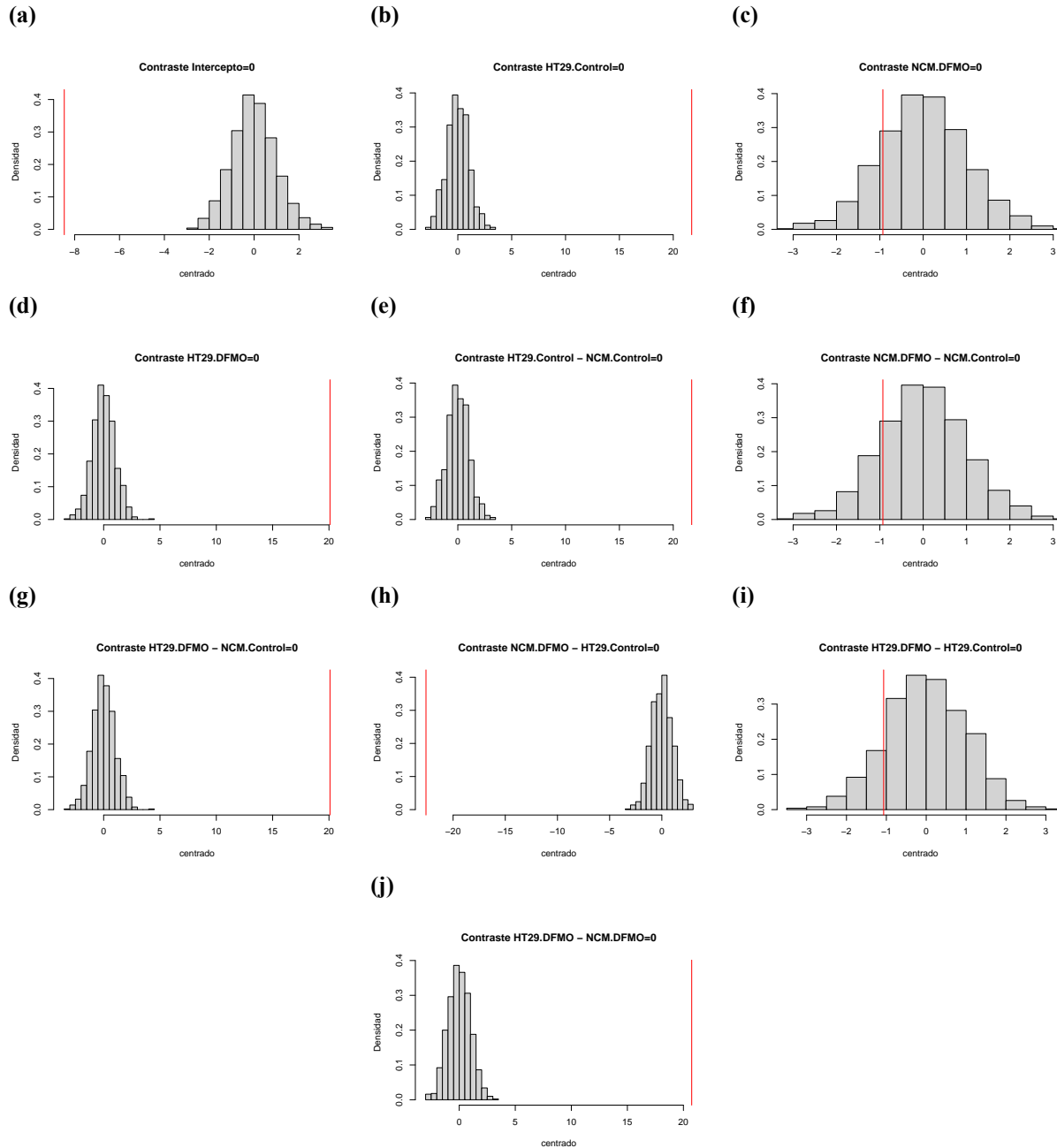


Figura C.11: Contrastes de hipótesis mediante *Bootstrapping*. Se muestran las distribuciones nulas, con ayuda de un histograma, de cada uno de los estadísticos correspondientes a los contrastes de hipótesis de interés, y el estadístico original se representa en cada uno de ellos mediante una línea vertical roja para evidenciar cómo de cerca o lejos está de dicha distribución. Los contrastes de hipótesis evaluados son: **a)** $B_0 = 0$; **b)** $B_{\text{HT29.Control}} = 0$; **c)** $B_{\text{NCM.DFMO}} = 0$; **d)** $B_{\text{HT29.DFMO}} = 0$; **e)** $\text{HT29.Control} - \text{NCM.Control} = 0$; **f)** $\text{NCM.DFMO} - \text{NCM.Control} = 0$; **g)** $\text{HT29.DFMO} - \text{NCM.Control} = 0$; **h)** $\text{NCM.DFMO} - \text{HT29.Control} = 0$; **i)** $\text{HT29.DFMO} - \text{HT29.Control} = 0$; **j)** $\text{HT29.DFMO} - \text{NCM.DFMO} = 0$.

Tabla C.59: Estimaciones de la parte fija y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo mixto *Bootstrap*.

| | Estimación | SEM | t | p-valor |
|-------------------------------|------------|-----------|------------|---------|
| β_0 | -1.0583105 | 0.1254004 | -8.4394531 | 0 |
| $\beta_{\text{HT29.Control}}$ | 0.4136723 | 0.0190618 | 21.7016333 | 0 |
| $\beta_{\text{NCM.DFMO}}$ | -0.0176163 | 0.0189268 | -0.9307596 | 0.366 |
| $\beta_{\text{HT29.DFMO}}$ | 0.3923188 | 0.0195243 | 20.0938560 | 0 |

Tabla C.60: Covarianzas entre los coeficientes del modelo mixto *Bootstrap*.

| | β_0 | β_{HT29} | β_{DFMO} | $\beta_{\text{HT29:DFMO}}$ |
|----------------------------|------------|-----------------------|-----------------------|----------------------------|
| β_0 | 0.0157253 | -0.0003016 | -0.0002686 | -0.0002859 |
| β_{HT29} | -0.0003016 | 0.0003634 | 0.0001785 | 0.0001671 |
| β_{DFMO} | -0.0002686 | 0.0001785 | 0.0003582 | 0.0001744 |
| $\beta_{\text{HT29:DFMO}}$ | -0.0002859 | 0.0001671 | 0.0001744 | 0.0003812 |

Tabla C.61: Diferencias dos a dos entre las condiciones experimentales del modelo lineal mixto *Bootstrap*.

| | Diferencias | SEM | pvalor |
|-----------------------------------|-------------|-----------|--------|
| HT29.Control - NCM.Control | 0.4136723 | 0.0190618 | 0 * |
| NCM.DFMO - NCM.Control | -0.0176163 | 0.0189268 | 1 |
| HT29.DFMO - NCM.Control | 0.3923188 | 0.0195243 | 0 * |
| NCM.DFMO - HT29.Control | -0.4312886 | 0.0190948 | 0 * |
| HT29.DFMO - HT29.Control | -0.0213535 | 0.0202555 | 1 |
| HT29.DFMO - NCM.DFMO | 0.4099351 | 0.0197627 | 0 * |

C.5. Modelado de la variable respuesta $[\text{Ca}^{2+}]_{\text{cit}}$ basal

Con el mismo procedimiento llevado a cabo en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos”, se han empleado modelos lineales mixtos estimados tanto mediante métodos clásicos como *bootstrapping* paramétrico para analizar el efecto del DFMO la $[\text{Ca}^{2+}]_{\text{cit}}$ basal y cómo de diferente es este efecto entre las líneas celulares de interés. En concreto, se ha modelizado la mediana de la señal durante el primer minuto del experimento, de cada célula, en medio Ca^{2+} 1 mM. Cabe destacar que los posibles factores aleatorios también son los mismos que en el caso de *AUCDepósitos* (apéndice “C.1”). En la tabla C.62 se muestra el valor de cAIC para cada una de las estructuras aleatorias evaluadas, donde la estructura aleatoria se expresa según el convenio usado en el paquete lme4 de R:

Tabla C.62: Selección de modelos para la variable respuesta $[\text{Ca}^{2+}]_{\text{cit}}$ basal con cAIC.

| modelo | Parte Aleatoria (Anotación según lme4) | cAIC |
|--------|---|-----------|
| C_DTiL | (1 Dia:Cristal) + (1 + Tratamiento * Linea Dia) | error |
| C_DTL | (1 Dia:Cristal) + (1 + Tratamiento + Linea Dia) | error |
| C_DT | ((1 Dia:Cristal) + (1 + Tratamiento Dia) | error |
| C_DL | (1 Dia:Cristal) + (1 + Linea Dia) | error |
| C_D | (1 Dia:Cristal) + (1 Dia) | error |
| C | (1 Cristal) | -1940.043 |
| DTiL | (1 + Tratamiento * Linea Dia) | error |
| DTL | (1 + Tratamiento + Linea Dia) | error |
| DT | (1 + Tratamiento Dia) | -1785.378 |
| DL | (1 + Linea Dia) | error |
| D | (1 Dia) | -1785.378 |
| Fijo | | -1785.378 |

Como puede comprobarse, el valor más pequeño de cAIC es para el modelo C, es decir, aquel cuya parte aleatoria tiene únicamente como efecto aleatorio al intercepto aleatorio debido al *Cristal* (tabla C.63).

Tabla C.63: Parte aleatoria del modelo seleccionado para $[\text{Ca}^{2+}]_{\text{cit}}$ basal. Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza.

```
#Modelo C_D
Linear mixed model fit by REML ['lmerMod']
Formula: CaBasal ~ Linea * Tratamiento + (1 | Cristal)
Data: datosr
REML criterion at convergence: -1882.803
Random effects:
Groups          Name          Std.Dev.
Dia:IDexperimento (Intercept) 0.04390
Residual                0.09164
Number of obs: 1015, groups: Dia:IDexperimento, 32
```

Posteriormente, mediante la AIC, se evalúa la parte fija que, al tratarse de un modelo de efectos fijos, es la única parte que contiene. Como resultado, en la tabla C.64 se muestra la selección de la parte fija que, como puede verse, resultan ser tanto los dos efectos principales y su interacción.

Tabla C.64: Selección de modelos para la variable respuesta $[\text{Ca}^{2+}]_{\text{cit}}$ basal con AIC.

| Parte Fija | AIC |
|--|---------|
| Linea + Tratamiento + Linea:Tratamiento | -1896.5 |
| Linea + Tratamiento | -1898.1 |
| Linea | -1895.0 |
| Tratamiento | -1894.2 |
| nulo | -1891.8 |

Como puede observarse, según el valor de AIC el modelo no debería tener ni el término de interacción ni el del factor *Tratamiento*. No obstante, debido al elevado número de observaciones, se ha creído conveniente dejarlo puesto que la pérdida de dos grado de libertad, en este caso, no se considera problemático. Tras evaluar el modelo sobreóptimo con el fin de encontrar tanto la parte fija como la parte aleatoria más adecuada, se obtuvo que el modelo más adecuado para modelizar MaxSOCE es (ecuación C.10). Cabe destacar que la variable respuesta ha sido transformada según la familia de Box-Cox (Hawkins and Weisberg (2017)), con un gamma de 0.48 y un lambda de -0.79.

$$T(\text{CaBasal})_{ijcr} = \beta_0 + \beta_1 \text{Linea}_i + \beta_2 \text{Tratamiento}_j + \beta_3 \text{Tejido}_i \text{Tratamiento}_j + b_{0,c} + \epsilon_{ijcr} \quad (\text{C.10})$$

donde $T(\text{CaBasal})_{ijcr}$ es el valor de la variable respuesta transformada medida en la réplica r de la muestra perteneciente a la línea celular i (NCM460 o HT29), tratamiento j (Control o DFMO)

medidas en el cristal c . En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel NCM460.Control); β_1 es el cambio de expresión en la línea HT29 respecto del de referencia, β_2 el cambio en $\log(\text{MaxSOCE})$ en las células tratadas con DFMO respecto de las de referencia (NCM460.Control); β_3 el efecto de la interacción entre los factores *Línea* y *Tratamiento* que indica si el cambio entre células tratadas con DFMO y sin tratar es diferente en NCM460 que en HT29 (en caso de ser significativo el término de interacción), y $b_{0,c}$ es el intercepto aleatorio debido a cristal o cubreobjetos c que indica la desviación respecto de β_0 que se observa en el cristal c .

En cuanto al diagnóstico del modelo o la evaluación del cumplimiento de los supuestos, en la tabla C.65 los métodos basados en contrastes de hipótesis y en la figura C.10 se muestran los métodos gráficos. Como puede observarse, aunque en los métodos gráficos los supuestos de normalidad y homocedasticidad no parecen violarse de forma preocupante, los tests sugieren que la violación es considerable. Por ello, tras realizar las estimaciones y contrastes de hipótesis pertinentes con los modelos lineales mixtos, se emplearán las técnicas de estimación *bootstrapping* para evaluar si el modelo es o no adecuado en función de la semejanza entre ambos y, en caso de que no lo sean, se emplearán las estimaciones *bootstrapping*.

Tabla C.65: Supuestos del modelo seleccionado (p valores).

| | Normalidad | | | Homocedasticidad | |
|----------------------------|------------|---------|----------|------------------|-----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 4.00e-07 | 0.00038 | 2.00e-07 | 0.5090 | 0.3743226 |
| Cristal intercepto | 8.99e-01 | 0.81600 | 4.67e-01 | | |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test. de Breush-Pagan.

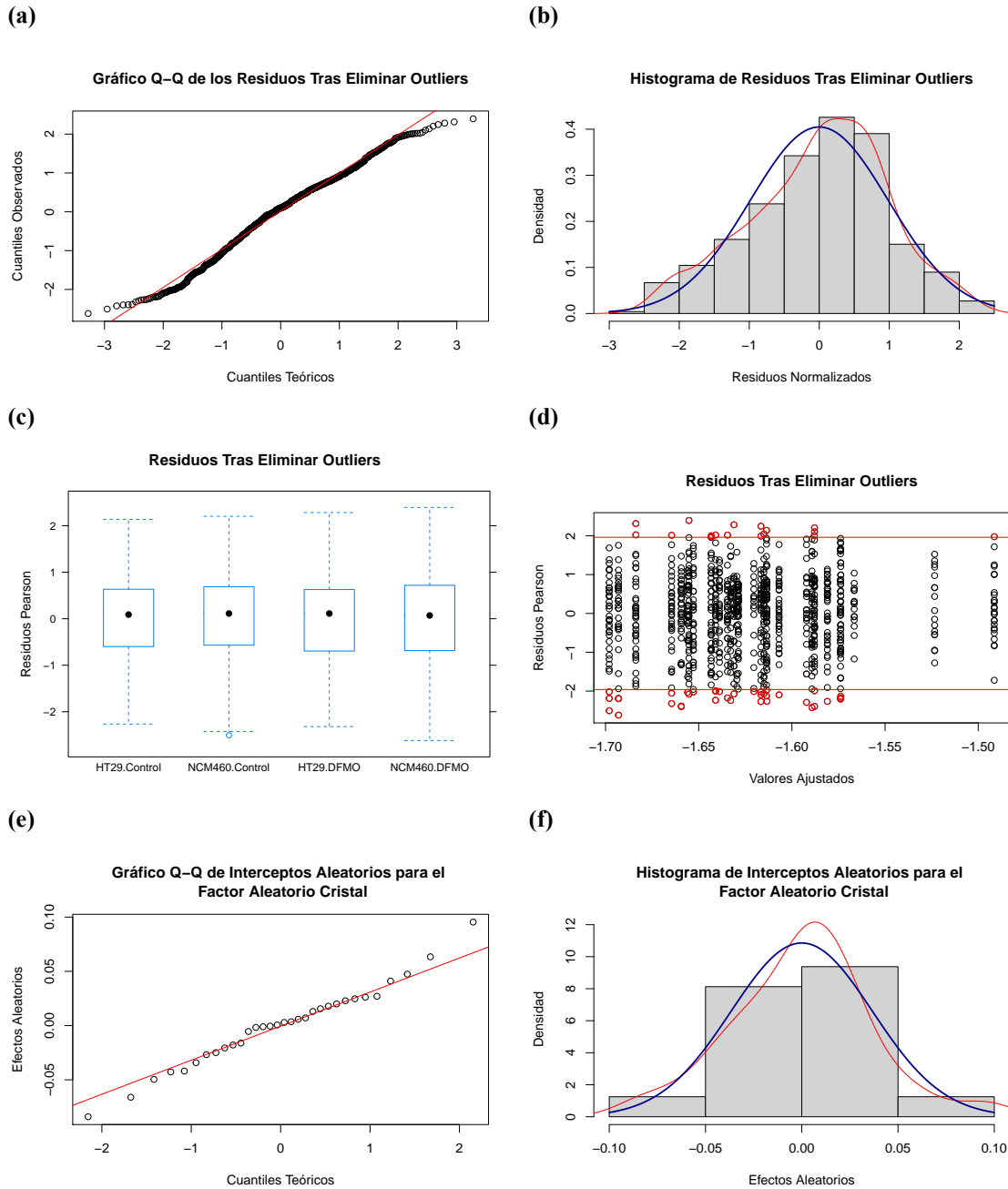


Figura C.12: Diagnóstico del modelo lineal de efectos fijos para la variable respuesta $T(\text{CaBasal})_{ijcr}$ tras eliminar los *outliers*. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como *outliers*, y f) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Cristal.

En la tabla C.66 se muestran las estimaciones de los coeficientes del modelo y en la tabla C.67 su matriz de varianzas y covarianzas.

Tabla C.66: Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para T (CaBasal).

| | Estimación | SEM | t | gdl | p-valor |
|----------------------------|------------|-----------|--------------|----------|-----------|
| β_0 | -1.5867731 | 0.0154537 | -102.6789992 | 27.93819 | 0 |
| β_{HT29} | -0.0271539 | 0.0218514 | -1.2426619 | 27.93819 | 0.2243216 |
| β_{DFMO} | -0.0281599 | 0.0219415 | -1.2834093 | 27.93819 | 0.2098925 |
| $\beta_{\text{HT29:DFMO}}$ | -0.0167015 | 0.0309130 | -0.5402738 | 27.93819 | 0.5932895 |

Abreviaturas:

gdl: grados de libertad

Tabla C.67: Covarianzas entre los coeficientes del modelo final para T (CaBasal).

| | β_0 | β_{HT29} | β_{DFMO} | $\beta_{\text{HT29:DFMO}}$ |
|----------------------------|------------|-----------------------|-----------------------|----------------------------|
| β_0 | 0.0002388 | -0.0002388 | -0.0002388 | 0.0002388 |
| β_{HT29} | -0.0002388 | 0.0004775 | 0.0002388 | -0.0004775 |
| β_{DFMO} | -0.0002388 | 0.0002388 | 0.0004814 | -0.0004814 |
| $\beta_{\text{HT29:DFMO}}$ | 0.0002388 | -0.0004775 | -0.0004814 | 0.0009556 |

La varianza de los efectos aleatorios se muestra en la tabla en la tabla C.68.

Tabla C.68: Estimaciones de la varianza de la parte aleatoria del modelo mixto final para T (CaBasal).

| Componente Aleatorio | LI | Estimación SEM | LS |
|---------------------------|-----------|----------------|-----------|
| Cristal Intercepto | 0.0289463 | 0.041113 | 0.0513076 |
| Sigma | 0.0752074 | 0.078657 | 0.0823825 |

La tabla ANOVA del modelo seleccionado con interacción se muestra en la tabla C.69, y sin interacción en la tabla C.56. Tanto el factor *Tratamiento* como el factor *Linea* son significativos, pero no el término de interacción, por lo que el DFMO afecta por igual a ambas líneas celulares, si bien es cierto que entre ellas son diferentes también.

Tabla C.69: Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para T (CaBasal).

| | Sum Sq | gdl | p-valor |
|--------------------------|-----------|-----|-----------|
| Linea | 5.2748341 | 1 | 0.0216358 |
| Tratamiento | 5.5994718 | 1 | 0.0179659 |
| Linea:Tratamiento | 0.2918958 | 1 | 0.5890082 |

Abreviaturas:

gdl: grados de libertad

Tabla C.70: Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado sin interacción para $\log(\text{CaBasal})$.

| | Sum Sq | gdl | p-valor |
|--------------------|----------|-----|-----------|
| Linea | 5.407616 | 1 | 0.0200491 |
| Tratamiento | 5.735271 | 1 | 0.0166275 |

Abreviaturas:

gdl: grados de libertad

Las estimaciones de CaBasal se muestran en la tabla C.71.

Tabla C.71: Estimaciones de $[\text{Ca}^{2+}]_{\text{cit}}$ basal con el modelo seleccionado.

| Condiciones | $\log(\text{Señal})$ | Límite Superior | Límite Inferior | $\log(\text{Señal}-\text{SEM})$ | $\log(\text{Señal}+\text{SEM})$ |
|---------------------|----------------------|-----------------|-----------------|---------------------------------|---------------------------------|
| NCM.Control | 0.1886032 | 0.1955506 | 0.1817616 | 0.1850996 | 0.1921343 |
| HT29.Control | 0.1948284 | 0.2018765 | 0.1878899 | 0.1912749 | 0.1984104 |
| NCM.DFMO | 0.1785036 | 0.1852502 | 0.1718561 | 0.1750998 | 0.1819331 |
| HT29.DFMO | 0.1883742 | 0.1953757 | 0.1814802 | 0.1848436 | 0.1919327 |

En este caso, puesto que la interacción no es significativa y dado que cada uno de los factores únicamente tiene dos niveles, es evidente que ambos factores son significativos; es decir, que el tratamiento afecta a ambas líneas celulares del mismo modo y, además, que ambas líneas celulares tienen una $[\text{Ca}^{2+}]_{\text{cit}}$ diferente. Así, el contraste de interés, que es si las células HT29 tratadas con DFMO son o no iguales de las NCM460 control, se muestra en la tabla C.72.

Tabla C.72: Contrastes de interés ($T(\text{CaBasal})$) estimados con el modelo seleccionado.

| | Diferencias | SEM | t | pvalor |
|--------------------------------|-------------|----------|-------|--------|
| HT29.DFMO - NCM.Control | 0.001006 | 0.021938 | 0.046 | 0.963 |

En este caso, no se han realizado estimaciones *bootstrapping* dado la suposición de homocedasticidad se cumple adecuadamente y que la desviación de la normalidad no es preocupante, y menos con un tamaño muestral superior a 1000 células. No obstante, puesto que el supuesto de homocedasticidad se viola en los modelos empleados, también se recurre a la estimación del modelo lineal mixto seleccionado mediante *bootstrapping* paramétrico. En la figura C.11 se muestran los gráficos de los contrastes de hipótesis, en la tabla C.59 las estimaciones de los coeficientes y los resultados del contraste de hipótesis, en la tabla C.29 la matriz de varianzas y covarianzas de los coeficientes estimados y en la tabla C.61 las diferencias dos a dos que, como puede comprobarse, son similares a las obtenidas mediante el método clásico de estimación de los modelos lineales mixtos. Seguramente, se debe a la robustez del modelo a pequeñas desviaciones de los supuestos.

Apéndice D

Análisis de datos de Western Blot

En este apéndice se muestran diferentes resultados de los modelos empleados para el análisis de expresión de proteínas mediante *Western Blot*. En particular, se muestra la selección de la parte aleatoria y fija de los modelos, la evaluación de los supuestos del modelo seleccionado, la tabla ANOVA de los factores fijos del modelo, la estimación de los coeficientes y su matriz de varianzas y covarianzas, las estimaciones de la parte aleatoria, las estimaciones de la expresión de cada una de las cuatro proteínas y las comparaciones dos a dos entre las diferentes condiciones experimentales.

D.1. Modelado de la expresión de la proteína ORAI1 en muestras de tejido humano de cáncer de colon y normal

Con el mismo procedimiento llevado a cabo en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos”, se ha modernizado la expresión de la proteína ORAI1. En primer lugar, se generó el modelo sobreóptimo con los dos factores fijos (*Tejido* y *Sexo*) y los dos factores aleatorios cruzados (*Paciente* y *Gel*).

Tras evaluar el modelo sobreóptimo con el fin de encontrar tanto la parte fija como la parte aleatoria más adecuada, se obtuvo que el modelo más adecuado para modelizar la expresión de ORAI1 es un modelo de efectos fijos (ecuación D.1):

$$\log \left(\frac{DO_{ORAI1}}{DO_{\beta\text{-actina}}}_{ijr} \right) = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{sexo}_j + \epsilon_{ijr} \quad (\text{D.1})$$

donde $\log \left(\frac{DO_{ORAI1}}{DO_{\beta\text{-actina}}}_{ijr} \right)$ es el valor de la expresión de ORAI1 medido en la réplica r de la muestra cuyo tejido es i (Tumoral o Normal) y sexo j (Hombre o Mujer). En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el

nivel Hombre.Normal); β_1 es el cambio de expresión en el tejido tumoral respecto del de referencia (Hombre.Normal), β_2 el cambio en la expresión de la proteína de interés en las muestras de mujer respecto de las de referencia (Hombre.Normal), y β_3 el efecto de la interacción entre los factores *Tejido* y *Sexo* que indica si el cambio entre el tejido tumoral y normal es diferente en hombres que en mujeres (en caso de ser significativo). En cuanto al diagnóstico del modelo o la evaluación del cumplimiento de los supuestos, en la figura D.1 se muestran los métodos gráficos y en la tabla D.1 los métodos basados en contrastes de hipótesis.

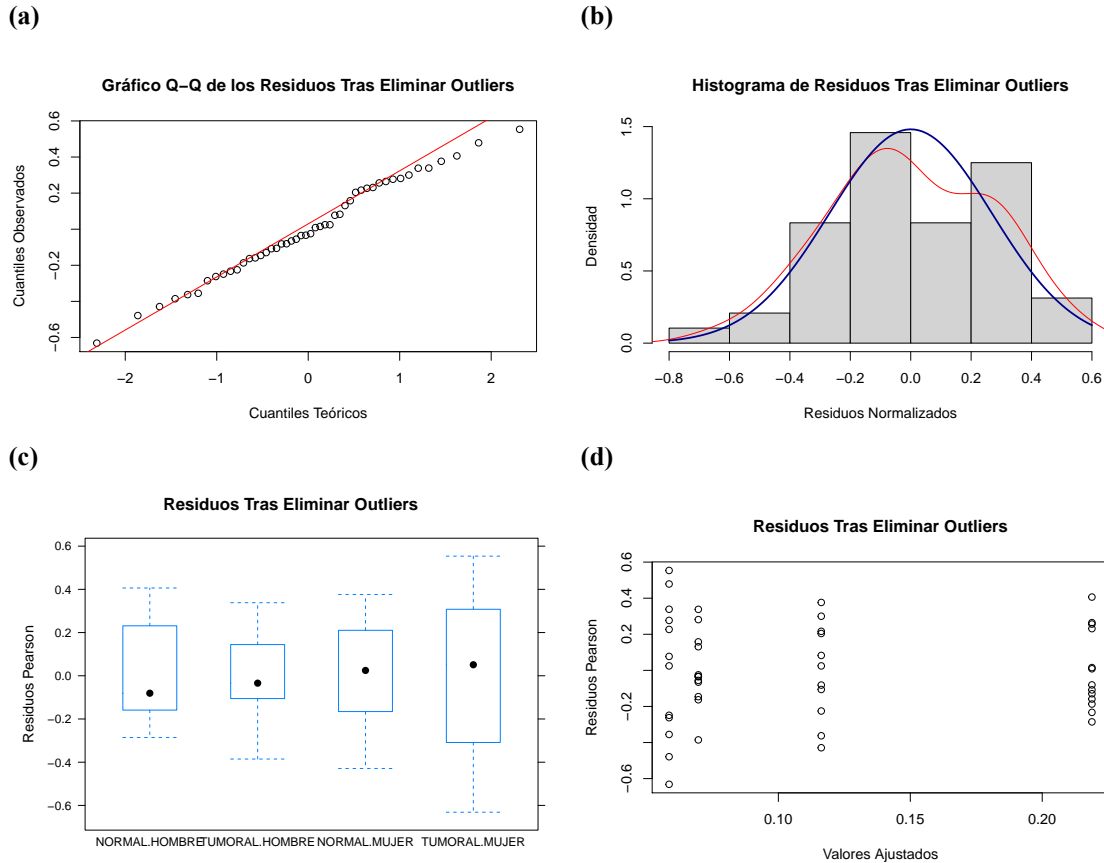


Figura D.1: Diagnóstico del modelo lineal de efectos fijos para la variable respuesta $\log(DO_{ORAI1}/DO_{\beta\text{-actina}})$ tras eliminar los outliers. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como outliers.

Tabla D.1: Supuestos del modelo seleccionado (p valores).

| | Normalidad | | | Homocedasticidad | |
|----------------------------|------------|---------|---------|------------------|----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 0.90000 | 0.69800 | 0.45600 | 0.33800 | 0.11190 |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test. de Breush-Pagan.

En la tabla D.2 se muestran las estimaciones de los coeficientes del modelo y en la tabla D.3 su matriz de varianzas y covarianzas.

Tabla D.2: Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo seleccionado para $\log(DO_{ORAI1}/DO_{\beta-actina})$.

| | Estimación | SEM | t | p-valor |
|-------------------------|------------|-----------|------------|-----------|
| β_0 | 0.2188366 | 0.0772070 | 2.8344153 | 0.0069075 |
| $\beta_{Tumoral}$ | -0.1491714 | 0.1114386 | -1.3385969 | 0.1875777 |
| β_{Mujer} | -0.1025796 | 0.1140422 | -0.8994881 | 0.3732893 |
| $\beta_{Tumoral:Mujer}$ | 0.0915008 | 0.1609998 | 0.5683287 | 0.5727010 |

Tabla D.3: Covarianzas entre los coeficientes del modelo seleccionado para $\log(DO_{ORAI1}/DO_{\beta-actina})$.

| | β_0 | $\beta_{Tumoral}$ | β_{Mujer} | $\beta_{Tumoral:Mujer}$ |
|-------------------------|-----------|-------------------|-----------------|-------------------------|
| β_0 | 0.00596 | -0.00596 | -0.00596 | 0.00596 |
| $\beta_{Tumoral}$ | -0.00596 | 0.01242 | 0.00596 | -0.01242 |
| β_{Mujer} | -0.00596 | 0.00596 | 0.01301 | -0.01301 |
| $\beta_{Tumoral:Mujer}$ | 0.00596 | -0.01242 | -0.01301 | 0.02592 |

La tabla ANOVA del modelo seleccionado se muestra en la tabla D.4. Evidentemente, ningún factor tiene efecto sobre la variable respuesta y, por ende, siendo estricto, habría que eliminar la interacción entre los factores y demás variables innecesarias. No obstante, puesto que se pretende estimar el valor de la variable respuesta para cada una de las condiciones experimentales y, además, que hay suficientes grados de libertad para estimar el error aleatorio, entonces no se han eliminado del modelo.

Tabla D.4: Tabla ANOVA (tipo II) del modelo seleccionado para $\log(DO_{ORAI1}/DO_{\beta-actina})$.

| | Sum Sq | gdl | F | p-valor |
|--------------------|-----------|-----|-----------|-----------|
| Tejido | 0.1329113 | 1 | 1.7151637 | 0.1971141 |
| Sexo | 0.0384038 | 1 | 0.4955848 | 0.4851558 |
| Tejido:Sexo | 0.0250297 | 1 | 0.3229976 | 0.5727010 |
| Residuals | 3.4096429 | 44 | | |

Abreviaturas:

gdl: grados de libertad

Las estimaciones de la expresión de ORAI1, expresado como $DO_{ORAI1}/DO_{\beta-actina}$, se muestran en la tabla D.5.

Tabla D.5: Estimaciones de la expresión de ORAI1 ($DO_{ORAI1}/DO_{\beta-actina}$) con el modelo seleccionado.

| Condiciones | ratio DO | Límite Superior | Límite Inferior | log(Señal-SEM) | log(Señal+SEM) | n |
|-----------------------|----------|-----------------|-----------------|----------------|----------------|----|
| Hombre.Normal | 1.244628 | 1.454172 | 1.0652784 | 1.1521498 | 1.344529 | 13 |
| Hombre.Tumoral | 1.072149 | 1.260640 | 0.9118417 | 0.9893626 | 1.161863 | 12 |
| Mujer.Normal | 1.123285 | 1.330310 | 0.9484763 | 1.0328522 | 1.221634 | 11 |
| Mujer.Tumoral | 1.060337 | 1.246750 | 0.9017953 | 0.9784621 | 1.149062 | 12 |

Las comparaciones múltiples se muestran en la tabla D.6.

Tabla D.6: Diferencias dos a dos entre las condiciones experimentales de sus valores $\log(DO_{ORAI1}/DO_{\beta-actina})$ estimados con el modelo seleccionado.

| | Diferencias | SEM | t | pvalor |
|-------------------------------------|-------------|---------|----------|----------|
| Hombre.Tumoral-Hombre.Normal | -0.14917 | 0.11144 | -1.33860 | 5.44E-01 |
| Mujer.Normal-Hombre.Normal | -0.10258 | 0.11404 | -0.89949 | 8.05E-01 |
| Mujer.Tumoral-Hombre.Normal | -0.16025 | 0.11144 | -1.43801 | 4.83E-01 |
| Mujer.Normal-Hombre.Tumoral | 0.04659 | 0.11620 | 0.40096 | 9.78E-01 |
| Mujer.Tumoral-Hombre.Tumoral | -0.01108 | 0.11365 | -0.09749 | 1.00E+00 |
| Mujer.Tumoral-Mujer.Normal | -0.05767 | 0.11620 | -0.49631 | 9.59E-01 |

D.2. Modelado de la expresión de la proteína ORAI2 en muestras de tejido humano de cáncer de colon y normal

Con el mismo procedimiento llevado a cabo en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos”, se ha modelizado la expresión de la proteína ORAI2. En primer lugar, se generó el modelo sobreóptimo con los dos factores fijos (*Tejido* y *Sexo*) y los dos factores aleatorios cruzados (*Paciente* y *Gel*), de tal manera que en el primero se consideró tanto el intercepto aleatorio como la pendiente aleatoria debida al factor fijo *Tejido* ($1 + \text{Tejido}|\text{Paciente}$) y en el segundo factor aleatorio se consideró el intercepto aleatorio, la pendiente aleatoria para *Tejido* y la pendiente aleatoria para *Sexo* ($1 + \text{Tejido} + \text{Sexo}|\text{Gel}$). En la tabla D.7 se muestra el valor de cAIC para cada una de las estructuras aleatorias evaluadas, donde la estructura aleatoria se expresa según el convenio usado en el paquete lme4 de R:

Tabla D.7: Selección de modelos para la variable respuesta $\log(\text{ORAI2})$ con cAIC.

| modelo | Parte Aleatoria (Anotación según lme4) | cAIC |
|------------------|--|----------|
| Pt_Gts | $(1 + \text{Tejido} \text{Paciente}) + (1 + \text{Tejido} + \text{Sexo} \text{Gel})$ | error |
| Pt_Gt | $(1 + \text{Tejido} \text{Paciente}) + (1 + \text{Tejido} \text{Gel})$ | 47.49817 |
| Pt_G | $(1 + \text{Tejido} \text{Paciente}) + (1 \text{Gel})$ | 46.51628 |
| P_Gt | $(1 \text{Paciente}) + (1 + \text{Tejido} \text{Gel})$ | 49.23382 |
| P_G | $(1 \text{Paciente}) + (1 \text{Gel})$ | 54.49607 |
| Pacientet | $(1 + \text{Tejido} \text{Paciente})$ | 68.07678 |
| Paciente | (1Paciente) | 66.99470 |
| Gelts | $(1 + \text{Tejido} + \text{Sexo} \text{Gel})$ | 67.97068 |
| Gelt | $(1 + \text{Tejido} \text{Gel})$ | 74.04612 |
| Gel | (1Gel) | 70.61029 |
| Fijo | | 80.67788 |

Como puede comprobarse, el valor más pequeño de cAIC es para el modelo cuasi sobreóptimo (Pt_Gts). No obstante, los modelos con menor cAIC estiman unas matrices de varianzas-covarianzas para la parte aleatoria peligrosas en el sentido de que hay efectos aleatorios con varianzas igual a cero y correlaciones entre efectos ± 1 , por lo que estos efectos aleatorios se eliminan del modelo (Barr et al. (2013), Matuschek et al. (2017)). Así, de los modelos evaluados, en las tablas D.8}-D.9} se muestran las estimaciones de la varianza de los diferentes efectos aleatorios y sus correlaciones para aquellos modelos en los que se han identificado algunos de estos problemas.

Tabla D.8: Parte aleatoria de los modelos problemáticos para $\log(DO_{ORAI2}/DO_{\beta\text{-actina}})$. Se muestra la salida de R de la función lmer usada para estimar el modelo lineal mixto

```
#Modelo Pt_Gts
Linear mixed model fit by REML ['lmerMod']
Formula: logORAI2 ~ Tejido * Sexo + (1 + Tejido | Paciente) + (1 + Tejido +      Sexo | Gel)
Data: datosr
REML criterion at convergence: 48.8491
Random effects:
Groups   Name                Std.Dev. Corr
Gel      (Intercept)          0.3743
         TejidoTUMORAL  0.4360  -0.26
         SexoMUJER      0.4616   0.44 -0.98
Paciente (Intercept)    0.3158
         TejidoTUMORAL  0.1870   1.00
Residual                    0.1770
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings

#Modelo Gelts
Linear mixed model fit by REML ['lmerMod']
Formula: logORAI2 ~ Tejido * Sexo + (1 + Tejido + Sexo | Gel)
Data: datosr
REML criterion at convergence: 66.6016
Random effects:
Groups   Name                Std.Dev. Corr
Gel      (Intercept)          0.5180
         TejidoTUMORAL  0.3710   0.06
         SexoMUJER      0.3524  -0.94 -0.40
Residual                    0.4027
```

No obstante, a pesar de que las estimaciones de la parte aleatoria no fueron problemáticas, sí que se obtuvieron problemas en la estimación de los intervalos de confianza de otros modelos: problemas de convergencia de los algoritmos, correlaciones cuyos intervalos de confianza eran ± 1 y varianzas cuyo límite inferior era 0. Por lo tanto, estos también fueron descargados:

Tabla D.9: Parte aleatoria de los modelos problemáticos para $\log(DO_{ORAI2}/DO_{\beta\text{-actina}})$. Se muestra la salida de R de la función lmer usada para estimar el modelo lineal mixto

```
#Modelo Pt_G
Linear mixed model fit by REML ['lmerMod']
Formula: logORAI2 ~ Tejido * Sexo + (1 + Tejido | Paciente) + (1 | GelORAI2)
Data: datosr
REML criterion at convergence: 58.4477
Random effects:
Groups   Name                Std.Dev. Corr
GelORAI2 (Intercept)  0.3296
Paciente (Intercept)  0.3036
          TejidoTUMORAL 0.4069  -0.13
Residual                    0.3126

          2.5 %    97.5 %
.sig01      0.15732909  0.6084496
.sig02      0.01021056  0.5745260
.sig03     -1.00000000  1.0000000
.sig04      0.00000000  0.7827021
.sigma      0.23250706  0.4469461
(Intercept) -1.03809891 -0.2128750
TejidoTUMORAL -0.58560392  0.1242176
SexoMUJER   -0.09364782  0.9625812
```

```
#Modelo P_Gt
Linear mixed model fit by REML ['lmerMod']
Formula: logORAI2 ~ Tejido * Sexo + (1 | Paciente) + (1 + Tejido | GelORAI2)
Data: datosr
REML criterion at convergence: 59.8493
Random effects:
Groups   Name                Std.Dev. Corr
GelORAI2 (Intercept)  0.3413
          TejidoTUMORAL 0.3464  -0.30
Paciente (Intercept)  0.3387

          2.5 %    97.5 %
.sig01      0.05635463  0.655106566
.sig02     -1.00000000  1.000000000
.sig03      0.00000000  0.679150698
.sig04      0.15017379  0.579332914
.sigma      0.22590076  0.448468373
(Intercept) -1.04626742 -0.178873484
TejidoTUMORAL -0.71073724  0.005583629
SexoMUJER   -0.16561903  0.950741862
TejidoTUMORAL:SexoMUJER -0.19368573  0.683963990
```

En consecuencia, tras descartar los modelos problemáticos, el modelo que presenta un menor valor de cAIC y no presenta ninguno de los problemas considerados es el *modelo P_G* (tabla D.10)

Tabla D.10: Parte aleatoria del modelo seleccionado para $\log(\text{DO}_{\text{ORAI2}}/\text{DO}_{\beta\text{-actina}})$. Se muestra la salida de R de la función lmer usada para estimar el modelo lineal mixto (*continuación*)

```
#Modelo_P_G
Linear mixed model fit by REML ['lmerMod']
Formula: logORAI2 ~ Tejido * Sexo + (1 | Paciente) + (1 | GelORAI2)
REML criterion at convergence: 61.9163
Random effects:
Groups   Name          Std.Dev.
GelORAI2 (Intercept) 0.3134
Paciente (Intercept) 0.3466
Residual                0.3756
Number of obs: 41, groups: GelORAI2, 10; Paciente, 8
```

El modelo más adecuado para modelizar la expresión de ORAI2 es lineal mixto (ecuación D.2):

$$\log\left(\frac{\text{DO}_{\text{ORAI2}}}{\text{DO}_{\beta\text{-actina}}}\right)_{ijmpr} = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{Sexo}_j + b_{0,\text{gel}m} + b_{0,\text{paciente}p} + \epsilon_{ijmpr} \quad (\text{D.2})$$

donde $\log(\text{DO}_{\text{ORAI2}}/\text{DO}_{\beta\text{-actina}})_{ijmpr}$ es el valor de la expresión de ORAI2 medido en la réplica r de la muestra cuyo tejido es i (Tumoral o Normal), sexo j (Hombre o Mujer) medidas en el gel m . En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel Hombre.Normal); β_1 es el cambio de expresión en el tejido tumoral respecto del de referencia (Hombre.Normal), β_2 el cambio en la expresión de la proteína de interés en las muestras de mujer respecto de las de referencia (Hombre.Normal); β_3 el efecto de la interacción entre los factores *Tejido* y *Sexo* que indica si el cambio entre el tejido tumoral y normal es diferente en hombres que en mujeres (en caso de ser significativo); $b_{0,\text{gel}m}$ es el intercepto aleatorio debido al gel que indica la desviación respecto de β_0 que hay en gel m , y $b_{0,\text{paciente}p}$ es el intercepto aleatorio debido a paciente p que indica la desviación respecto de β_0 que muestra el paciente p . En la tabla D.7 se muestra la selección de la parte aleatoria y en la tabla D.11 la selección de la parte fija.

Tabla D.11: Selección de modelos para la variable respuesta $\log(\text{ORAI2})$ con AIC, BIC y LR.

| Parte Fija | AIC |
|-----------------------------|--------|
| Tejido + Sexo + Tejido:Sexo | 69.761 |
| Tejido + Sexo | 68.905 |
| Tejido | 70.255 |
| nulo | 71.997 |

En cuanto al diagnóstico del modelo o la evaluación del cumplimiento de los supuestos, en la figura D.2 se muestran los métodos gráficos y en la tabla D.12 los métodos basados en contrastes de hipótesis.

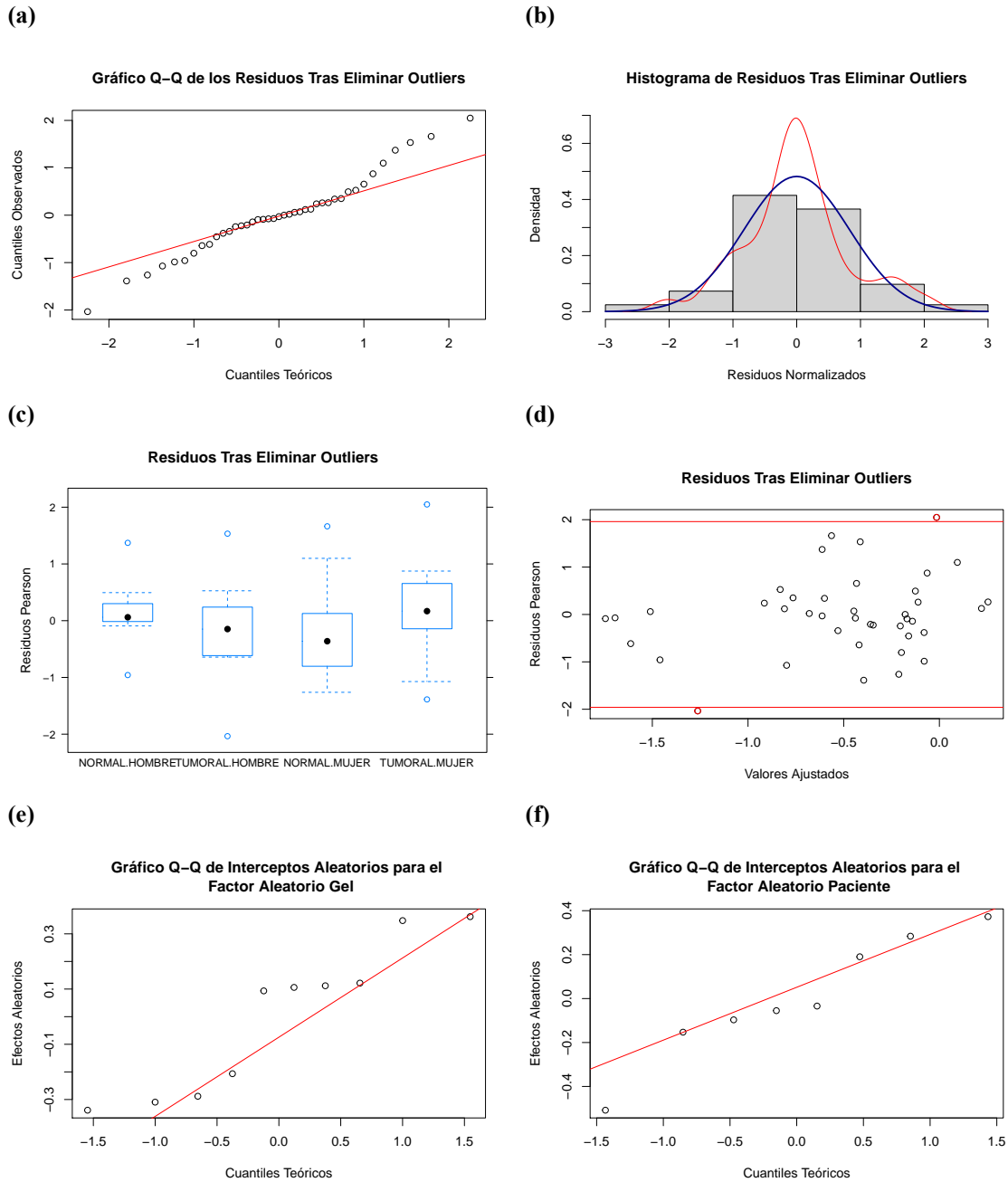


Figura D.2: Diagnóstico del modelo lineal de efectos fijos para la variable respuesta $\log(DO_{ORAI2}/DO_{\beta-actina})$ tras eliminar los *outliers*. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como *outliers*; e) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Paciente, y f) Gráfico Q-Q de los efectos aleatorios del factor aleatorio Gel.

Tabla D.12: Supuestos del modelo seleccionado (p valores).

| | Normalidad | | | Homocedasticidad | |
|----------------------------|------------|-------|-------|------------------|-----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 0.426 | 0.699 | 0.175 | 0.9530 | 0.3270352 |
| Paciente intercepto | 0.690 | 0.871 | 0.682 | | |
| Gel intercepto | 0.114 | 0.653 | 0.114 | | |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test. de Breush-Pagan.

En la tabla D.13 se muestran las estimaciones de los coeficientes del modelo y en la tabla D.14 su matriz de varianzas y covarianzas.

Tabla D.13: Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para $\log(DO_{ORAI2}/DO_{\beta-actina})$.

| | Estimación | SEM | t | gdl | p-valor |
|-------------------|------------|-----------|-----------|----------|-----------|
| β_0 | -0.6636667 | 0.2269292 | -2.924554 | 6.197904 | 0.0254967 |
| $\beta_{Tumoral}$ | -0.2346966 | 0.1174812 | -1.997738 | 6.197904 | 0.0911945 |
| β_{Mujer} | 0.5064257 | 0.2790302 | 1.814949 | 6.197904 | 0.1178802 |

Abreviaturas:

gdl: grados de libertad

Tabla D.14: Covarianzas entre los coeficientes del modelo final para $\log(DO_{ORAI2}/DO_{\beta-actina})$.

| | β_0 | $\beta_{Tumoral}$ | β_{Mujer} |
|-------------------|------------|-------------------|-----------------|
| β_0 | 0.0514969 | -0.0066387 | -0.0379441 |
| $\beta_{Tumoral}$ | -0.0066387 | 0.0138018 | -0.0004035 |
| β_{Mujer} | -0.0379441 | -0.0004035 | 0.0778579 |

En cuanto a la estructura de la parte aleatoria, la varianza de los efectos aleatorios se muestra en la tabla en la tabla D.15.

Tabla D.15: Estimaciones de la varianza de la parte aleatoria del modelo mixto final para $\log(DO_{ORAI2}/DO_{\beta-actina})$.

| Componente Aleatorio | LI | Estimación SEM | LS |
|----------------------------|-----------|----------------|-----------|
| Paciente intercepto | 0.1432189 | 0.3493150 | 0.6023073 |
| Gel intercepto | 0.1420544 | 0.3231921 | 0.6023117 |
| sigma | 0.2848465 | 0.3729616 | 0.4892397 |

La tabla ANOVA del modelo seleccionado se muestra en la tabla D.16. Evidentemente, el factor *Sexo* no es significativo, si bien es cierto que el criterio AIC indica que sí ha de estar en el modelo. Por lo tanto, este se mantiene y se eleva el nivel de significación a 0.07. En cuanto a la interacción, en ningún caso resulta ser necesaria en el modelo; es decir, no es significativa.

Tabla D.16: Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado para $\log(DO_{ORAI2}/DO_{\beta\text{-actina}})$.

| | Sum Sq | gdl | p-valor |
|--------------|----------|-----|-----------|
| Tumor | 3.990956 | 1 | 0.0457451 |
| Sexo | 3.294041 | 1 | 0.0695317 |

Abreviaturas:

gdl: grados de libertad

Las estimaciones de la expresión de ORAI2, expresado como $DO_{ORAI2}/DO_{\beta\text{-actina}}$, se muestran en la tabla D.17.

Tabla D.17: Estimaciones de la expresión de ORAI2 ($DO_{ORAI2}/DO_{\beta\text{-actina}}$) con el modelo seleccionado.

| Condiciones | ratio DO | Límite Superior | Límite Inferior | log(Señal-SEM) | log(Señal+SEM) | n |
|-----------------------|-----------|-----------------|-----------------|----------------|----------------|----|
| Hombre.Normal | 0.5457812 | 0.8595578 | 0.3465470 | 0.4328900 | 0.6881128 | 11 |
| Hombre.Tumoral | 0.3824155 | 0.6050913 | 0.2416852 | 0.3025933 | 0.4832944 | 10 |
| Mujer.Normal | 0.8021318 | 1.2773285 | 0.5037196 | 0.6326383 | 1.0170352 | 10 |
| Mujer.Tumoral | 0.7220771 | 1.1495340 | 0.4535711 | 0.5695789 | 0.9154050 | 10 |

Puesto que no existe término de interacción y que el tumor es significativo y que el sexo está muy cercano al nivel de significación 0.05 pero inferior al nivel de significación 0.1, se consideran que ambos son significativos (en los modelos lineales mixtos los contrastes de F y de t hay que tomarlos con cautela cerca del nivel de significación y, por ello, en este caso se ha considerado la posibilidad de que sea significativo en este caso). Además, puesto que cada uno de los dos factores únicamente tiene dos niveles, no es necesario realizar comparaciones múltiples dentro de cada factor. Por lo tanto, en este caso se consideran que las cuatro condiciones experimentales son diferentes entre sí.

D.3. Modelado de la expresión de la proteína TRPC1 en muestras de tejido humano de cáncer de colon y normal

Con el mismo procedimiento llevado a cabo en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos”, se ha modernizado la expresión de la proteína TRPC1. En primer lugar, se generó el modelo sobreóptimo con los dos factores fijos (*Tejido* y *Sexo*) y los dos factores aleatorios cruzados (*Paciente* y *Gel*).

Tras evaluar el modelo sobreóptimo con el fin de encontrar tanto la parte fija como la parte aleatoria más adecuada, se obtuvo que el modelo más adecuado para modelizar la expresión de TRPC1 es un modelo de efectos fijos (ecuación D.3):

$$T \left(\frac{DO_{TRPC1}}{DO_{\beta\text{-actina}}}_{ijr} \right) = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{sexo}_j + \epsilon_{ijr} \quad (D.3)$$

donde $T \left(\frac{DO_{TRPC1}}{DO_{\beta\text{-actina}}}_{ijr} \right)$ es el valor de la variable respuesta transformada según la familia de Box-Cox propuesto por Hawkins and Weisberg (2017) medido en la réplica r de la muestra cuyo tejido es i (Tumoral o Normal) y sexo j (Hombre o Mujer). En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel Hombre.Normal); β_1 es el cambio de expresión en el tejido tumoral respecto del de referencia (Hombre.Normal), β_2 el cambio en la expresión de la proteína de interés en las muestras de mujer respecto de las de referencia (Hombre.Normal), y β_3 el efecto de la interacción entre los factores *Tejido* y *Sexo* que indica si el cambio entre el tejido tumoral y normal es diferente en hombres que en mujeres (en caso de ser significativo). En cuanto al diagnóstico del modelo o la evaluación del cumplimiento de los supuestos, en la tabla D.18 se muestran los métodos basados en contrastes de hipótesis y en la figura D.3 los métodos gráficos .

Tabla D.18: Supuestos del modelo seleccionado (p valores).

| | Normalidad | | | Homocedasticidad | |
|----------------------------|------------|-------|---------|------------------|-----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 0.00262 | 0.152 | 6.5e-06 | 0.0937 | 0.6198766 |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test. de Breush-Pagan.

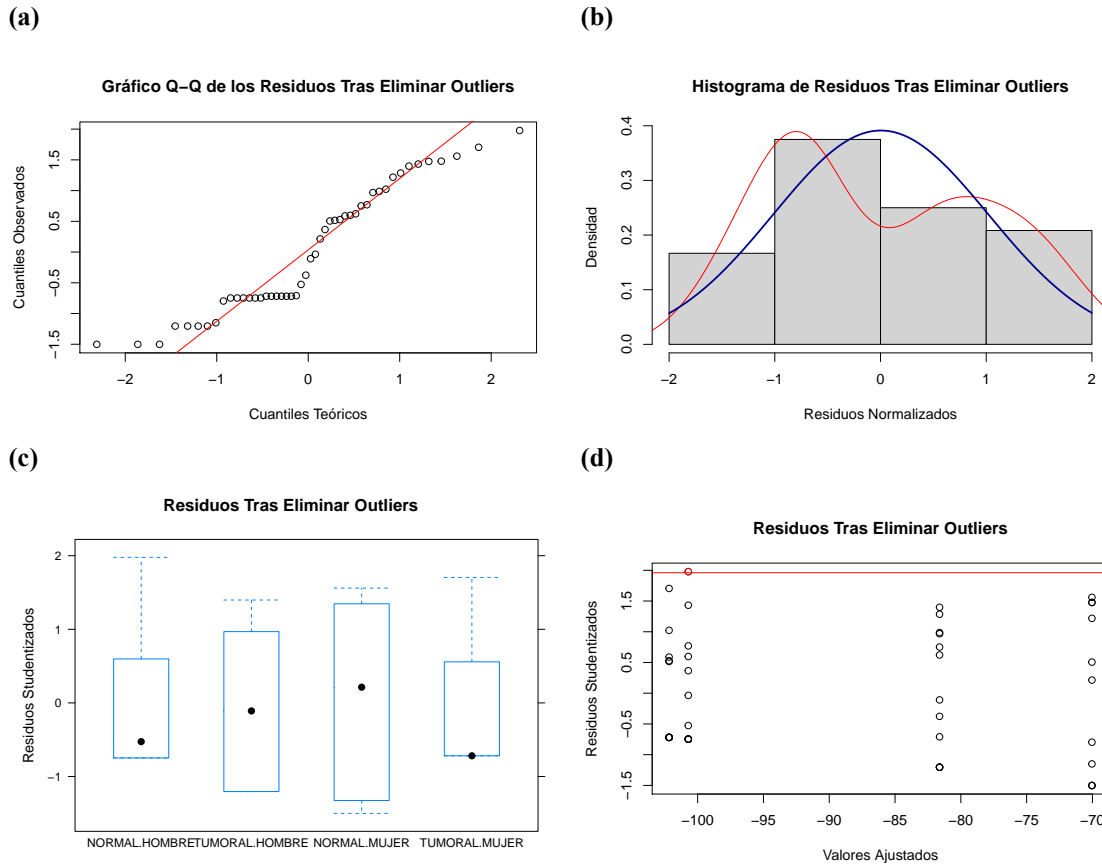


Figura D.3: Diagnóstico del modelo lineal de efectos fijos para la variable respuesta T ($DO_{TRPC1}/DO_{\beta-actina}$) tras eliminar los *outliers*. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como *outliers*.

En la tabla D.19 se muestran las estimaciones de los coeficientes del modelo y en la tabla D.20 su matriz de varianzas y covarianzas.

Tabla D.19: Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo seleccionado para T ($DO_{TRPC1}/DO_{\beta-actina}$).

| | Estimación | SEM | t | p-valor |
|-------------------------|------------|----------|-----------|-----------|
| β_0 | -100.71085 | 12.33542 | -8.164362 | 0 |
| $\beta_{Tumoral}$ | 19.09530 | 17.44492 | 1.094605 | 0.2796457 |
| β_{Mujer} | 30.67329 | 18.22062 | 1.683438 | 0.0993735 |
| $\beta_{Tumoral:Mujer}$ | -51.22785 | 25.76785 | -1.988053 | 0.0530498 |

Tabla D.20: Covarianzas entre los coeficientes del modelo seleccionado para T ($DO_{TRPC1}/DO_{\beta-actina}$).

| | β_0 | $\beta_{Tumoral}$ | β_{Mujer} | $\beta_{Tumoral:Mujer}$ |
|-------------------------|-----------|-------------------|-----------------|-------------------------|
| β_0 | 152.1626 | -152.1626 | -152.1626 | -152.1626 |
| $\beta_{Tumoral}$ | -152.1626 | 304.3252 | 152.1626 | 152.1626 |
| β_{Mujer} | -152.1626 | 152.1626 | 331.9912 | 152.16261 |
| $\beta_{Tumoral:Mujer}$ | -152.1626 | 152.1626 | 152.1626 | 331.9912 |

La tabla ANOVA del modelo seleccionado se muestra en la tabla D.21. Evidentemente, ningún factor tiene efecto sobre la variable respuesta y, por ende, siendo estricto, habría que eliminar la interacción entre los factores y demás variables innecesarias. No obstante, puesto que se pretende estimar el valor de la variable respuesta para cada una de las condiciones experimentales y, además, que hay suficientes grados de libertad para estimar el error aleatorio, entonces no se han eliminado del modelo.

Tabla D.21: Tabla ANOVA (tipo II) del modelo seleccionado para T ($DO_{TRPC1}/DO_{\beta-actina}$).

| | Sum Sq | gdl | F | p-valor |
|--------------------|------------|-----|-----------|-----------|
| Tejido | 230.6475 | 1 | 0.1165997 | 0.7343775 |
| Sexo | 305.0337 | 1 | 0.1542043 | 0.6964455 |
| Tejido:Sexo | 7818.2044 | 1 | 3.9523528 | 0.0530498 |
| Residuals | 87037.0156 | 44 | | |

Abreviaturas:

gdl: grados de libertad

Las estimaciones de la expresión de TRPC1, expresado como $DO_{TRPC1}/DO_{\beta-actina}$, se muestran en la tabla D.22.

Tabla D.22: Estimaciones de la expresión de TRPC1 ($DO_{TRPC1}/DO_{\beta-actina}$) con el modelo seleccionado.

| Condiciones | ratio DO | Límite Superior | Límite Inferior | T(Señal-SEM) | T(Señal+SEM) | n |
|-----------------------|-----------|-----------------|-----------------|--------------|--------------|----|
| Hombre.Normal | 0.0243586 | 0.0495199 | 0.0049338 | 0.0141765 | 0.0359154 | 13 |
| Hombre.Tumoral | 0.0429851 | 0.0757183 | 0.0194498 | 0.0304972 | 0.0576435 | 13 |
| Mujer.Normal | 0.0566660 | 0.1014249 | 0.0276157 | 0.0410101 | 0.0759241 | 11 |
| Mujer.Tumoral | 0.0230893 | 0.0503567 | 0.0024316 | 0.0122258 | 0.0355303 | 11 |

Las comparaciones múltiples se muestran en la tabla D.23.

Tabla D.23: Diferencias dos a dos entre las condiciones experimentales de sus valores $T(DO_{TRPC1}/DO_{\beta\text{-actina}})$ estimados con el modelo seleccionado.

| | Diferencias | SEM | t | pvalor |
|-------------------------------------|--------------------|------------|------------|---------------|
| Hombre.Tumoral-Hombre.Normal | 19.095297 | 17.44492 | 1.0946050 | 0.6944973 |
| Mujer.Normal-Hombre.Normal | 30.673295 | 18.22062 | 1.6834382 | 0.3442098 |
| Mujer.Tumoral-Hombre.Normal | -1.459256 | 18.22062 | -0.0800881 | 0.9998127 |
| Mujer.Normal-Hombre.Tumoral | 11.577998 | 18.22062 | 0.6354336 | 0.9199362 |
| Mujer.Tumoral-Hombre.Tumoral | -20.554553 | 18.22062 | -1.1280927 | 0.6741924 |
| Mujer.Tumoral-Mujer.Normal | -32.132551 | 18.96463 | -1.6943413 | 0.3386319 |

D.4. Modelado de la expresión de la proteína STIM1 en muestras de tejido humano de cáncer de colon y normal

Con el mismo procedimiento llevado a cabo en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos”, se ha modelizado la expresión de la proteína STIM1. En primer lugar, se generó el modelo sobreóptimo con los dos factores fijos (*Tejido* y *Sexo*) y los dos factores aleatorios cruzados (*Paciente* y *Gel*), de tal manera que en el primero se consideró tanto el intercepto aleatorio como la pendiente aleatoria debida al factor fijo *Tejido* ($1 + \text{Tejido}|\text{Paciente}$) y en el segundo factor aleatorio se consideró el intercepto aleatorio, la pendiente aleatoria para *Tejido* y la pendiente aleatoria para *Sexo* ($1 + \text{Tejido} + \text{Sexo}|\text{Gel}$). En la tabla D.24 se muestra el valor de cAIC para cada una de las estructuras aleatorias evaluadas, donde la estructura aleatoria se expresa según el convenio usado en el paquete lme4 de R:

Tabla D.24: Selección de modelos para la variable respuesta $\log(\text{STIM1})$ con cAIC.

| modelo | Parte Aleatoria (Anotación según lme4) | cAIC |
|------------------|---|------------|
| Pt_Gts | (1 + Tejido Paciente) + (1 + Tejido + Sexo Gel) | -76.794943 |
| Pt_Gt | (1 + Tejido Paciente) + (1 + Tejido Gel) | -48.688705 |
| Pt_G | (1 + Tejido Paciente) + (1 Gel) | 3.720895 |
| P_Gt | (1 Paciente) + (1 + Tejido Gel) | error |
| P_G | (1 Paciente) + (1 Gel) | 82.999081 |
| Pacientet | (1 + Tejido Paciente) | 41.000314 |
| Paciente | (1 Paciente) | 78.109132 |
| Gelts | (1 + Tejido + Sexo Gel) | 62.064711 |
| Gelt | (1 + Tejido Gel) | 75.138184 |
| Gel | (1 Gel) | 82.999081 |
| Fijo | | 80.109132 |

Como puede comprobarse, el valor más pequeño de cAIC es para el modelo cuasi sobreóptimo que puede (modelo DTL). No obstante, algunos de los modelos con menor cAIC estiman unas matrices de varianzas-covarianzas para la parte aleatoria peligrosas en el sentido de que hay efectos aleatorios con varianzas igual a cero (o su límite inferior es cero) y correlaciones entre efectos ± 1 (o al menos uno de sus límites inferior y/o superior). En consecuencia, estos modelos se descartan (Barr et al. (2013), Matuschek et al. (2017)). Así, de los modelos evaluados (en orden creciente de cAIC), en las tablas D.25-D.27 se muestran las estimaciones de la varianza de los diferentes efectos aleatorios y sus correlaciones (y sus intervalos de confianza) para aquellos modelos con un cAIC inferior al modelo seleccionado en los que se han identificado algunos de estos problemas.

Tabla D.25: Parte aleatoria de los modelos problemáticos para $\log(DO_{STIM1}/DO_{\beta-actina})$. Se muestra la salida de R de la función `lmer()` usada para estimar el modelo lineal mixto y la función `confint(level = 0.95)` para los intervalos de confianza.

```
#Modelo Pt_Gts
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM1 ~ Tejido * Sexo + (1 + Tejido | Paciente) + (1 + Tejido + Sexo | Gel)
REML criterion at convergence: 20.2331
Random effects:
  Groups      Name                Std.Dev. Corr
  Gel         (Intercept)          0.24603
              TejidoTUMORAL      0.45645  -0.01
              SexoMUJER          0.36932  -0.39 -0.91
  Paciente    (Intercept)          0.34695
              TejidoTUMORAL      1.03995  -0.96
  Residual                                0.08417
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings

#Modelo Pt_Gt
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM1 ~ Tejido * Sexo + (1 + Tejido | Paciente) + (1 + Tejido | Gel)
REML criterion at convergence: 27.911
Random effects:
  Groups      Name                Std.Dev. Corr
  Gel         (Intercept)          0.2218
              TejidoTUMORAL      0.5829  -0.76
  Paciente    (Intercept)          0.5200
              TejidoTUMORAL      1.1043  -0.97
  Residual                                0.0950
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings

#Modelo Pt_G
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM1 ~ Tejido * Sexo + (1 + Tejido | Paciente) + (1 | Gel)
  Data: datosr
REML criterion at convergence: 41.9295
Random effects:
  Groups      Name                Std.Dev. Corr
  Gel         (Intercept)          0.3490
  Paciente    (Intercept)          0.5258
              TejidoTUMORAL      1.1537  -0.99
  Residual                                0.1825

#Modelo P_Gt
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM1 ~ Tejido * Sexo + (1 | Paciente) + (1 + Tejido | Gel)
  Data: datosr
REML criterion at convergence: 73.7811
Random effects:
  Groups      Name                Std.Dev. Corr
  Gel         (Intercept)          0.2565
              TejidoTUMORAL      0.6559  -0.71
  Paciente    (Intercept)          0.0000
  Residual                                0.4274
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
```

Tabla D.26: Parte aleatoria de los modelos problemáticos para $\log(DO_{STIM1}/DO_{\beta\text{-actina}})$. Se muestra la salida de R de la función lmer usada para estimar el modelo lineal mixto (*continuación*)

```
#Modelo P_G
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM1 ~ Tejido * Sexo + (1 | Paciente) + (1 | Gel)
Data: datosr
REML criterion at convergence: 76.6303
Random effects:
Groups   Name          Std.Dev.
Gel      (Intercept)  0.2070
Paciente (Intercept) 0.0000
Residual                0.5179
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings

#Modelo Pacientet
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM1 ~ Tejido * Sexo + (1 + Tejido | Paciente)
Data: datosr
REML criterion at convergence: 51.5562
Random effects:
Groups   Name          Std.Dev. Corr
Paciente (Intercept)  0.4929
          TejidoTUMORAL 1.0565  -0.97
Residual                0.3090

#Modelo Paciente
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM1 ~ Tejido * Sexo + (1 | Paciente)
Data: datosr
REML criterion at convergence: 77.1537
Random effects:
Groups   Name          Std.Dev.
Paciente (Intercept) 0.0000
Residual                0.5513
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings

#Modelo Gelts
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM1 ~ Tejido * Sexo + (1 + Tejido + Sexo | Gel)
Data: datosr
REML criterion at convergence: 69.5994
Random effects:
Groups   Name          Std.Dev. Corr
Gel      (Intercept)  0.1925
          TejidoTUMORAL 0.5220  -0.93
          SexoMUJER    1.0763   0.79 -0.96
Residual                0.3700
```

Tabla D.27: Parte aleatoria de los modelos problemáticos para $\log(\text{DO}_{\text{STIM1}}/\text{DO}_{\beta\text{-actina}})$. Se muestra la salida de R de la función lmer usada para estimar el modelo lineal mixto (*continuación*)

```
#Modelo Gelt

Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM1 ~ Tejido * Sexo + (1 + Tejido | Gel)
Data: datosr
REML criterion at convergence: 73.7811
Random effects:
Groups   Name                Std.Dev. Corr
Gel      (Intercept)          0.2565
         TejidoTUMORAL  0.6558   -0.99
Residual                    0.4274
```

En consecuencia, tras descartar los modelos problemáticos, se tiene que el único modelo lineal mixto posible tiene una cAIC superior a la del modelo lineal ordinario (de efectos fijos), por lo que se selecciona un modelo sin estructura aleatoria más allá de la del puro error experimental. Posteriormente, mediante la AIC, se evalúa la parte fija que, al tratarse de un modelo de efectos fijos, es la única parte que contiene. Como resultado, en la tabla D.28 se muestra la selección de la parte fija que, como puede verse, resultan ser tanto los dos efectos principales y su interacción.

Tabla D.28: Selección de modelos para la variable respuesta $\log(\text{STIM1})$ con AIC, BIC y LR.

| Parte Fija | AIC |
|------------------------------------|--------|
| Tejido + Sexo + Tejido:Sexo | 79.920 |
| Tejido + Sexo | 83.405 |
| Tejido | 81.981 |
| Sexo | 88.475 |
| nulo | 86.657 |

Así, tras evaluar el modelo sobreóptimo con el fin de encontrar tanto la parte fija como la parte aleatoria más adecuada (ninguna en este caso), se obtuvo que el modelo más adecuado para modelizar la expresión de STIM1 es (ecuación D.4):

$$\log \left(\frac{\text{DO}_{\text{STIM1}}}{\text{DO}_{\beta\text{-actina}}_{ijr}} \right) = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{sexo}_j + \epsilon_{ijr} \quad (\text{D.4})$$

donde $\log \left(\text{DO}_{\text{STIM1}}/\text{DO}_{\beta\text{-actina}} \right)_{ijr}$ es el valor de la expresión de STIM1 medido en la réplica r de la muestra cuyo tejido es i (Tumoral o Normal) y sexo j (Hombre o Mujer). En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel Hombre.Normal); β_1 es el cambio de expresión en el tejido tumoral respecto del de referencia (Hombre.Normal), β_2 el cambio en la expresión de la proteína de interés en las muestras de mujer

respecto de las de referencia (Hombre.Normal), y β_3 el efecto de la interacción entre los factores *Tejido* y *Sexo* que indica si el cambio entre el tejido tumoral y normal es diferente en hombres que en mujeres (en caso de ser significativo).

En cuanto al diagnóstico del modelo o la evaluación del cumplimiento de los supuestos, en la figura D.4 se muestran los métodos gráficos y en la tabla D.29 los métodos basados en contrastes de hipótesis.

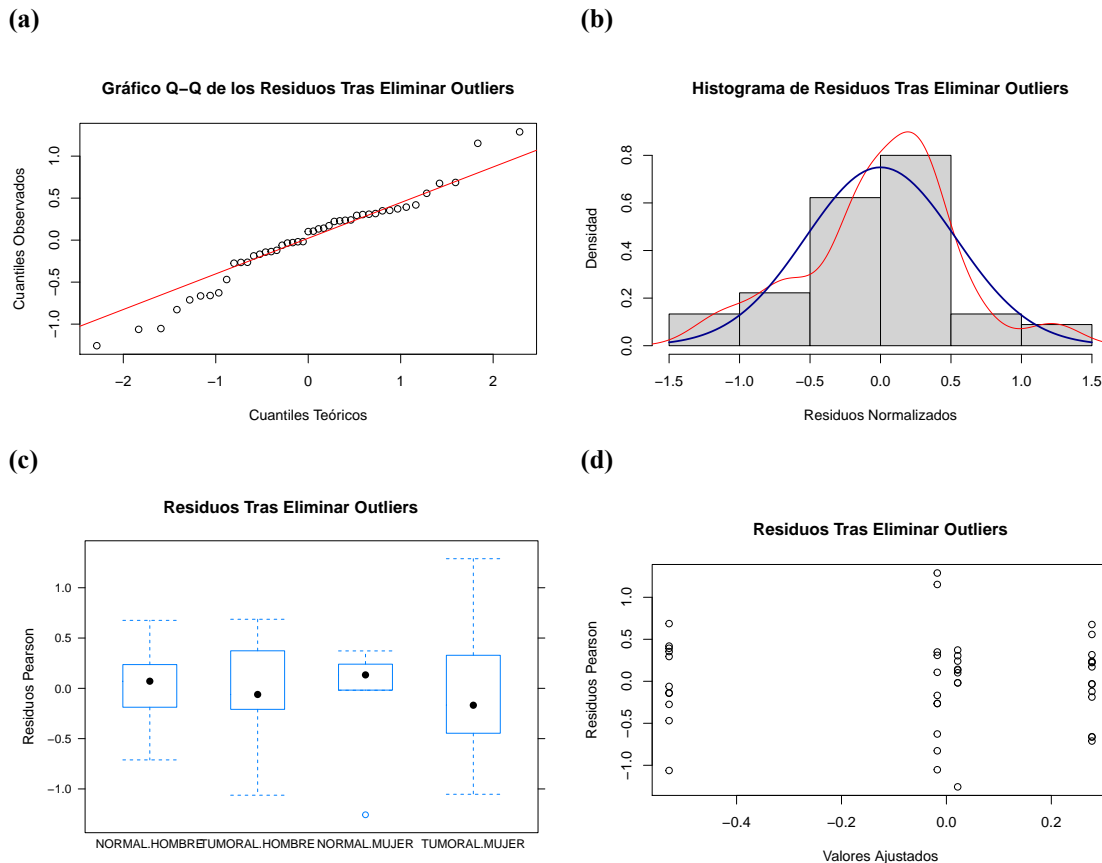


Figura D.4: Diagnóstico del modelo lineal de efectos fijos para la variable respuesta $\log(DO_{STIM1}/DO_{\beta\text{-actina}})$ tras eliminar los *outliers*. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como *outliers*.

Tabla D.29: Supuestos del modelo seleccionado (p valores).

| | Normalidad | | | Homocedasticidad | |
|----------------------------|------------|-------|-------|------------------|-----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 0.219 | 0.762 | 0.251 | 0.872 | 0.2938139 |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test. de Breush-Pagan.

En la tabla D.30 se muestran las estimaciones de los coeficientes del modelo y en la tabla D.31 su matriz de varianzas y covarianzas.

Tabla D.30: Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo seleccionado para $\log(DO_{STIM1}/DO_{\beta-actina})$.

| | Estimación | SEM | t | p-valor |
|-------------------------|------------|-----------|-----------|-----------|
| β_0 | 0.2774002 | 0.1473371 | 1.882759 | 0.0668418 |
| $\beta_{Tumoral}$ | -0.8060665 | 0.2221190 | -3.628985 | 0.0007813 |
| β_{Mujer} | -0.2559009 | 0.2355346 | -1.086468 | 0.2836183 |
| $\beta_{Tumoral:Mujer}$ | 0.7671689 | 0.3327669 | 2.305424 | 0.0262773 |

Abreviaturas:

gdl: grados de libertad

Tabla D.31: Covarianzas entre los coeficientes del modelo seleccionado para $\log(DO_{STIM1}/DO_{\beta-actina})$.

| | β_0 | $\beta_{Tumoral}$ | β_{Mujer} | $\beta_{Tumoral:Mujer}$ |
|-------------------------|------------|-------------------|-----------------|-------------------------|
| β_0 | 0.0217082 | -0.0217082 | -0.0217082 | 0.0217082 |
| $\beta_{Tumoral}$ | -0.0217082 | 0.0493368 | 0.0217082 | -0.0493368 |
| β_{Mujer} | -0.0217082 | 0.0217082 | 0.0554765 | -0.0554765 |
| $\beta_{Tumoral:Mujer}$ | 0.0217082 | -0.0493368 | -0.0554765 | 0.1107338 |

La tabla ANOVA del modelo seleccionado se muestra en la tabla D.32. Evidentemente, ningún factor tiene efecto sobre la variable respuesta y, por ende, siendo estricto, habría que eliminar la interacción entre los factores y demás variables innecesarias. No obstante, puesto que se pretende estimar el valor de la variable respuesta para cada una de las condiciones experimentales y, además, que hay suficientes grados de libertad para estimar el error aleatorio, entonces no se han eliminado del modelo.

Tabla D.32: Tabla ANOVA (tipo II) del modelo seleccionado para $\log(DO_{STIM1}/DO_{\beta-actina})$.

| | Sum Sq | gdl | p-valor |
|-------------------|------------|-----|-----------|
| Tumor | 2.3946049 | 1 | 0.0076190 |
| Sexo | 0.1811154 | 1 | 0.4445579 |
| Tumor:Sexo | 1.6153023 | 1 | 0.0262773 |
| Residual | 12.4605120 | 41 | |

Abreviaturas:

gdl: grados de libertad

Las estimaciones de la expresión de STIM1, expresado como $DO_{STIM1}/DO_{\beta-actina}$, se muestran en la tabla D.33.

Tabla D.33: Estimaciones de la expresión de STIM1 ($DO_{STIM1}/DO_{\beta-actina}$) con el modelo seleccionado para $\log(DO_{STIM1}/DO_{\beta-actina})$.

| Condiciones | ratio DO | Límite Superior | Límite Inferior | log(Señal-SEM) | log(Señal+SEM) | n |
|-----------------------|-----------|-----------------|-----------------|----------------|----------------|----|
| Hombre.Normal | 1.3196945 | 1.7770478 | 0.9800487 | 1.1389003 | 1.5291886 | 14 |
| Hombre.Tumoral | 0.5893906 | 0.8244977 | 0.4213247 | 0.4991319 | 0.6959708 | 11 |
| Mujer.Normal | 1.0217321 | 1.4808458 | 0.7049596 | 0.8502182 | 1.2278454 | 9 |
| Mujer.Tumoral | 0.9827523 | 1.3747709 | 0.7025185 | 0.8322546 | 1.1604646 | 11 |

Las comparaciones múltiples se muestran en la tabla D.34.

Tabla D.34: Diferencias dos a dos entre las condiciones experimentales de sus valores $\log(DO_{STIM1}/DO_{\beta-actina})$ estimados con el modelo seleccionado.

| | Diferencias | SEM | t | pvalor |
|-------------------------------------|-------------|-----------|------------|-------------|
| Hombre.Tumoral-Hombre.Normal | -0.8060665 | 0.2221190 | -3.6289852 | 0.0040101 * |
| Mujer.Normal-Hombre.Normal | -0.2559009 | 0.2355346 | -1.0864684 | 0.6990534 |
| Mujer.Tumoral-Hombre.Normal | -0.2947985 | 0.2221190 | -1.3272097 | 0.5506161 |
| Mujer.Normal-Hombre.Tumoral | 0.5501656 | 0.2477841 | 2.2203427 | 0.1344659 |
| Mujer.Tumoral-Hombre.Tumoral | 0.5112680 | 0.2350686 | 2.1749734 | 0.1468817 |
| Mujer.Tumoral-Mujer.Normal | -0.0388976 | 0.2477841 | -0.1569817 | 0.9985952 |

D.5. Modelado de la expresión de la proteína STIM2 en muestras de tejido humano de cáncer de colon y normal

Con el mismo procedimiento llevado a cabo en el apéndice “C.1 Caso Práctico: modelado de la variable respuesta *Área Bajo la Curva* (AUC) de la señal correspondiente al tamaño de los depósitos”, se ha modelizado la expresión de la proteína STIM2. En primer lugar, se generó el modelo sobreóptimo con los dos factores fijos (*Tejido* y *Sexo*) y los dos factores aleatorios cruzados (*Paciente* y *Gel*), de tal manera que en el primero se consideró tanto el intercepto aleatorio como la pendiente aleatoria debida al factor fijo *Tejido* ($1 + \text{Tejido}|\text{Paciente}$) y en el segundo factor aleatorio se consideró el intercepto aleatorio, la pendiente aleatoria para *Tejido* y la pendiente aleatoria para *Sexo* ($1 + \text{Tejido} + \text{Sexo}|\text{Gel}$). En la tabla D.35 se muestra el valor de cAIC para cada una de las estructuras aleatorias evaluadas, donde la estructura aleatoria se expresa según el convenio usado en el paquete lme4 de R:

Tabla D.35: Selección de modelos para la variable respuesta $\log(\text{STIM2})$ con cAIC.

| modelo | Parte Aleatoria (Anotación según lme4) | cAIC |
|------------------|---|------------|
| Pt_Gts | (1 + Tejido Paciente) + (1 + Tejido + Sexo Gel) | error |
| Pt_Gt | (1 + Tejido Paciente) + (1 + Tejido Gel) | error |
| Pt_G | (1 + Tejido Paciente) + (1 Gel) | 1.6897228 |
| P_Gt | (1 Paciente) + (1 + Tejido Gel) | error |
| P_G | (1 Paciente) + (1 Gel) | 28.6127987 |
| Pacientet | (1 + Tejido Paciente) | error |
| Paciente | (1 Paciente) | 28.6127987 |
| Gelts | (1 + Tejido + Sexo Gel) | error |
| Gelt | (1 + Tejido Gel) | -0.9797899 |
| Gel | (1 Gel) | 26.6849549 |
| Fijo | | 60.61730 |

Como puede comprobarse, el valor más pequeño de cAIC es para el *modelo Gelt* seguido del modelo *Pt_G*. No obstante, algunos de los modelos con menor cAIC estiman unas matrices de varianzas-covarianzas para la parte aleatoria peligrosas en el sentido de que hay efectos aleatorios con varianzas igual a cero (o su límite inferior es cero) y correlaciones entre efectos ± 1 (o al menos uno de sus límites inferior y/o superior). En consecuencia, estos modelos se descartan (Barr et al. (2013), Matuschek et al. (2017)). Así, de los modelos evaluados, en las tablas D.36-D.37 se muestran las estimaciones de la varianza de los diferentes efectos aleatorios y sus correlaciones para aquellos modelos en los que se han identificado algunos de estos problemas.

Tabla D.36: Parte aleatoria de los modelos problemáticos para $\log(DO_{STIM2}/DO_{\beta\text{-actina}})$. Se muestra la salida de R de la función `lmer()` usada para estimar el modelo lineal mixto y la función `confint(level = 0.95)` para los intervalos de confianza. Con una x se indican las estimaciones problemáticas.

```
#Modelo Pt_G
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM2 ~ Tejido * Sexo + (1 + Tejido | Paciente) + (1 | GelSTIM2)
Data: datosr
REML criterion at convergence: 17.973
Random effects:
Groups   Name                Std.Dev. Corr
GelSTIM2 (Intercept)  0.18230
Paciente (Intercept)  0.08256
          TejidoTUMORAL 0.44324  -0.14
Residual                    0.17864

          2.5 %      97.5 % Problemática
.sig01          0.0000000  0.32803232
.sig02          0.0000000  0.21982517
.sig03         -1.0000000  1.00000000          x
.sig04          0.17185555  0.74580933
.sigma          0.13073361  0.27038113

#Modelo_Gelt
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM2 ~ Tejido * Sexo + (1 + Tejido | GelSTIM2)
Data: datosr
REML criterion at convergence: 24.7612
Random effects:
Groups   Name                Std.Dev. Corr
GelSTIM2 (Intercept)  0.2472
          TejidoTUMORAL 0.3817  -1.00
Residual                    0.2500
Number of obs: 42, groups: GelSTIM2, 13
Fixed Effects:
          (Intercept)          TejidoTUMORAL          SexoMUJER  TejidoTUMORAL:SexoMUJER
          -0.20415          0.05316          0.12525          0.35036
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
```


Tabla D.37: Parte aleatoria de los modelos problemáticos para $\log(DO_{STIM2}/DO_{\beta-actina})$. Se muestra la salida de R de la función `lmer()` usada para estimar el modelo lineal mixto y la función `confint(level = 0.95)` para los intervalos de confianza. Con una x se indican las estimaciones problemáticas.

```
#Modelo Gel
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM2 ~ Tejido * Sexo + (1 | GelSTIM2)
Data: datosr
REML criterion at convergence: 29.9071
Random effects:
Groups   Name          Std.Dev.
GelSTIM2 (Intercept) 0.000
Residual          0.317
Number of obs: 42, groups: GelSTIM2, 13
Fixed Effects:
              (Intercept)          TejidoTUMORAL          SexoMUJER TejidoTUMORAL:SexoMUJER
              -0.21003              0.05195              0.16343              0.30988
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings

#Modelo P_G
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM2 ~ Tejido * Sexo + (1 | Paciente) + (1 | GelSTIM2)
Data: datosr
REML criterion at convergence: 28.8539
Random effects:
Groups   Name          Std.Dev.
GelSTIM2 (Intercept) 0.0000
Paciente (Intercept) 0.1213
Residual          0.2980
Number of obs: 42, groups: GelSTIM2, 13; Paciente, 8
Fixed Effects:
              (Intercept)          TejidoTUMORAL          SexoMUJER TejidoTUMORAL:SexoMUJER
              -0.21003              0.06885              0.17378              0.26858
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
```

En consecuencia, tras descartar los modelos problemáticos, el modelo que presenta un menor valor de cAIC y no presenta ninguno de los problemas considerados es el *modelo Paciente* (tabla D.38).

Tabla D.38: Parte aleatoria del modelo seleccionado para STIM2. Se muestra la salida de R de la función lmer() usada para estimar el modelo lineal mixto y la función confint(level = 0.95) para los intervalos de confianza.

```
#Modelo Paciente
Linear mixed model fit by REML ['lmerMod']
Formula: logSTIM2 ~ Tejido * Sexo + (1 | Paciente)
Data: datosr
REML criterion at convergence: 28.8539
Random effects:
Groups   Name          Std.Dev.
Paciente (Intercept) 0.1213
Residual          0.2980
Number of obs: 42, groups: Paciente, 8
```

Posteriormente, mediante la AIC, se evalúa la parte fija que, al tratarse de un modelo de efectos fijos, es la única parte que contiene. Como resultado, en la tabla D.39 se muestra la selección de la parte fija que, como puede verse, resultan ser tanto los dos efectos principales y su interacción.

Tabla D.39: Selección de modelos para la variable respuesta T(STIM2) con AIC, BIC y LR.

| Parte Fija | AIC |
|------------------------------------|--------|
| Tejido + Sexo + Tejido:Sexo | 30.063 |
| Tejido + Sexo | 30.401 |
| Tejido | 32.866 |
| Sexo | 33.079 |
| nulo | 35.875 |

Así, tras evaluar el modelo sobreóptimo con el fin de encontrar tanto la parte fija como la parte aleatoria más adecuada, se obtuvo que el modelo más adecuado para modelizar la expresión de STIM2 es (ecuación D.5):

$$T \left(\frac{DO_{STIM2}}{DO_{\beta\text{-actina}}_{ijdr}} \right) = \beta_0 + \beta_1 \text{Tejido}_i + \beta_2 \text{Sexo}_j + \beta_3 \text{Tejido}_i \text{sexo}_j + b_{0,d} + \epsilon_{ijdr} \quad (D.5)$$

donde $\left(\frac{DO_{STIM2}}{DO_{\beta\text{-actina}}_{ijdr}} \right)$ es el valor de la variable respuesta transformada según la familia de Box-Cox propuesto por Hawkins and Weisberg (2017), medido en la réplica r de la muestra cuyo tejido es i (Tumoral o Normal) y el sexo j (Hombre o Mujer) medidas en el paciente d . En cuanto a los parámetros: β_0 es el intercepto o término independiente del modelo (referencia que, en este caso es el nivel Hombre.Normal); β_1 es el cambio de expresión en el tejido tumoral respecto del de referencia (Hombre.Normal), β_2 el cambio en la expresión de la proteína de interés en las muestras de mujer respecto de las de referencia (Hombre.Normal); β_3 el efecto de la interacción entre los

factores *Tejido* y *Sexo* que indica si el cambio entre el tejido tumoral y normal es diferente en hombres que en mujeres (en caso de ser significativo), y $b_{0,d}$ es el intercepto aleatorio debido al día que indica la desviación respecto de β_0 que hay en el paciente d .

En cuanto al diagnóstico del modelo o la evaluación del cumplimiento de los supuestos, en la tabla D.40 se muestran los métodos basados en contrastes de hipótesis y en la figura D.5 los métodos gráficos.

Tabla D.40: Supuestos del modelo seleccionado (p valores).

| | Normalidad | | | Homocedasticidad | |
|----------------------------|------------|-------|-------|------------------|-----------|
| | S-W | J-B | Lil | B-P ¹ | Bartlett |
| Residuos de Pearson | 0.464 | 0.595 | 0.233 | 0.555 | 0.0443964 |
| Paciente intercepto | 0.743 | 0.769 | 0.874 | | |

Abreviaturas:

S-W: Shapiro-Wilks; J-B: Jarque Bera; Lil: Lilliefors; B-P: Breush-Pagan

Notas:

¹ Se ha realizado mediante una modificación del test de .

² Se han ajustado los residuos a los valores ajustados y se a realizado el test. de Breush-Pagan.

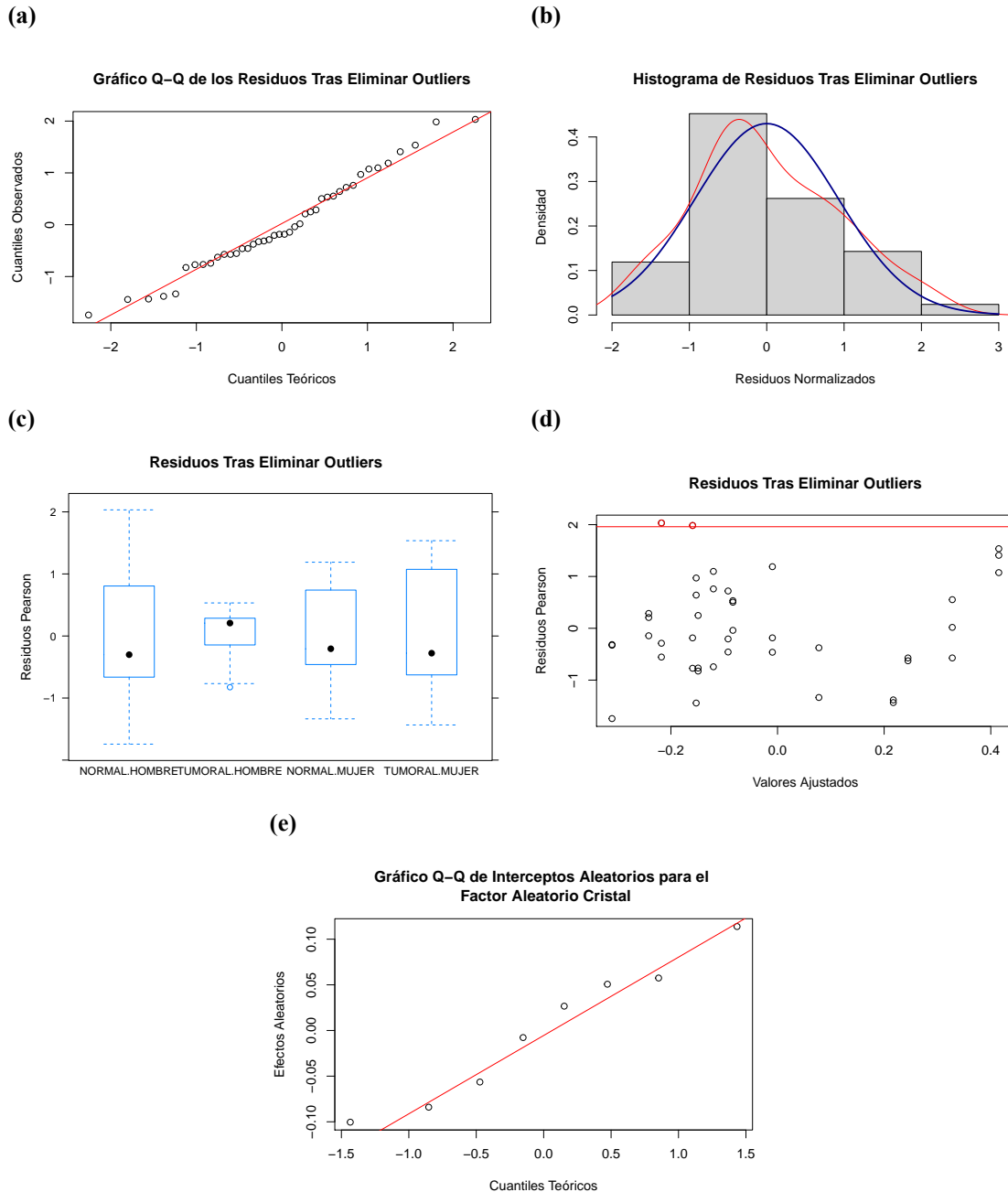


Figura D.5: Diagnóstico del modelo lineal de efectos fijos para la variable respuesta T (STIM2) tras eliminar los outliers. a) Gráfico Q-Q de los residuos de Pearson, donde la violación de la normalidad es tanto más cuanto mayor sea la desviación de los residuos respecto de la diagonal principal; b) Histograma de los residuos, en el que se ha superpuesto en rojo la estimación de la distribución de densidad y en azul la distribución teórica de una $N(0, 1)$; c) Gráfico de cajas y bigotes de los residuos de Pearson vs. condiciones experimentales, donde se esperaría ver una distribución similar en cada una de ellas; d) Gráfico de dispersión de los residuos de Pearson vs. los valores ajustados por el modelo, en los que se esperaría una distribución aleatoria e independiente del valor ajustado, a la par de unos valores absolutos de los residuos no superiores a 2 (señalados en rojo) o 3, pues, en caso contrario, serían considerados como outliers; e) Gráfico Q-Q de los efectos aleatorios (intercepto) del factor aleatorio Paciente.

En la tabla D.41 se muestran las estimaciones de los coeficientes del modelo y en la tabla D.42 su matriz de varianzas y covarianzas.

Tabla D.41: Estimaciones de los coeficientes y contraste de hipótesis $H_0 : \beta_i = 0$ del modelo final para T (STIM2).

| | Estimación | SEM | t | gdl | p-valor |
|--------------------------------|------------|-----------|------------|----------|-----------|
| β_0 | -0.2100348 | 0.1052497 | -1.9955863 | 11.07832 | 0.0711585 |
| β_{Tumoral} | 0.0688477 | 0.1344816 | 0.5119485 | 11.07832 | 0.6187379 |
| β_{Mujer} | 0.1737825 | 0.1512537 | 1.1489473 | 11.07832 | 0.2747798 |
| $\beta_{\text{Tumoral:Mujer}}$ | 0.2685807 | 0.1877318 | 1.4306619 | 11.07832 | 0.1801233 |

Abreviaturas:

gdl: grados de libertad

Tabla D.42: Covarianzas entre los coeficientes del modelo seleccionado para T ($DO_{\text{STIM2}}/DO_{\beta\text{-actina}}$).

| | β_0 | β_{Tumoral} | β_{Mujer} | $\beta_{\text{Tumoral:Mujer}}$ |
|--------------------------------|------------|--------------------------|------------------------|--------------------------------|
| β_0 | 0.0110775 | -0.0073999 | -0.0110775 | 0.0073999 |
| β_{Tumoral} | -0.0073999 | 0.0180853 | 0.0073999 | -0.0180853 |
| β_{Mujer} | -0.0110775 | 0.0073999 | 0.0228777 | -0.0155604 |
| $\beta_{\text{Tumoral:Mujer}}$ | 0.0073999 | -0.0180853 | -0.0155604 | 0.0352432 |

En cuanto a la estructura de la parte aleatoria, la varianza de los efectos aleatorios se muestra en la tabla en la tabla D.43.

Tabla D.43: Estimaciones de la varianza de la parte aleatoria del modelo mixto final para T (STIM2).

| Componente Aleatorio | LI | Estimación SEM | LS |
|----------------------|-----------|----------------|----------|
| Paciente Intercepto | 0.022807 | 0.1212865 | 0.223501 |
| Sigma | 0.2323555 | 0.2979911 | 0.374375 |

La tabla ANOVA del modelo seleccionado con interacción se muestra en la tabla D.44. En este caso, para un nivel de significación del 0.05, los efectos principales son significativos, pero no la interacción. No obstante, según AIC, la interacción ha de estar en el modelo.

Tabla D.44: Tabla ANOVA (tipo II) del modelo seleccionado del modelo seleccionado con interacción para T (STIM2).

| | Sum Sq | gdl | p-valor |
|--------------------|----------|-----|-----------|
| Tejido | 4.851177 | 1 | 0.0276273 |
| Sexo | 5.339808 | 1 | 0.0208438 |
| Tejido:Sexo | 2.046794 | 1 | 0.1525271 |

Abreviaturas:

gdl: grados de libertad

Las estimaciones de la expresión de STIM2, expresado como $DO_{STIM2}/DO_{\beta\text{-actina}}$, se muestran en la tabla D.45.

Tabla D.45: Estimaciones de la expresión de STIM2 ($DO_{STIM2}/DO_{\beta\text{-actina}}$) con el modelo lineal seleccionado.

| Condiciones | ratio DO | Límite Superior | Límite Inferior | log(Señal-SEM) | log(Señal+SEM) | n |
|-----------------------|-----------|-----------------|-----------------|----------------|----------------|----|
| Hombre.Normal | 0.6395361 | 0.8414151 | 0.4869476 | 0.5569782 | 0.7346501 | 12 |
| Hombre.Tumoral | 0.7001178 | 0.9635075 | 0.5138848 | 0.5981585 | 0.8215336 | 9 |
| Mujer.Normal | 0.8051392 | 1.0872198 | 0.6075451 | 0.6967158 | 0.9348716 | 11 |
| Mujer.Tumoral | 1.3218293 | 2.0070237 | 0.9460961 | 1.1071467 | 1.6111548 | 10 |

Las comparaciones múltiples se muestran en la tabla D.46.

Tabla D.46: Diferencias dos a dos entre las condiciones experimentales de sus valores T ($DO_{STIM2}/DO_{\beta\text{-actina}}$) estimados con el modelo seleccionado.

| | Diferencias | SEM | t | pvalor |
|-------------------------------------|-------------|-----------|-----------|-------------|
| Hombre.Tumoral-Hombre.Normal | 0.0688477 | 0.1344816 | 0.5119485 | 0.9554362 |
| Mujer.Normal-Hombre.Normal | 0.1737825 | 0.1512537 | 1.1489473 | 0.6548242 |
| Mujer.Tumoral-Hombre.Normal | 0.5112109 | 0.1539954 | 3.3196498 | 0.0049590 * |
| Mujer.Normal-Hombre.Tumoral | 0.1049349 | 0.1617504 | 0.6487456 | 0.9145146 |
| Mujer.Tumoral-Hombre.Tumoral | 0.4423633 | 0.1643171 | 2.6921313 | 0.0352105 * |
| Mujer.Tumoral-Mujer.Normal | 0.3374284 | 0.1309883 | 2.5760200 | 0.0480485 * |

Parte VIII

Bibliografía

Bibliografía

- Abeele, Fabien Vanden, Morad Roudbaraki, Yaroslav Shuba, Roman Skryma, and Natalia Prevarskaya. 2003. "Store-Operated Ca²⁺ Current in Prostate Cancer Epithelial Cells: Role of Endogenous Ca²⁺ Transporter Type 1." *Journal of Biological Chemistry* 278 (17): 15381–89.
- Adams, Jerry M, and Suzanne Cory. 2007. "The Bcl-2 Apoptotic Switch in Cancer Development and Therapy." *Oncogene* 26 (9): 1324–37.
- Ahmed, D, PW Eide, IA Eilertsen, SA Danielsen, M Eknaes, M Hektoen, GE Lind, and RA Lothe. 2013. "Epigenetic and Genetic Features of 24 Colon Cancer Cell Lines." *Oncogenesis* 2 (9): e71–71.
- Albarrán, Letizia, Jose J Lopez, Natalia Dionisio, Tarik Smani, Gines M Salido, and Juan A Rosado. 2013. "Transient Receptor Potential Ankyrin-1 (Trpa1) Modulates Store-Operated Ca²⁺ Entry by Regulation of Stim1-Orai1 Association." *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* 1833 (12): 3025–34.
- Alberts, Bruce, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter, John Wilson, and Tim Hunt. 2010. *Biología Molecular de La célula*. Ediciones Omega.
- Aldea, Marcos, Kisun Jun, Hee-Sup Shin, Eva Andrés-Mateos, Luisa M Solís-Garrido, Carmen Montiel, Antonio G García, and Almudena Albillos. 2002. "A Perforated Patch-Clamp Study of Calcium Currents and Exocytosis in Chromaffin Cells of Wild-Type and α 1a Knockout Mice." *Journal of Neurochemistry* 81 (5): 911–21.
- Alkhani, Hazim, Ariel R Ase, Rebecca Grant, Dajan O'Donnell, Klaus Groschner, and Philippe Séguéla. 2014. "Contribution of Trpc3 to Store-Operated Calcium Entry and Inflammatory Transductions in Primary Nociceptors." *Molecular Pain* 10: 1744–8069.
- Almasi, Shekoufeh, and Yassine El Hiani. 2020. "Exploring the Therapeutic Potential of Membrane Transport Proteins: Focus on Cancer and Chemoresistance." *Cancers* 12 (6): 1624.
- Amantini, Consuelo, M Mosca, M Nabissi, R Lucciarini, S Caprodossi, A Arcella, F Giangaspero, and G Santoni. 2007. "Capsaicin-Induced Apoptosis of Glioma Cells Is Mediated by Trpv1 Vanilloid Receptor and Requires P38 MAPK Activation." *Journal of Neurochemistry* 102 (3): 977–90.

-
- Amit, Sharon, Ada Hatzubai, Yaara Birman, Jens S Andersen, Etti Ben-Shushan, Matthias Mann, Yinon Ben-Neriah, and Irit Alkalay. 2002. "Axin-Mediated CKI Phosphorylation of β -Catenin at Ser 45: A Molecular Switch for the Wnt Pathway." *Genes & Development* 16 (9): 1066–76.
- Aplan, Peter D. 2006. "Causes of Oncogenic Chromosomal Translocation." *TRENDS in Genetics* 22 (1): 46–55.
- Arbajian, Atousa, Jean-Philippe Brouland, Ágota Apáti, Katalin Pászty, Luca Hegedűs, Ágnes Enyedi, Christine Chomienne, and Béla Papp. 2013. "Modulation of Endoplasmic Reticulum Calcium Pump Expression During Lung Cancer Cell Differentiation." *The FEBS Journal* 280 (21): 5408–18.
- Arbajian, Atousa, Mircea Iftinca, Christophe Altier, Param Priya Singh, Hervé Isambert, and Sylvie Coscoy. 2020. "Mutations in Calmodulin-Binding Domains of Trpv4/6 Channels Confer Invasive Properties to Colon Adenocarcinoma Cells." *Channels* 14 (1): 101–9.
- Aung, Cho S, Weilan Ye, Greg Plowman, Amelia A Peters, Gregory R Monteith, and Sarah J Roberts-Thomson. 2009. "Plasma Membrane Calcium ATPase 4 and the Remodeling of Calcium Homeostasis in Human Colon Cancer Cells." *Carcinogenesis* 30 (11): 1962–69.
- Aydar, Ebru, Syn Yeo, Mustafa Djamgoz, and Christopher Palmer. 2009. "Abnormal Expression, Localization and Interaction of Canonical Transient Receptor Potential Ion Channels in Human Breast Cancer Cell Lines and Tissues: A Potential Target for Breast Cancer Diagnosis and Therapy." *Cancer Cell International* 9 (1): 1–12.
- Aytes, Alvaro, David G Molleví, Maria Martinez-Iniesta, Marga Nadal, August Vidal, Albert Morales, Ramón Salazar, Gabriel Capellà, and Alberto Villanueva. 2012. "Stromal Interaction Molecule 2 (Stim2) Is Frequently Overexpressed in Colorectal Tumors and Confers a Tumor Cell Growth Suppressor Phenotype." *Molecular Carcinogenesis* 51 (9): 746–53.
- Baker, Mariah R, Guizhen Fan, and Irina I Serysheva. 2017. "Structure of Ip3r Channel: High-Resolution Insights from Cryo-EM." *Current Opinion in Structural Biology* 46: 38–47.
- Baraniak Jr, James H, Yandong Zhou, Robert M Nwonkonko, Michelle R Jennette, Sarah A Kazzaz, Jazmin M Stenson, Abigale L Whitsell, Youjun Wang, Mohamed Trebak, and Donald Gill. 2020. "The Orai Channel c-Terminal Peptide Defines the STIM-Orai Coupling Interface." *Available at SSRN 3707179*.
- Barr, Dale J, Roger Levy, Christoph Scheepers, and Harry J Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal." *Journal of Memory and Language* 68 (3): 255–78.
- Bass, Adam J, Michael S Lawrence, Lear E Brace, Alex H Ramos, Yotam Drier, Kristian Cibulskis, Carrie Sougnez, et al. 2011. "Genomic Sequencing of Colorectal Adenocarcinomas Identifies a Recurrent Vt1a-Tcf712 Fusion." *Nature Genetics* 43 (10): 964–68.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss>.

v067.i01.

- Benjamini, Yoav, and Yoel Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300.
- Ben-Kasus Nissim, Tsipi, Xuexin Zhang, Assaf Elazar, Soumitra Roy, Judith A Stolwijk, Yandong Zhou, Rajender K Motiani, et al. 2017. "Mitochondria Control Store-Operated Ca²⁺ Entry Through Na⁺ and Redox Signals." *The EMBO Journal* 36 (6): 797–815.
- Benzerdjeb, Nazim, Henri Sevestre, Ahmed Ahidouch, and Halima Ouadid-Ahidouch. 2016. "Orai3 Is a Predictive Marker of Metastasis and Survival in Resectable Lung Adenocarcinoma." *Oncotarget* 7 (49): 81588.
- Berger, C, Y Qian, and Xinbin Chen. 2013. "The P53-Estrogen Receptor Loop in Cancer." *Current Molecular Medicine* 13 (8): 1229–40.
- Bergsmann, Judith, Isabella Derler, Martin Muik, Irene Frischauf, Marc Fahrner, Philipp Pollheimer, Clemens Schwarzingler, Hermann J Gruber, Klaus Groschner, and Christoph Romanin. 2011. "Molecular Determinants Within n Terminus of Orai3 Protein That Control Channel Activation and Gating." *Journal of Biological Chemistry* 286 (36): 31565–75.
- Berlansky, Sascha, Christina Humer, Matthias Sallinger, and Irene Frischauf. 2021. "More Than Just Simple Interaction Between STIM and Orai Proteins: CRAC Channel Function Enabled by a Network of Interactions with Regulatory Proteins." *International Journal of Molecular Sciences* 22 (1): 471.
- Berna-Erro, Alejandro, Isaac Jardin, Gines M Salido, and Juan A Rosado. 2017. "Role of Stim2 in Cell Function and Physiopathology." *The Journal of Physiology* 595 (10): 3111–28.
- Berridge, Michael. 2014. "Ion Channels." *Cell Signalling Biology* 2014 (1): 1–74.
- Berridge, Michael J. 1995. "Capacitative Calcium Entry." *Biochemical Journal* 312 (Pt 1): 1.
- . 2012. "Calcium Signalling Remodelling and Disease." *Biochemical Society Transactions* 40 (2): 297–309.
- Berridge, Michael J, and Robin F Irvine. 1984. "Inositol Trisphosphate, a Novel Second Messenger in Cellular Signal Transduction." *Nature* 312 (5992): 315–21.
- Berridge, Michael J, Peter Lipp, and Martin D Bootman. 2000. "The Versatility and Universality of Calcium Signalling." *Nature Reviews Molecular Cell Biology* 1 (1): 11–21.
- Bhardwaj, Rajesh, Matthias A Hediger, and Nicolas Demaurex. 2016. "Redox Modulation of STIM-ORAI Signaling." *Cell Calcium* 60 (2): 142–52.
- Bhowmick, Neil A, Eric G Neilson, and Harold L Moses. 2004. "Stromal Fibroblasts in Cancer Initiation and Progression." *Nature* 432 (7015): 332–37.
- Bidaux, Gabriel, Matthieu Flourakis, Stéphanie Thebault, Alexander Zholos, Benjamin Beck, Dimitra Gkika, Morad Roudbaraki, et al. 2007. "Prostate Cell Differentiation Status Determines

-
- Transient Receptor Potential Melastatin Member 8 Channel Subcellular Localization and Function.” *The Journal of Clinical Investigation* 117 (6): 1647–57.
- Bidaux, Gabriel, Morad Roudbaraki, Corinne Merle, Alexandre Crépin, Philippe Delcourt, Christian Slomianny, Stéphanie Thébault, et al. 2005. “Evidence for Specific Trpm8 Expression in Human Prostate Secretory Epithelial Cells: Functional Androgen Receptor Requirement.” *Endocrine-Related Cancer* 12 (2): 367–82.
- Blackshaw, L Ashley, Amanda J Page, and Richard L Young. 2011. “Metabotropic Glutamate Receptors as Novel Therapeutic Targets on Visceral Sensory Pathways.” *Frontiers in Neuroscience* 5: 40.
- Bleich, Stefan, Konstanze Römer, Jens Wiltfang, and Johannes Kornhuber. 2003. “Glutamate and the Glutamate Receptor System: A Target for Drug Action.” *International Journal of Geriatric Psychiatry* 18 (S1): S33–40.
- Bode, Katrin, and Damien M O’Halloran. 2018. “NCX-DB: A Unified Resource for Integrative Analysis of the Sodium Calcium Exchanger Super-Family.” *BMC Neuroscience* 19 (1): 1–10.
- Bolstad, Ben. 2008. “Preprocessing and Normalization for Affymetrix GeneChip Expression Microarrays.” *Methods in Microarray Normalization* 41: 41–60.
- Bolstad, Benjamin M. 2004. “Low Level Analysis of High-Density Oligonucleotide Array Data: Background, Normalization and Summarization.” PhD thesis, University of California, Berkeley.
- Bolstad, Benjamin M, Francois Collin, Julia Brettschneider, Ken Simpson, Leslie Cope, Rafael A Irizarry, and Terence P Speed. 2005. “Quality Assessment of Affymetrix GeneChip Data.” In *Bioinformatics and Computational Biology Solutions Using r and Bioconductor*, edited by R. Gentleman, V. Carey, W. Huber, R Irizarry, and S Dudoit, 33–47. New York: Springer.
- Bomben, Valerie C, Kathryn L Turner, Tia-Tabitha C Barclay, and Harald Sontheimer. 2011. “Transient Receptor Potential Canonical Channels Are Essential for Chemotactic Migration of Human Malignant Gliomas.” *Journal of Cellular Physiology* 226 (7): 1879–88.
- Bomfim, Guilherme H Souza, Veronica Costiniti, Yi Li, Youssef Idaghdour, and Rodrigo S Lacruz. 2020. “Trpm7 Activation Potentiates SOCE in Enamel Cells but Requires ORAI.” *Cell Calcium* 87: 102187.
- Bonhenry, Daniel, Romana Schober, Tony Schmidt, Linda Waldherr, Rüdiger H Ettrich, and Rainer Schindl. 2019. “Mechanistic Insights into the Orai Channel by Molecular Dynamics Simulations.” In *Seminars in Cell & Developmental Biology*, 94:50–58. Elsevier.
- Bonora, Massimo, and Paolo Pinton. 2014. “The Mitochondrial Permeability Transition Pore and Cancer: Molecular Mechanisms Involved in Cell Death.” *Frontiers in Oncology* 4: 302.
- Borrelli, Francesca, Ester Pagano, Barbara Romano, Stefania Panzera, Francesco Maiello, Diana Coppola, Luciano De Petrocellis, Lorena Buono, Pierangelo Orlando, and Angelo A Izzo. 2014.

-
- “Colon Carcinogenesis Is Inhibited by the Trpm8 Antagonist Cannabigerol, a Cannabis-Derived Non-Psychotropic Cannabinoid.” *Carcinogenesis* 35 (12): 2787–97.
- Box, George EP, and David R Cox. 1964. “An Analysis of Transformations.” *Journal of the Royal Statistical Society: Series B (Methodological)* 26 (2): 211–43.
- Boyman, Liron, George SB Williams, Daniel Khananshvili, Israel Sekler, and WJ Lederer. 2013. “NCLX: The Mitochondrial Sodium Calcium Exchanger.” *Journal of Molecular and Cellular Cardiology* 59: 205–13.
- Brettschneider, Julia, Francois Collin, Benjamin M Bolstad, and Terence P Speed. 2007. “Quality Assessment for Short Oligonucleotide Arrays.” *Technometrics* In press.
- Bréchar, Sabrina, Chantal Melchior, Sébastien Plançon, Véronique Schenten, and EJ Tschirhart. 2008. “Store-Operated Ca²⁺ Channels Formed by Trpc1, Trpc6 and Orail and Non-Store-Operated Channels Formed by Trpc3 Are Involved in the Regulation of NADPH Oxidase in HL-60 Granulocytes.” *Cell Calcium* 44 (5): 492–506.
- Brini, Marisa, Tito Cali, Denis Ottolini, and Ernesto Carafoli. 2013. “Intracellular Calcium Homeostasis and Signaling.” In *Metallomics and the Cell*, 119–68. Springer.
- Brocardo, Mariana, Inke S Näthke, and Beric R Henderson. 2005. “Redefining the Subcellular Location and Transport of APC: New Insights Using a Panel of Antibodies.” *EMBO Reports* 6 (2): 184–90.
- Brouland, Jean-Philippe, Pascal Gélébart, Tünde Kovács, Jocelyne Enouf, Johannes Grossmann, and Béla Papp. 2005. “The Loss of Sarco/Endoplasmic Reticulum Calcium Transport ATPase 3 Expression Is an Early Event During the Multistep Process of Colon Carcinogenesis.” *The American Journal of Pathology* 167 (1): 233–42.
- Bubien, James K, Liang-Ji Zhou, P Darwin Bell, Raymond A Frizzell, and Thomas F Tedder. 1993. “Transfection of the Cd20 Cell Surface Molecule into Ectopic Cell Types Generates a Ca²⁺ Conductance Found Constitutively in b Lymphocytes.” *Journal of Cell Biology* 121 (5): 1121–32.
- Burn, John, Anne-Marie Gerdes, Finlay Macrae, Jukka-Pekka Mecklin, Gabriela Moeslein, Sylviane Olschwang, Diane Eccles, et al. 2011. “Long-Term Effect of Aspirin on Cancer Risk in Carriers of Hereditary Colorectal Cancer: An Analysis from the Capp2 Randomised Controlled Trial.” *The Lancet* 378 (9809): 2081–87.
- Butcher, David, Khatab Hassanein, Maureen Dudgeon, James Rhodes, and Frederick F Holmes. 1985. “Female Gender Is a Major Determinant of Changing Subsite Distribution of Colorectal Cancer with Age.” *Cancer* 56 (3): 714–16.
- Butorac, Carmen, Adéla Krizova, and Isabella Derler. 2020. “Structure and Activation Mechanisms of CRAC Channels.” *Calcium Signaling*, 547–604.
- Cai, Rong, Xia Ding, Kechun Zhou, Yu Shi, Ruiliang Ge, Gang Ren, Yening Jin, and Yizheng

-
- Wang. 2009. “Blockade of Trpc6 Channels Induced G2/m Phase Arrest and Suppressed Growth in Human Gastric Cancer Cells.” *International Journal of Cancer* 125 (10): 2281–87.
- Camara, Amadou KS, YiFan Zhou, Po-Chao Wen, Emad Tajkhorshid, and Wai-Meng Kwok. 2017. “Mitochondrial Vdac1: A Key Gatekeeper as Potential Therapeutic Target.” *Frontiers in Physiology* 8: 460.
- Capelle, Lisette G, Nicole CT Van Grieken, Hester F Lingsma, Ewout W Steyerberg, Willem J Klokman, Marco J Bruno, Hans FA Vasen, and Ernst J Kuipers. 2010. “Risk and Epidemiological Time Trends of Gastric Cancer in Lynch Syndrome Carriers in the Netherlands.” *Gastroenterology* 138 (2): 487–92.
- Carafoli, Ernesto. 2003. “The Calcium-Signalling Saga: Tap Water and Protein Crystals.” *Nature Reviews Molecular Cell Biology* 4 (4): 326–32.
- Carmeliet, Peter, and Rakesh K Jain. 2011a. “Molecular Mechanisms and Clinical Applications of Angiogenesis.” *Nature* 473 (7347): 298–307.
- . 2011b. “Principles and Mechanisms of Vessel Normalization for Cancer and Other Angiogenic Diseases.” *Nature Reviews Drug Discovery* 10 (6): 417–27.
- Carvalho, Benilton S, and Rafael A Irizarry. 2010. “A Framework for Oligonucleotide Microarray Preprocessing.” *Bioinformatics* 26 (19): 2363–67. <https://doi.org/10.1093/bioinformatics/btq431>.
- Carvalho, Benilton, and Robert Scharpf. 2020. *oligoClasses: Classes for High-Throughput Arrays Supported by Oligo and Crlmm*.
- Casero, Robert A, and Laurence J Marton. 2007. “Targeting Polyamine Metabolism and Function in Cancer and Other Hyperproliferative Diseases.” *Nature Reviews Drug Discovery* 6 (5): 373–90.
- Castiglioni, Sara, Alessandra Cazzaniga, Valentina Trapani, Concettina Cappadone, Giovanna Farruggia, Lucia Merolle, Federica I Wolf, Stefano Iotti, and Jeanette AM Maier. 2015. “Magnesium Homeostasis in Colon Carcinoma LoVo Cells Sensitive or Resistant to Doxorubicin.” *Scientific Reports* 5 (1): 1–9.
- Chalhoub, Nader, and Suzanne J Baker. 2009. “PTEN and the Pi3-Kinase Pathway in Cancer.” *Annual Review of Pathology: Mechanisms of Disease* 4: 127–50.
- Chalmers, Silke B, and Gregory R Monteith. 2018. “ORAI Channels and Cancer.” *Cell Calcium* 74: 160–67.
- Chang, David W, Zheng Xing, Yi Pan, Alicia Algeciras-Schimnich, Bryan C Barnhart, Shoshanit Yaish-Ohad, Marcus E Peter, and Xiaolu Yang. 2002. “C-FLIPL Is a Dual Function Regulator for Caspase-8 Activation and Cd95-Mediated Apoptosis.” *The EMBO Journal* 21 (14): 3704–14.
- Chang, Hsin-Han, Yu-Chen Cheng, Wen-Chiuan Tsai, Min-Jen Tsao, and Ying Chen. 2018. “Pyr3 Induces Apoptosis and Inhibits Migration in Human Glioblastoma Cells.” *Cellular Physiology and Biochemistry* 48 (4): 1694–1702.

-
- Chawla, Ajay, Joyce J Repa, Ronald M Evans, and David J Mangelsdorf. 2001. "Nuclear Receptors and Lipid Physiology: Opening the x-Files." *Science* 294 (5548): 1866–70.
- Chen, Liang, Rui Cao, Gang Wang, Lushun Yuan, Guofeng Qian, Zhongqiang Guo, Chin-Lee Wu, Xinghuan Wang, and Yu Xiao. 2017. "Downregulation of Trpm7 Suppressed Migration and Invasion by Regulating Epithelial–Mesenchymal Transition in Prostate Cancer Cells." *Medical Oncology* 34 (7): 1–11.
- Chen, Qi, Jian Kang, and Caiyun Fu. 2018. "The Independence of and Associations Among Apoptosis, Autophagy, and Necrosis." *Signal Transduction and Targeted Therapy* 3 (1): 1–11.
- Chen, Shu-jean, Lei Bao, Kerry Keefer, Santhanam Shanmughapriya, Longgui Chen, John Lee, JuFang Wang, et al. 2020. "Transient Receptor Potential Ion Channel Trpm2 Promotes AML Proliferation and Survival Through Modulation of Mitochondrial Function, ROS, and Autophagy." *Cell Death & Disease* 11 (4): 1–17.
- Chen, Yan, Junzhi Li, Lei Jin, Kai Lei, Haiwang Liu, and Yong Yang. 2019. "Fibulin-5 Contributes to Colorectal Cancer Cell Apoptosis via the ROS/MAPK and Akt Signal Pathways by Downregulating Transient Receptor Potential Cation Channel Subfamily v Member 1." *Journal of Cellular Biochemistry* 120 (10): 17838–46.
- Chen, Yih-Fung, Wen-Tai Chiu, Ying-Ting Chen, Pey-Yun Lin, Huey-Jy Huang, Cheng-Yang Chou, Hsien-Chang Chang, Ming-Jer Tang, and Meng-Ru Shen. 2011. "Calcium Store Sensor Stromal-Interaction Molecule 1-Dependent Signaling Plays an Important Role in Cervical Cancer Growth, Migration, and Angiogenesis." *Proceedings of the National Academy of Sciences* 108 (37): 15225–30.
- Chen, Zhen, Chunlei Tang, Yaodan Zhu, Mingxu Xie, Dongxu He, Qiongxi Pan, Peng Zhang, et al. 2017a. "TrpC5 Regulates Differentiation Through the Ca²⁺/Wnt5a Signalling Pathway in Colorectal Cancer." *Clinical Science* 131 (3): 227–37.
- Chen, Zhen, Yaodan Zhu, Yongfei Dong, Peng Zhang, Xiping Han, Jian Jin, and Xin Ma. 2017b. "Overexpression of TrpC5 Promotes Tumor Metastasis via the HIF-1 α –Twist Signaling Pathway in Colon Cancer." *Clinical Science* 131 (19): 2439–50.
- Cheng, Nikki, Anna Chytil, Yu Shyr, Alison Joly, and Harold L Moses. 2008. "Transforming Growth Factor- β Signaling–Deficient Fibroblasts Enhance Hepatocyte Growth Factor Signaling in Mammary Carcinoma Cells to Promote Scattering and Invasion." *Molecular Cancer Research* 6 (10): 1521–33.
- Chlebowski, Rowan T, Jean Wactawski-Wende, Cheryl Ritenbaugh, F Allan Hubbell, Joao Ascensao, Rebecca J Rodabough, Carol A Rosenberg, et al. 2004. "Estrogen Plus Progestin and Colorectal Cancer in Postmenopausal Women." *New England Journal of Medicine* 350 (10): 991–1004.
- Choi, Jung Kyoon, Ungsik Yu, Sangsoo Kim, and Ook Joon Yoo. 2003. "Combining Multiple Microarray Studies and Modeling Interstudy Variation." *Bioinformatics* 19 (suppl_1): i84–90.

-
- Chovancova, Barbora, Veronika Liskova, Petr Babula, and Olga Krizanova. 2020. "Role of Sodium/Calcium Exchangers in Tumors." *Biomolecules* 10 (9): 1257.
- Chung, Fu-Yen, Shiu-Ru Lin, Chien-Yu Lu, Ching-Sheng Yeh, Fang-Ming Chen, Jan-Sing Hsieh, Tsung-Jen Huang, and Jaw-Yuan Wang. 2006. "Sarco/Endoplasmic Reticulum Calcium-ATPase 2 Expression as a Tumor Marker in Colorectal Cancer." *The American Journal of Surgical Pathology* 30 (8): 969–74.
- Clevers, Hans, and Roel Nusse. 2012. "Wnt/ β -Catenin Signaling and Disease." *Cell* 149 (6): 1192–1205.
- Cojocaru, Florentina, Tudor Șelescu, Dan Domocoș, Luminița Măruțescu, Gabriela Chiritoiu, Nicoleta-Raluca Chelaru, Simona Dima, Dan Mihăilescu, Alexandru Babes, and Dana Cucu. 2021. "Functional Expression of the Transient Receptor Potential Ankyrin Type 1 Channel in Pancreatic Adenocarcinoma Cells." *Scientific Reports* 11 (1): 1–14.
- Colomer, Marián Carretero. 2005. "Cáncer Colorrectal Metastásico: Tratamiento Con Un Nuevo Agente Biológico (Cetuximab)." *Offarm: Farmacia y Sociedad* 24 (5): 114–16.
- Connaughton, Victoria. 2007. "Glutamate and Glutamate Receptors in the Vertebrate Retina." *Webvision: The Organization of the Retina and Visual System [Internet]*.
- Cosens, DJ, and Aubrey Manning. 1969. "Abnormal Electroretinogram from a Drosophila Mutant." *Nature* 224 (5216): 285–87.
- Cramer, William. 1918. "On the Biochemical Mechanism of Growth. The Effect of Sodium and Calcium Ions on the Growth of a Transplantable Mouse Carcinoma." *Biochemical Journal* 12 (3): 210.
- Curry, Merril C, Nicole A Luk, Paraic A Kenny, Sarah J Roberts-Thomson, and Gregory R Monteith. 2012. "Distinct Regulation of Cytoplasmic Calcium Signals and Cell Death Pathways by Different Plasma Membrane Calcium ATPase Isoforms in MDA-MB-231 Breast Cancer Cells." *Journal of Biological Chemistry* 287 (34): 28598–608.
- D'Errico, I, and A Moschetta. 2008. "Nuclear Receptors, Intestinal Architecture and Colon Cancer: An Intriguing Link." *Cellular and Molecular Life Sciences* 65 (10): 1523–43.
- Dadsena, Shashank, Svenja Bockelmann, John GM Mina, Dina G Hassan, Sergei Korneev, Guilherme Razzera, Helene Jahn, et al. 2019. "Ceramide Bind Vdac2 to Trigger Mitochondrial Apoptosis." *Nature Communications* 10 (1): 1–12.
- Dai, James Y, Bo Wang, Xiaoyu Wang, Anqi Cheng, Suzanne Kolb, Janet L Stanford, and Jonathan L Wright. 2019a. "Vigorous Physical Activity Is Associated with Lower Risk of Metastatic–Lethal Progression in Prostate Cancer and Hypomethylation in the Cracr2a Gene." *Cancer Epidemiology and Prevention Biomarkers* 28 (2): 258–64.
- . 2019b. "Vigorous Physical Activity Is Associated with Metastatic-Lethal Progression in Prostate Cancer and Differential Tumor DNA Methylation in the Cracr2a Gene." *Cancer*

-
- Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 28 (2): 258–64.
- Dang, Donna, and Rajini Rao. 2016. “Calcium-ATPases: Gene Disorders and Dysregulation in Cancer.” *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* 1863 (6): 1344–50.
- Daudén, E. 2007. “Farmacogenética II. Métodos Moleculares de Estudio, Bioinformática y Aspectos éticos.” *Actas Dermo-Sifiliográficas* 98 (1): 3–13.
- Davies, MA, and Y Samuels. 2010. “Analysis of the Genome to Personalize Therapy for Melanoma.” *Oncogene* 29 (41): 5545–55.
- Davis, Felicity M, Michelle T Parsonage, Peter J Cabot, Marie-Odile Parat, Erik W Thompson, Sarah J Roberts-Thomson, and Gregory R Monteith. 2013. “Assessment of Gene Expression of Intracellular Calcium Channels, Pumps and Exchangers with Epidermal Growth Factor-Induced Epithelial-Mesenchymal Transition in a Breast Cancer Cell Line.” *Cancer Cell International* 13 (1): 1–7.
- De Marchi, Umberto, Jaime Santo-Domingo, Cyril Castelbou, Israel Sekler, Andreas Wiederkehr, and Nicolas Demaurex. 2014. “NCLX Protein, but Not Letm1, Mediates Mitochondrial Ca²⁺ Extrusion, Thereby Limiting Ca²⁺-Induced NAD (p) h Production and Modulating Matrix Redox State.” *Journal of Biological Chemistry* 289 (29): 20377–85.
- Deb, Bipan K, and Gaiti Hasan. 2016. “Regulation of Store-Operated Ca²⁺ Entry by Septins.” *Frontiers in Cell and Developmental Biology* 4: 142.
- DeHaven, Wayne I, Bertina F Jones, John G Petranka, Jeremy T Smyth, Takuro Tomita, Gary S Bird, and James W Putney Jr. 2009. “TRPC Channels Function Independently of Stim1 and Orai1.” *The Journal of Physiology* 587 (10): 2275–98.
- Delierneux, Céline, Sana Kouba, Santhanam Shanmughapriya, Marie Potier-Cartereau, Mohamed Trebak, and Nadine Hempel. 2020. “Mitochondrial Calcium Regulation of Redox Signaling in Cancer.” *Cells* 9 (2): 432.
- Denko, Nicholas C. 2008. “Hypoxia, Hif1 and Glucose Metabolism in the Solid Tumour.” *Nature Reviews Cancer* 8 (9): 705–13.
- Derler, Isabella, Isaac Jardin, and Christoph Romanin. 2016. “Molecular Mechanisms of STIM/Orai Communication.” *American Journal of Physiology-Cell Physiology* 310 (8): C643–62.
- Derler, Isabella, Josef Madl, Gerhard Schütz, and Christoph Romanin. 2012. “Structure, Regulation and Biophysics of i CRAC, STIM/Orai1.” *Calcium Signaling*, 383–410.
- Dhennin-Duthille, Isabelle, Mathieu Gautier, Malika Faouzi, Arnaud Guilbert, Marie Brevet, David Vaudry, Ahmed Ahidouch, Henri Sevestre, and Halima Ouadid-Ahidouch. 2011. “High Expression of Transient Receptor Potential Channels in Human Breast Cancer Epithelial Cells and Tissues: Correlation with Pathological Parameters.” *Cellular Physiology and Biochemistry* 28 (5): 813–22.

-
- Ding, Xia, Zhuohao He, Kechun Zhou, Ju Cheng, Hailan Yao, Dongliang Lu, Rong Cai, et al. 2010. "Essential Role of Trpc6 Channels in G2/m Phase Transition and Development of Human Glioma." *JNCI: Journal of the National Cancer Institute* 102 (14): 1052–68.
- Dixon, L Beth, Helena F Balder, Mikko J Virtanen, Bahram Rashidkhani, Satu Männistö, Vittorio Krogh, Piet A van Den Brandt, et al. 2004. "Dietary Patterns Associated with Colon and Rectal Cancer: Results from the Dietary Patterns and Cancer (DIETSCAN) Project." *The American Journal of Clinical Nutrition* 80 (4): 1003–11.
- Dong, Hao, Yiming Zhang, Ruiheng Song, Jingjie Xu, Yigao Yuan, Jindou Liu, Jia Li, et al. 2019. "Toward a Model for Activation of Orai Channel." *Iscience* 16: 356–67.
- Dong, Hui, Ki-Nam Shim, Jenny MJ Li, Christine Estrema, Tiffany A Ornelas, Flang Nguyen, Shanglei Liu, et al. 2010. "Molecular Mechanisms Underlying Ca²⁺-Mediated Motility of Human Pancreatic Duct Cells." *American Journal of Physiology-Cell Physiology* 299 (6): C1493–503.
- Dong, Xian-Ping, Xiang Wang, and Haoxing Xu. 2010. "TRP Channels of Intracellular Membranes." *Journal of Neurochemistry* 113 (2): 313–28.
- Dubois, Charlotte, Fabien Vanden Abeele, V'yacheslav Lehen'kyi, Dimitra Gkika, Basma Guarmit, Gilbert Lepage, Christian Slomianny, et al. 2014. "Remodeling of Channel-Forming ORAI Proteins Determines an Oncogenic Switch in Prostate Cancer." *Cancer Cell* 26 (1): 19–32.
- Duitama, Milena, Viviana Vargas-López, Zulma Casas, Sonia L Albarracin, Jhon-Jairo Sutachan, and Yolima P Torres. 2020. "TRP Channels Role in Pain Associated with Neurodegenerative Diseases." *Frontiers in Neuroscience* 14.
- Dupont, Geneviève, Martin Falcke, Vivien Kirk, and James Sneyd. 2016. *Models of Calcium Signalling*. Vol. 43. Springer.
- Elbaz, Mohamad, Dinesh Ahirwar, Zhang Xiaoli, Xinyu Zhou, Maryam Lustberg, Mohd W Nasser, Konstantin Shilo, and Ramesh K Ganju. 2018. "Trpv2 Is a Novel Biomarker and Therapeutic Target in Triple Negative Breast Cancer." *Oncotarget* 9 (71): 33459.
- Elmets, Craig A, and Mohammad Athar. 2010. "Targeting Ornithine Decarboxylase for the Prevention of Nonmelanoma Skin Cancer in Humans." *Cancer Prevention Research* 3 (1): 8–11.
- Elzamazzy, Osama M, Reinhold Penner, and Lori A Hazlehurst. 2020. "The Role of Trpc1 in Modulating Cancer Progression." *Cells* 9 (2): 388.
- Endo, Yosuke, Katsuhiko Uzawa, Yoshiyuki Mochida, Masashi Shiiba, Hiroki Bukawa, Hidetaka Yokoe, and Hideki Tanzawa. 2004. "Sarcoendoplasmic Reticulum Ca²⁺ ATPase Type 2 Downregulated in Human Oral Squamous Cell Carcinoma." *International Journal of Cancer* 110 (2): 225–31.
- Erdmann, Kati, Knut Kaulke, Cathleen Thomae, Doreen Huebner, Mildred Sergon, Michael Froehner, Manfred P Wirth, and Susanne Fuessel. 2014. "Elevated Expression of Prostate

-
- Cancer-Associated Genes Is Linked to down-Regulation of microRNAs.” *BMC Cancer* 14 (1): 1–14.
- Ernst J. Kuipers, David Lieberman, William M. Grady. 2015. “Colorectal Cancer.” *Nature Reviews Disease Primers*, no. 15065.
- Fahrner, Marc, Saurabh K Pandey, Martin Muik, Lukas Traxler, Carmen Butorac, Michael Stadlbauer, Vasilina Zayats, et al. 2018. “Communication Between n Terminus and Loop2 Tunes Orai Activation.” *Journal of Biological Chemistry* 293 (4): 1271–85.
- Fahrner, Marc, Rainer Schindl, Martin Muik, Isabella Derler, and Christoph Romanin. 2017. “The STIM-Orai Pathway: The Interactions Between STIM and Orai.” In *Store-Operated Ca⁺ Entry (SOCE) Pathways: Emerging Signaling Concepts in Human (Patho)physiology*, edited by Klaus Groschner, Wolfgang F. Graier, and Christoph Romanin, 59–81. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57732-6_4.
- Fan, Lu, Ang Li, Wanshuai Li, Peifen Cai, Baofang Yang, Minxia Zhang, Yanhong Gu, et al. 2014. “Novel Role of Sarco/Endoplasmic Reticulum Calcium ATPase 2 in Development of Colorectal Cancer and Its Regulation by F36, a Curcumin Analog.” *Biomedicine & Pharmacotherapy* 68 (8): 1141–48.
- Fang, Diana, and Eduardo N Maldonado. 2018. “VDAC Regulation: A Mitochondrial Target to Stop Cell Proliferation.” *Advances in Cancer Research* 138: 41–69.
- Faouzi, Malika, Frederic Hague, Marie Potier, Ahmed Ahidouch, Henri Sevestre, and Halima Ouadid-Ahidouch. 2011. “Down-Regulation of Orai3 Arrests Cell-Cycle Progression and Induces Apoptosis in Breast Cancer Cells but Not in Normal Breast Epithelial Cells.” *Journal of Cellular Physiology* 226 (2): 542–51.
- Faouzi, Malika, Philippe Kischel, Frédéric Hague, Ahmed Ahidouch, Nazim Benzerdjeb, Henri Sevestre, Reinhold Penner, and Halima Ouadid-Ahidouch. 2013. “Orai3 Silencing Alters Cell Proliferation and Cell Cycle Progression via c-Myc Pathway in Breast Cancer Cells.” *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* 1833 (3): 752–60.
- Faraway, Julian J. 2004. *Linear Models with r*. Chapman; Hall/CRC.
- . 2016. *Extending the Linear Model with r: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman; Hall/CRC.
- Feng, Mingye, Desma M Grice, Helen M Faddy, Nguyen Nguyen, Sharon Leitch, Yingyu Wang, Sabina Muend, et al. 2010. “Store-Independent Activation of Orai1 by Spca2 in Mammary Tumors.” *Cell* 143 (1): 84–98.
- Feng, Ming-Ye, and Rajini Rao. 2013. “New Insights into Store-Independent Ca²⁺ Entry: Secretory Pathway Calcium ATPase 2 in Normal Physiology and Cancer.” *International Journal of Oral Science* 5 (2): 71–74.
- Fenwick, Elizabeth M, A Marty, and E Neher. 1982. “A Patch-Clamp Study of Bovine Chromaffin Cells and of Their Sensitivity to Acetylcholine.” *The Journal of Physiology* 331 (1): 577–97.

-
- Ferreira, Leonardo MR. 2010. "Cancer Metabolism: The Warburg Effect Today." *Experimental and Molecular Pathology* 89 (3): 372–80.
- Feske, Stefan, Yousang Gwack, Murali Prakriya, Sonal Srikanth, Sven-Holger Puppel, Bogdan Tanasa, Patrick G Hogan, Richard S Lewis, Mark Daly, and Anjana Rao. 2006. "A Mutation in Orai1 Causes Immune Deficiency by Abrogating CRAC Channel Function." *Nature* 441 (7090): 179–85.
- Fioro Pla, Alessandra, and Dimitra Gkika. 2013. "Emerging Role of TRP Channels in Cell Migration: From Tumor Vascularization to Metastasis." *Frontiers in Physiology* 4: 311.
- FitzMaurice Garrett, M, M Laird Nan, and H Ware James. 2004. "Applied Longitudinal Analysis." New Jersey: John Wiley & Sons.
- Flourakis, Matthieu, V Lehen'kyi, Benjamin Beck, Maylis Raphael, Matthieu Vandenberghe, F Vanden Abeele, Morad Roudbaraki, et al. 2010. "Orai1 Contributes to the Establishment of an Apoptosis-Resistant Phenotype in Prostate Cancer Cells." *Cell Death & Disease* 1 (9): e75–75.
- Fogh, J, and G Trempe. 1975. "In Fogh j.(ed) Human Tumor Cells in Vitro." *New York Plenum Press* 1 (9): 7.
- Foskett, J Kevin, and Benjamin Philipson. 2015. "The Mitochondrial Ca²⁺ Uniporter Complex." *Journal of Molecular and Cellular Cardiology* 78: 3–8.
- Fouillioux, F AND Rivera, C AND Contreras. 2004. "Receptores de glutamato: Implicaciones terapé." *Archivos Venezolanos de Farmacología y Terapé* 23 (July): 99–108. http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0798-02642004000200002&nrm=iso.
- Fourbon, Yann, Maxime Guéguinou, Romain Félix, Bruno Constantin, Arnaud Uguen, Gaëlle Fromont, Laurie Lajoie, et al. 2017. "Ca²⁺ Protein Alpha 1d of CaV1.3 Regulates Intracellular Calcium Concentration and Migration of Colon Cancer Cells Through a Non-Canonical Activity." *Scientific Reports* 7 (1): 1–13.
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Frischauf, Irene, Monika Litviňuková, Romana Schober, Vasilina Zayats, Barbora Svobodová, Daniel Bonhenry, Victoria Lunz, et al. 2017. "Transmembrane Helix Connectivity in Orai1 Controls Two Gates for Calcium-Dependent Transcription." *Science Signaling* 10 (507).
- Fuente, Sergio de la, Rosalba I Fonteriz, Mayte Montero, and Javier Alvarez. 2013. "Ca²⁺ Homeostasis in the Endoplasmic Reticulum Measured with a New Low-Ca²⁺-Affinity Targeted Aequorin." *Cell Calcium* 54 (1): 37–45.
- Fuertes, L, C Santonja, H Kutzner, and L Requena. 2013. "Inmunohistoquímica En Dermatopatología: Revisión de Los Anticuerpos Utilizados Con Mayor Frecuencia (Parte i)." *Actas Dermo-Sifiliográficas* 104 (2): 99–127.
- Fusi, Camilla, Serena Materazzi, Daiana Minocci, Vincenza Maio, Teresa Oranges, Daniela Massi, and Romina Nassini. 2014. "Transient Receptor Potential Vanilloid 4 (Trpv4) Is Downregulated

-
- in Keratinocytes in Human Non-Melanoma Skin Cancer.” *Journal of Investigative Dermatology* 134 (9): 2408–17.
- Gackiere, Florian, Gabriel Bidaux, Philippe Delcourt, Fabien Van Coppenolle, Maria Katsogiannou, Etienne Dewailly, Alexis Bavencoffe, et al. 2008. “CaV3. 2 t-Type Calcium Channels Are Involved in Calcium-Dependent Secretion of Neuroendocrine Prostate Cancer Cells.” *Journal of Biological Chemistry* 283 (15): 10162–73.
- Galeano-Otero, Isabel, Raquel Del Toro, Abdel-Majid Khatib, Juan Antonio Rosado, Antonio Ordóñez-Fernández, and Tarik Smani. 2021. “SARAF and Orail Contribute to Endothelial Cell Activation and Angiogenesis.” *Frontiers in Cell and Developmental Biology* 9: 300.
- Gao, Nannan, Feng Yang, Siyuan Chen, Hanxing Wan, Xiaoyan Zhao, and Hui Dong. 2020. “The Role of Trpv1 Ion Channels in the Suppression of Gastric Cancer Development.” *Journal of Experimental & Clinical Cancer Research* 39 (1): 1–17.
- Gao, Sheng-Lin, Chui-Ze Kong, Zhe Zhang, Ze-Liang Li, Jian-Bin Bi, and Xian-Kui Liu. 2017. “Trpm7 Is Overexpressed in Bladder Cancer and Promotes Proliferation, Migration, Invasion and Tumor Growth.” *Oncology Reports* 38 (4): 1967–76.
- Gautier, Laurent, Leslie Cope, Benjamin M. Bolstad, and Rafael A. Irizarry. 2004. “Affy—Analysis of Affymetrix GeneChip Data at the Probe Level.” *Bioinformatics* 20 (3): 307–15. <https://doi.org/10.1093/bioinformatics/btg405>.
- Geistlinger, Ludwig, Gergely Csaba, and Ralf Zimmer. 2016. “Bioconductor’s EnrichmentBrowser: Seamless Navigation Through Combined Results of Set- & Network-Based Enrichment Analysis.” *BMC Bioinformatics* 17: 45. <https://doi.org/10.1186/s12859-016-0884-1>.
- Gentleman, R., V. Carey, W. Huber, and F. Hahne. 2021. *Genefilter: Genefilter: Methods for Filtering Genes from High-Throughput Experiments*.
- Gerner, Eugene W, Natalia A Ignatenko, and David G Besselsen. 2003. “Preclinical Models for Chemoprevention of Colon Cancer.” *Tumor Prevention and Genetics*, 58–71.
- Gélébart, Pascal, Tünde Kovács, Jean-Philippe Brouland, Roosje van Gorp, Johannes Grossmann, Nathalie Rivard, Yves Panis, et al. 2002. “Expression of Endomembrane Calcium Pumps in Colon and Gastric Cancer Cells: Induction of Serca3 Expression During Differentiation.” *Journal of Biological Chemistry* 277 (29): 26310–20.
- Giardiello, Francis M, Stanley R Hamilton, Linda M Hyland, Vincent W Yang, Pamela Tamez, and Robert A Casero. 1997. “Ornithine Decarboxylase and Polyamines in Familial Adenomatous Polyposis.” *Cancer Research* 57 (2): 199–201.
- Giardiello, Francis M, Stanley R Hamilton, Anne J Krush, Steven Piantadosi, Linda M Hyland, Paul Celano, Susan V Booker, C Rahj Robinson, and G Johan A Offerhaus. 1993. “Treatment of Colonic and Rectal Adenomas with Sulindac in Familial Adenomatous Polyposis.” *New England Journal of Medicine* 328 (18): 1313–16.

-
- Gil, R, C López, and A Pellin. 1989. “Cultivo in Vitro de células Procedentes de Tumores Mamarios: Posibilidades y Problemas Metodológicos a La Luz de Nuestra Experiencia.” *Revista de Senología y Patología Mamaria* 2 (3): 161–67.
- Gill, AF. 1935. “The Relation of Calcium and Magnesium to Cancer.” *Canadian Medical Association Journal* 33 (1): 98.
- Giorgi, Carlotta, Federica Baldassari, Angela Bononi, Massimo Bonora, Elena De Marchi, Saverio Marchi, Sonia Missiroli, et al. 2012. “Mitochondrial Ca²⁺ and Apoptosis.” *Cell Calcium* 52 (1): 36–43.
- Giorgi, Carlotta, Massimo Bonora, Giovanni Sorrentino, Sonia Missiroli, Federica Poletti, Jan M Suski, Fabian Galindo Ramirez, et al. 2015. “P53 at the Endoplasmic Reticulum Regulates Apoptosis in a Ca²⁺-Dependent Manner.” *Proceedings of the National Academy of Sciences* 112 (6): 1779–84.
- Goeman, Jelle J, and Peter Bühlmann. 2007. “Analyzing Gene Expression Data in Terms of Gene Sets: Methodological Issues.” *Bioinformatics* 23 (8): 980–87.
- Goodwani, Sunil, Hannah Saternos, Fawaz Alasmari, and Youssef Sari. 2017. “Metabotropic and Ionotropic Glutamate Receptors as Potential Targets for the Treatment of Alcohol Use Disorder.” *Neuroscience & Biobehavioral Reviews* 77: 14–31.
- Gopal Krishnan, Priya D, Emily Golden, Eleanor A Woodward, Nathan J Pavlos, and Pilar Blancafort. 2020. “Rab GTPases: Emerging Oncogenes and Tumor Suppressive Regulators for the Editing of Survival Pathways in Cancer.” *Cancers* 12 (2): 259.
- Gou, Wen-Feng, Zhe-Feng Niu, Shuang Zhao, Yasuo Takano, and Hua-Chuan Zheng. 2014. “Aberrant Serca3 Expression During the Colorectal Adenoma-Adenocarcinoma Sequence.” *Oncology Reports* 31 (1): 232–40.
- Grabmayr, Herwig, Christoph Romanin, and Marc Fahrner. 2021. “STIM Proteins: An Ever-Expanding Family.” *International Journal of Molecular Sciences* 22 (1): 378.
- Grice, Desma M, Irina Vetter, Helen M Faddy, Paraic A Kenny, Sarah J Roberts-Thomson, and Gregory R Monteith. 2010. “Golgi Calcium Pump Secretory Pathway Calcium ATPase 1 (Spca1) Is a Key Regulator of Insulin-Like Growth Factor Receptor (Igf1r) Processing in the Basal-Like Breast Cancer Cell Line MDA-MB-231.” *Journal of Biological Chemistry* 285 (48): 37458–66.
- Grynkiwicz, Grzegorz, Martin Poenie, and Roger Y Tsien. 1985. “A New Generation of Ca²⁺ Indicators with Greatly Improved Fluorescence Properties.” *Journal of Biological Chemistry* 260 (6): 3440–50.
- Gu, Peng, Yi-Bin Zhou, Dong-Rong Yang, Yu-Xi Shan, and Bo-Xin Xue. 2014. “Inhibition of Stromal Interaction Molecule 1 and the Expression of Apoptosis-Related Proteins in Prostate Cancer PC-3 Cells.” *Zhonghua Nan Ke Xue= National Journal of Andrology* 20 (3): 225–28.

-
- Guilbert, Arnaud, Isabelle Dhennin-Duthille, Yassine EL Hiani, Nathalie Haren, Hafida Khorsi, Henri Sevestre, Ahmed Ahidouch, and Halima Ouadid-Ahidouch. 2008. "Expression of Trpc6 Channels in Human Epithelial Breast Cancer Cells." *BMC Cancer* 8 (1): 1–11.
- Guía de Usuario de Microarray Affymetrix (EXPRESION)*. 2021. Unidad genómica CABIMER.
- Gutiérrez, Lucía G, Miriam Hernández-Morales, Lucía Núñez, and Carlos Villalobos. 2019. "Inhibition of Polyamine Biosynthesis Reverses Ca²⁺ Channel Remodeling in Colon Cancer Cells." *Cancers* 11 (1): 83.
- Haben Fesseha, MVSc. 2020. "Principles and Applications of Deoxyribonucleic Acid Microarray: A Review."
- Han, Jing, Xue Zhang, Yan Liu, Li Jing, Yi-bing Liu, and Li Feng. 2020. "Clca4 and Ms4a12 as the Significant Gene Biomarkers of Primary Colorectal Cancer." *Bioscience Reports* 40 (8): BSR20200963.
- Han, Jung Woo, Woon Heo, Donghyuk Lee, Choeun Kang, Hye-Yeon Kim, Ikhyun Jun, Insuk So, et al. 2021. "Plasma Membrane Localized GCaMP-Ms4a12 by Orai1 Co-Expression Shows Thapsigargin-and Ca²⁺-Dependent Fluorescence Increases." *Molecules and Cells* 44 (4): 223–32.
- Hanahan, Douglas, and Robert A Weinberg. 2000. "The Hallmarks of Cancer." *Cell* 100 (1): 57–70.
- . 2011. "Hallmarks of Cancer: The Next Generation." *Cell* 144 (5): 646–74.
- Hanlon, MR, and BA Wallace. 2002. "Structure and Function of Voltage-Dependent Ion Channel Regulatory β Subunits." *Biochemistry* 41 (9): 2886–94.
- Hardie, Roger C. 2011. "A Brief History of Trp: Commentary and Personal Perspective." *Pflügers Archiv-European Journal of Physiology* 461 (5): 493–98.
- Hart, Matthew J, Rico De Los Santos, Iris N Albert, Bonnee Rubinfeld, and Paul Polakis. 1998. "Downregulation of β -Catenin by Human Axin and Its Association with the APC Tumor Suppressor, β -Catenin and Gsk3 β ." *Current Biology* 8 (10): 573–81.
- Hasan, Raquibul, and Xuming Zhang. 2018. "Ca²⁺ Regulation of TRP Ion Channels." *International Journal of Molecular Sciences* 19 (4): 1256.
- Hasna, Jessy, Nazim Benzerdjeb, Malika Faouzi, Anne-Sophie Ay, Philippe Kischel, Frédéric Hague, Henri Sevestre, Ahmed Ahidouch, and Halima Ouadid-Ahidouch. 2015. "Orai3 (ORAI Calcium Release-Activated Calcium Modulator 3)." *Atlas of Genetics and Cytogenetics in Oncology and Haematology*.
- Hasna, Jessy, Frédéric Hague, Lise Rodat-Despoix, Dirk Geerts, Catherine Leroy, David Tulasne, Halima Ouadid-Ahidouch, and Philippe Kischel. 2018. "Orai3 Calcium Channel and Resistance to Chemotherapy in Breast Cancer Cells: The P53 Connection." *Cell Death & Differentiation* 25 (4): 693–707.

-
- Hawkins, DM, and S Weisberg. 2017. "Combining the Box-Cox Power and Generalised Log Transformations to Accommodate Nonpositive Responses in Linear and Mixed-Effects Linear Models." *South African Statistical Journal* 51 (2): 317–28.
- Hayes, John D, Albena T Dinkova-Kostova, and Kenneth D Tew. 2020. "Oxidative Stress in Cancer." *Cancer Cell*.
- He, Benfu, Feiye Liu, Jian Ruan, Aimin Li, Jinzhang Chen, Rong Li, Jie Shen, Dayong Zheng, and Rongcheng Luo. 2012. "Silencing Trpc1 Expression Inhibits Invasion of Cne2 Nasopharyngeal Tumor Cells." *Oncology Reports* 27 (5): 1548–54.
- He, L, HY Deng, and others. 2017. "Decreased Expression of Ms4a12 Inhibits Differentiation and Predicts Early Stage Survival in Colon Cancer." *Neoplasia* 64 (1): 65–73.
- He, Wenfang, and Zhiping Hu. 2012. "The Role of the Golgi-Resident SPCA Ca²⁺/Mn²⁺ Pump in Ionic Homeostasis and Neural Function." *Neurochemical Research* 37 (3): 455–68.
- Hedgepeth, SC, MI Garcia, and LE Wagner. 2015. "II, Rodriguez AM, Chintapalli SV, Snyder RR, Et Al. The Brca1 Tumor Suppressor Binds to Inositol 1, 4, 5-Trisphosphate Receptors to Stimulate Apoptotic Calcium Release." *J Biol Chem* 290 (11): 7304–13.
- Hedges, Larry V. 1981. "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators." *Journal of Educational Statistics* 6 (2): 107–28.
- Hegedűs, Luca, Boglárka Zámbo, Katalin Pászty, Rita Padányi, Karolina Varga, John T Penniston, and Ágnes Enyedi. 2020. "Molecular Diversity of Plasma Membrane Ca²⁺ Transporting ATPases: Their Function Under Normal and Pathological Conditions." *Calcium Signaling*, 93–129.
- Heighway, Jim, Daniel C Betticher, Paul R Hoban, Hans J Altermatt, and Rachel Cowen. 1996. "Coamplification in Tumors of KRAS2, Type 2 Inositol 1, 4, 5 Triphosphate Receptor Gene, and a Novel Human Gene, KRAG." *Genomics* 35 (1): 207–14.
- Hernández-Morales, Miriam, Diego Sobradillo, Ruth A Valero, Eva Muñoz, Daniel Ubierna, Mary P Moyer, Lucía Núñez, and Carlos Villalobos. 2017. "Mitochondria Sustain Store-Operated Currents in Colon Cancer Cells but Not in Normal Colonic Cells: Reversal by Non-Steroidal Anti-Inflammatory Drugs." *Oncotarget* 8 (33): 55332.
- Hébert-Croteau, Nicole. 1998. "A Meta-Analysis of Hormone Replacement Therapy and Colon Cancer in Women." *Cancer Epidemiology and Prevention Biomarkers* 7 (8): 653–59.
- Hollenbeck, Peter J, and William M Saxton. 2005. "The Axonal Transport of Mitochondria." *Journal of Cell Science* 118 (23): 5411–19.
- Holzmann, Christian, Sven Kappel, Tatiana Kilch, Marcus Martin Jochum, Sabine Katharina Urban, Volker Jung, Michael Stöckle, Karen Rother, Markus Greiner, and Christine Peinelt. 2015. "Transient Receptor Potential Melastatin 4 Channel Contributes to Migration of Androgen-Insensitive Prostate Cancer Cells." *Oncotarget* 6 (39): 41783.

-
- Hooper, Robert, Elsie Samakai, Joseph Kedra, and Jonathan Soboloff. 2013. “Multifaceted Roles of STIM Proteins.” *Pflügers Archiv-European Journal of Physiology* 465 (10): 1383–96.
- Hoth, Markus, and Reinhold Penner. 1992. “Depletion of Intracellular Calcium Stores Activates a Calcium Current in Mast Cells.” *Nature* 355 (6358): 353–56.
- Hothorn, Torsten, Frank Bretz, and Peter Westfall. 2008. “Simultaneous Inference in General Parametric Models.” *Biometrical Journal* 50 (3): 346–63.
- Hou, Nengyi, Xuelai He, Yuhui Yang, Junwen Fu, Wei Zhang, Zhiyi Guo, Yang Hu, et al. 2019. “Trpv1 Induced Apoptosis of Colorectal Cancer Cells by Activating Calcineurin-Nfat2-P53 Signaling Pathway.” *BioMed Research International* 2019.
- Hou, Xiaowei, Shana R Burstein, and Stephen Barstow Long. 2018. “Structures Reveal Opening of the Store-Operated Calcium Channel Orai.” *Elife* 7: e36758.
- Hsu, Wen-Li, Ming-Hsien Tsai, Ching-Ying Wu, Jui-Lin Liang, Jian-He Lu, Jennifer S Kahle, Hsin-Su Yu, et al. 2020. “Nociceptive Transient Receptor Potential Canonical 7 (Trpc7) Mediates Aging-Associated Tumorigenesis Induced by Ultraviolet b.” *Aging Cell* 19 (1): e13075.
- Hu, Jijia, Kunhua Qin, Yan Zhang, Junbo Gong, Na Li, Dan Lv, Rong Xiang, and Xiaoyue Tan. 2011. “Downregulation of Transcription Factor Oct4 Induces an Epithelial-to-Mesenchymal Transition via Enhancement of Ca²⁺ Influx in Breast Cancer Cells.” *Biochemical and Biophysical Research Communications* 411 (4): 786–91.
- Huang HK, Chang HA, Lin YH. 2020. “Chemoresistant Ovarian Cancer Enhances Its Migration Abilities by Increasing Store-Operated Ca²⁺ Entry-Mediated Turnover of Focal Adhesions.” *Journal of Biomedical Science* 27 (1): 36.
- Huang, Peng, Mengnan Xu, Yi Wu, Alia Kazim Rizvi Syeda, and Xian-Ping Dong. 2020. “Multiple Facets of Trpm11 in Autophagy.” *Cell Calcium* 88: 102196.
- Hubert, Mia, Peter J Rousseeuw, and Tim Verdonck. 2012. “A Deterministic Algorithm for Robust Location and Scatter.” *Journal of Computational and Graphical Statistics* 21 (3): 618–37.
- Hwang, Sung-Min, Ji Yeon Lee, Chul-Kyu Park, and Yong Ho Kim. 2021. “The Role of TRP Channels and PMCA in Brain Disorders: Intracellular Calcium and pH Homeostasis.” *Frontiers in Cell and Developmental Biology* 9: 57.
- Ibrahim, Sajida, Hassan Dakik, Christophe Vandier, Romain Chautard, Gilles Paintaud, Frédéric Mazurier, Thierry Lecomte, Maxime Guéguinou, and William Raoul. 2019. “Expression Profiling of Calcium Channels and Calcium-Activated Potassium Channels in Colorectal Cancer.” *Cancers* 11 (4): 561.
- Ignatenko, Natalia A, Naveen Babbar, Dipti Mehta, Robert A Casero Jr, and Eugene W Gerner. 2004. “Suppression of Polyamine Catabolism by Activated Ki-Ras in Human Colon Cancer Cells.” *Molecular Carcinogenesis: Published in Cooperation with the University of Texas MD Anderson Cancer Center* 39 (2): 91–102.

-
- Irizarry, Rafael A, Benjamin M Bolstad, Francois Collin, Leslie M Cope, Bridget Hobbs, and Terence P Speed. 2003. "Summaries of Affymetrix GeneChip Probe Level Data." *Nucleic Acids Research* 31 (4): e15–15.
- Iwai, Miwako, Takayuki Michikawa, Ivan Bosanac, Mitsuhiko Ikura, and Katsuhiko Mikoshiba. 2007. "Molecular Basis of the Isoform-Specific Ligand-Binding Affinity of Inositol 1, 4, 5-Trisphosphate Receptors." *Journal of Biological Chemistry* 282 (17): 12755–64.
- Iyikesici, Mehmet Salih. 2019. "Feasibility Study of Metabolically Supported Chemotherapy with Weekly Carboplatin/Paclitaxel Combined with Ketogenic Diet, Hyperthermia and Hyperbaric Oxygen Therapy in Metastatic Non-Small Cell Lung Cancer." *International Journal of Hyperthermia* 36 (1): 445–54.
- Jain, Shanu, and Kenneth A Jacobson. 2021. "Purinergic Signaling in Liver Pathophysiology." *Frontiers in Endocrinology* 12.
- James, Briana, Sheldon Milstien, and Sarah Spiegel. 2019. "Ormdl3 and Allergic Asthma: From Physiology to Pathology." *Journal of Allergy and Clinical Immunology* 144 (3): 634–40.
- Jang, Meishi, Lining Cai, George O Udeani, Karla V Slowing, Cathy F Thomas, Christopher WW Beecher, Harry HS Fong, et al. 1997. "Cancer Chemopreventive Activity of Resveratrol, a Natural Product Derived from Grapes." *Science* 275 (5297): 218–20.
- Jansen, Natalie, and Harald Walach. 2016. "The Development of Tumours Under a Ketogenic Diet in Association with the Novel Tumour Marker Tktl1: A Case Series in General Practice." *Oncology Letters* 11 (1): 584–92.
- Jardin, Diez-Bello, I. 2021. "Melatonin Downregulates Trpc6, Impairing Store-Operated Calcium Entry in Triple-Negative Breast Cancer Cells." *Journal of Biological Chemistry* 296 (8): 764–69.
- Jardin, Isaac, Raquel Diez-Bello, Jose J Lopez, Pedro C Redondo, Ginés M Salido, Tarik Smani, and Juan A Rosado. 2018. "Trpc6 Channels Are Required for Proliferation, Migration and Invasion of Breast Cancer Cell Lines by Modulation of Orail and Orail3 Surface Exposure." *Cancers* 10 (9): 331.
- Jeffers, John NR. 1967. "Two Case Studies in the Application of Principal Component Analysis." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 16 (3): 225–36.
- Jeong, Jaekwang, Joshua N VanHouten, Pamela Dann, Wonnam Kim, Catherine Sullivan, Herbert Yu, Lance Liotta, et al. 2016. "Pmca2 Regulates Her2 Protein Kinase Localization and Signaling and Promotes Her2-Mediated Breast Cancer." *Proceedings of the National Academy of Sciences* 113 (3): E282–90.
- Jiang, Bing-Hua, and Ling-Zhi Liu. 2009. "Pi3k/PTEN Signaling in Angiogenesis and Tumorigenesis." *Advances in Cancer Research* 102: 19–65.
- Jiang, Dawei, Linlin Zhao, and David E Clapham. 2009. "Genome-Wide RNAi Screen Identifies Letm1 as a Mitochondrial Ca²⁺/h⁺ Antiporter." *Science* 326 (5949): 144–47.

-
- Jin, Liting, Bingchen Han, Emily Siegel, Yukun Cui, Armando Giuliano, and Xiaojiang Cui. 2018. “Breast Cancer Lung Metastasis: Molecular Biology and Therapeutic Implications.” *Cancer Biology & Therapy* 19 (10): 858–68.
- Jolliffe, Ian T. 1973. “Discarding Variables in a Principal Component Analysis. II: Real Data.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 22 (1): 21–31.
- Jones, Brandon C, Laura C Kelley, Yuriy V Loskutov, Kristina M Marinak, Varvara K Kozyreva, Matthew B Smolkin, and Elena N Pugacheva. 2017. “Dual Targeting of Mesenchymal and Amoeboid Motility Hinders Metastatic Behavior.” *Molecular Cancer Research* 15 (6): 670–82.
- Jones, Carrie K, Nellie Byun, and Michael Bubser. 2012. “Muscarinic and Nicotinic Acetylcholine Receptor Agonists and Allosteric Modulators for the Treatment of Schizophrenia.” *Neuropsychopharmacology* 37 (1): 16–42.
- Julio-Pieper, Marcela, Peter J Flor, Timothy G Dinan, and John F Cryan. 2011. “Exciting Times Beyond the Brain: Metabotropic Glutamate Receptors in Peripheral and Non-Neural Tissues.” *Pharmacological Reviews* 63 (1): 35–58.
- Junttila, Melissa R, and Gerard I Evan. 2009. “P53—a Jack of All Trades but Master of None.” *Nature Reviews Cancer* 9 (11): 821–29.
- Kaeberlein, Matt. 2009. “Spermidine Surprise for a Long Life.” *Nature Cell Biology* 11 (11): 1277–78.
- Kalogris, Cristina, Sara Caprodossi, Consuelo Amantini, Federica Lambertucci, Massimo Nabissi, Maria Beatrice Morelli, Valerio Farfariello, et al. 2010. “Expression of Transient Receptor Potential Vanilloid-1 (Trpv1) in Urothelial Cancers of Human Bladder: Relation to Clinicopathological and Molecular Parameters.” *Histopathology* 57 (5): 744–52.
- Kang, Min H, and C Patrick Reynolds. 2009. “Bcl-2 Inhibitors: Targeting Mitochondrial Apoptotic Pathways in Cancer Therapy.” *Clinical Cancer Research* 15 (4): 1126–32.
- Kang, Sang Soo, Kyung-Seok Han, Bo Mi Ku, Yeon Kyung Lee, Jinpyo Hong, Hye Young Shin, Antoine G Almonte, et al. 2010. “Caffeine-Mediated Inhibition of Calcium Release Channel Inositol 1, 4, 5-Trisphosphate Receptor Subtype 3 Blocks Glioblastoma Invasion and Extends Survival.” *Cancer Research* 70 (3): 1173–83.
- Kann, Oliver, and Richard Kovács. 2007. “Mitochondria and Neuronal Activity.” *American Journal of Physiology-Cell Physiology* 292 (2): C641–57.
- Kappel, Sven, Anna Borgström, Paulina Stokłosa, Kathrin Dörr, and Christine Peinelt. 2019. “Store-Operated Calcium Entry in Disease: Beyond STIM/Orai Expression Levels.” In *Seminars in Cell & Developmental Biology*, 94:66–73. Elsevier.
- Kappel, Sven, Paulina Stokłosa, Barbara Hauert, Daniela Ross-Kaschitzka, Anna Borgström, Roland Baur, José A Galván, Inti Zlobec, and Christine Peinelt. 2019. “Trpm4 Is Highly Expressed in Human Colorectal Tumor Buds and Contributes to Proliferation, Cell Cycle, and Invasion of Colorectal Cancer Cells.” *Molecular Oncology* 13 (11): 2393–2405.

-
- Kawai, Kanji, Carrie Viars, Karen Arden, David Tarin, Virginia Urquidi, and Steve Goodison. 2002. "Comprehensive Karyotyping of the HT-29 Colon Adenocarcinoma Cell Line." *Genes, Chromosomes and Cancer* 34 (1): 1–8.
- Kádár, Kristóf, Viktória Juhász, Anna Földes, Róbert Rácz, Yan Zhang, Heike Löchli, Erzsébet Kató, et al. 2021. "Trpm7-Mediated Calcium Transport in HAT-7 Ameloblasts." *International Journal of Molecular Sciences* 22 (8): 3992.
- Kchouk, Mehdi, Jean-Francois Gibrat, and Mourad Elloumi. 2017. "Generations of Sequencing Technologies: From First to Next Generation." *Biology and Medicine* 9 (3).
- Khorshidi, Azam, Preet Dhaliwal, and Burton B Yang. 2016. "Noncoding RNAs in Tumor Angiogenesis." *The Long and Short Non-Coding RNAs in Cancer Biology*, 217–41.
- Kim, Chul Hoon, Stephanie Braud, John TR Isaac, and Katherine W Roche. 2005. "Protein Kinase c Phosphorylation of the Metabotropic Glutamate ReceptormGluR5 on Serine 839 Regulates Ca²⁺ Oscillations." *Journal of Biological Chemistry* 280 (27): 25409–15.
- Kim, Sang Geol, Annie On-On Chan, Tsung-Teh Wu, Jean-Pierre J Issa, Stanley R Hamilton, and Asif Rashid. 2003. "Epigenetic and Genetic Alterations in Duodenal Carcinomas Are Distinct from Biliary and Ampullary Carcinomas." *Gastroenterology* 124 (5): 1300–1310.
- Kitamura, Takanori, Bin-Zhi Qian, and Jeffrey W Pollard. 2015. "Immune Cell Promotion of Metastasis." *Nature Reviews Immunology* 15 (2): 73–86.
- Klement, Rainer J. 2017. "Beneficial Effects of Ketogenic Diets for Cancer Patients: A Realist Review with Focus on Evidence and Confirmation." *Medical Oncology* 34 (8): 1–15.
- Klumpp, Dominik, Stephanie C Frank, Lukas Klumpp, Efe C Sezgin, Marita Eckert, Lena Edalat, Martin Bastmeyer, Daniel Zips, Peter Ruth, and Stephan M Huber. 2017. "Trpm8 Is Required for Survival and Radioresistance of Glioblastoma Cells." *Oncotarget* 8 (56): 95896.
- Koslowski, Michael, Ugur Sahin, Karl Dhaene, Christoph Huber, and Özlem Türeci. 2008. "Ms4a12 Is a Colon-Selective Store-Operated Calcium Channel Promoting Malignant Cell Processes." *Cancer Research* 68 (9): 3458–66.
- Koslowski, Michael, Özlem Türeci, Christoph Huber, and Ugur Sahin. 2009. "Selective Activation of Tumor Growth-Promoting Ca²⁺ Channel Ms4a12 in Colon Cancer by Caudal Type Homeobox Transcription Factor Cdx2." *Molecular Cancer* 8 (1): 1–6.
- Kranias, EG, and Donald M Bers. 2007. "Calcium and Cardiomyopathies." *Calcium Signalling and Disease*, 523–37.
- Kretz, Oliver, Wolfgang Nastainczyk, Markus Hoth, Stephan Philipp, Claudia Trost, Jan Warnat, Julia Rautmann, et al. 2000. "Trp4 (Cce1) Protein Is Part of Native Calcium Release-Activated Ca²⁺-Like Channels in Adrenal Cells." *Journal of Biological Chemistry* 275 (31): 23965–72.
- Kudryavtseva, Anna V, Anastasia V Lipatova, Andrew R Zaretsky, Alexey A Moskalev, Maria S Fedorova, Anastasiya S Rasskazova, Galina A Shibukhova, et al. 2016. "Important Molecular Genetic Markers of Colorectal Cancer." *Oncotarget* 7 (33): 53959.

-
- Kuhn, Max, and Ross Quinlan. 2021. *C50: C5.0 Decision Trees and Rule-Based Models*. <https://CRAN.R-project.org/package=C50>.
- Kurusamy, Sathishkumar, Dolores López-Maderuelo, Robert Little, David Cadagan, Aaron M Savage, Jude C Ihugba, Rhiannon R Baggott, et al. 2017. “Selective Inhibition of Plasma Membrane Calcium ATPase 4 Improves Angiogenesis and Vascular Reperfusion.” *Journal of Molecular and Cellular Cardiology* 109: 38–47.
- Kushner, Peter J, David A Agard, Geoffrey L Greene, Thomas S Scanlan, Andrew K Shiau, Rosalie M Uht, and Paul Webb. 2000. “Estrogen Receptor Pathways to AP-1.” *The Journal of Steroid Biochemistry and Molecular Biology* 74 (5): 311–17.
- Kusner, Linda L, Åse Mygland, and Henry J Kaminski. 1998. “Ryanodine Receptor Gene Expression Thymomas.” *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine* 21 (10): 1299–1303.
- Labelle, Myriam, and Richard O Hynes. 2012. “The Initial Hours of Metastasis: The Importance of Cooperative Host–Tumor Cell Interactions During Hematogenous Dissemination.” *Cancer Discovery* 2 (12): 1091–99.
- Laird, Nan M, and James H Ware. 1982. “Random-Effects Models for Longitudinal Data.” *Biometrics*, 963–74.
- Lange, Ingo, and Dana-Lynn T Koomoa. 2014. “MycN Promotes Trpm7 Expression and Cell Migration in Neuroblastoma Through a Process That Involves Polyamines.” *FEBS Open Bio* 4: 966–75.
- Lantz, Brett. 2013. *Machine Learning with r*. Packt Publishing Ltd.
- Law, Charity W, Yunshun Chen, Wei Shi, and Gordon K Smyth. 2014. “Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts.” *Genome Biology* 15 (2): 1–17.
- Lawson, Kathryn R, Natalia A Ignatenko, Gary A Piazza, Hiayan Cui, and Eugene W Gerner. 2000. “Influence of k-Ras Activation on the Survival Responses of Caco-2 Cells to the Chemopreventive Agents Sulindac and Difluoromethylornithine.” *Cancer Epidemiology and Prevention Biomarkers* 9 (11): 1155–62.
- Lee, Kyu Pil, Joseph P Yuan, Weizhong Zeng, Insuk So, Paul F Worley, and Shmuel Muallem. 2009. “Molecular Determinants of Fast Ca²⁺-Dependent Inactivation and Gating of the Orai Channels.” *Proceedings of the National Academy of Sciences* 106 (34): 14687–92.
- Lee, Wen Hsin, Lee Yee Choong, Naing Naing Mon, SsuYi Lu, Qingsong Lin, Brendan Pang, Benedict Yan, et al. 2016. “Trpv4 Regulates Breast Cancer Cell Extravasation, Stiffness and Actin Cortex.” *Scientific Reports* 6 (1): 1–16.
- Lee, WH, LY Choong, TH Jin, NN Mon, S Chong, CS Liew, T Putti, SY Lu, C Harteneck, and YP Lim. 2017. “Trpv4 Plays a Role in Breast Cancer Cell Migration via Ca²⁺-Dependent

-
- Activation of AKT and Downregulation of e-Cadherin Cell Cortex Protein.” *Oncogenesis* 6 (5): e338–38.
- Lee, Won Jae, Sarah J Roberts-Thomson, Nicola A Holman, Fiona J May, Gillian M Lehrbach, and Gregory R Monteith. 2002. “Expression of Plasma Membrane Calcium Pump Isoform mRNAs in Breast Cancer Cell Lines.” *Cellular Signalling* 14 (12): 1015–22.
- Lee, Won Jae, Sarah J Roberts-Thomson, and Gregory R Monteith. 2005. “Plasma Membrane Calcium-ATPase 2 and 4 in Human Breast Cancer Cell Lines.” *Biochemical and Biophysical Research Communications* 337 (3): 779–83.
- Lefranc, Florence. 2021. “Transient Receptor Potential (TRP) Ion Channels Involved in Malignant Glioma Cell Death and Therapeutic Perspectives.” *Frontiers in Cell and Developmental Biology* 9.
- Lemmon, Mark A, and Joseph Schlessinger. 2010. “Cell Signaling by Receptor Tyrosine Kinases.” *Cell* 141 (7): 1117–34.
- Levine, Beth, and Guido Kroemer. 2008. “Autophagy in the Pathogenesis of Disease.” *Cell* 132 (1): 27–42.
- Lewis, Nathan E, and Alyaa M Abdel-Haleem. 2013. “The Evolution of Genome-Scale Models of Cancer Metabolism.” *Frontiers in Physiology* 4: 237.
- Li, Hong, Herman AJ Schut, Philip Conran, Paula M Kramer, Ronald A Lubet, Vernon E Steele, Ernest E Hawk, Gary J Kelloff, and Michael A Pereira. 1999. “Prevention by Aspirin and Its Combination with α -Difluoromethylornithine of Azoxymethane-Induced Tumors, Aberrant Crypt Foci and Prostaglandin E2 Levels in Rat Colon.” *Carcinogenesis* 20 (3): 425–30.
- Li, Shengnan, Kuo Jiang, Jia Li, Xiaohua Hao, Wenguang Chu, Ceng Luo, Yuanyuan Zhu, Rougang Xie, and Biliang Chen. 2020. “Estrogen Enhances the Proliferation and Migration of Ovarian Cancer Cells by Activating Transient Receptor Potential Channel C3.” *Journal of Ovarian Research* 13 (1): 1–11.
- Li, Su W, John Westwick, and Chris T Poll. 2002. “Receptor-Operated Ca^{2+} Influx Channels in Leukocytes: A Therapeutic Target?” *Trends in Pharmacological Sciences* 23 (2): 63–70.
- Li, Wenjun, Minhong Zhang, Lei Xu, Danmiao Lin, Shaoxi Cai, and Fei Zou. 2013. “The Apoptosis of Non-Small Cell Lung Cancer Induced by Cisplatin Through Modulation of Stim1.” *Experimental and Toxicologic Pathology* 65 (7-8): 1073–81.
- Liao, Jun, Hua Li, Weizhong Zeng, David B Sauer, Ricardo Belmares, and Youxing Jiang. 2012. “Structural Insight into the Ion-Exchange Mechanism of the Sodium/Calcium Exchanger.” *Science* 335 (6069): 686–90.
- Liberzon, Arthur, Aravind Subramanian, Reid Pinchback, Helga Thorvaldsdóttir, Pablo Tamayo, and Jill P Mesirov. 2011. “Molecular Signatures Database (MSigDB) 3.0.” *Bioinformatics* 27 (12): 1739–40.

-
- Liou, Jen, Man Lyang Kim, Won Do Heo, Joshua T Jones, Jason W Myers, James E Ferrell Jr, and Tobias Meyer. 2005. "STIM Is a Ca²⁺ Sensor Essential for Ca²⁺-Store-Depletion-Triggered Ca²⁺ Influx." *Current Biology* 15 (13): 1235–41.
- Lis, Annette, Christine Peinelt, Andreas Beck, Suhel Parvez, Mahealani Monteilh-Zoller, Andrea Fleig, and Reinhold Penner. 2007. "Cracm1, Cracm2, and Cracm3 Are Store-Operated Ca²⁺ Channels with Distinct Functional Properties." *Current Biology* 17 (9): 794–800.
- Liu, Chunming, Yiming Li, Mikhail Semenov, Chun Han, Gyeong-Hun Baeg, Yi Tan, Zhuohua Zhang, Xinhua Lin, and Xi He. 2002. "Control of β -Catenin Phosphorylation/Degradation by a Dual-Kinase Mechanism." *Cell* 108 (6): 837–47.
- Liu, Huiling, Joshua D Hughes, Shadon Rollins, Betty Chen, and Eddie Perkins. 2011. "Calcium Entry via Orail Regulates Glioblastoma Cell Proliferation and Apoptosis." *Experimental and Molecular Pathology* 91 (3): 753–60.
- Liu, Jun, Jia Huang, Wei-Yan Yao, Qi-Wen Ben, Da-Fan Chen, Xiang-Yi He, Lei Li, and Yao-Zong Yuan. 2012. "The Origins of Vascularization in Tumors." *Front Biosci* 17 (1): 2559–65.
- Liu, Xiaofen, Guangyan Wu, Yi Yu, Xiaozhe Chen, Renci Ji, Jing Lu, Xin Li, Xing Zhang, Xue Yang, and Yuequan Shen. 2019. "Molecular Understanding of Calcium Permeation Through the Open Orail Channel." *PLoS Biology* 17 (4): e3000096.
- Liu, Xiaoyu, Peng Zhang, Chuanming Xie, Kathy WY Sham, Simon SM Ng, Yangchao Chen, and Christopher HK Cheng. 2019. "Activation of PTEN by Inhibition of Trpv4 Suppresses Colon Cancer Development." *Cell Death & Disease* 10 (6): 1–14.
- Liu, Xibao, Brij B Singh, and Indu S Ambudkar. 2003. "Trpc1 Is Required for Functional Store-Operated Ca²⁺ Channels: Role of Acidic Amino Acid Residues in the S5-S6 Region." *Journal of Biological Chemistry* 278 (13): 11337–43.
- Liu, Yitao, Tak Pan Wong, Michelle Aarts, Amanda Rooyakkers, Lidong Liu, Ted Weita Lai, Dong Chuan Wu, et al. 2007. "NMDA Receptor Subunits Have Differential Roles in Mediating Excitotoxic Neuronal Death Both in Vitro and in Vivo." *Journal of Neuroscience* 27 (11): 2846–57.
- Locatelli, Laura, Alessandra Cazzaniga, Giorgia Fedele, Monica Zocchi, Roberta Scrimieri, Claudia Moscheni, Sara Castiglioni, and Jeanette A Maier. 2021. "A Comparison of Doxorubicin-Resistant Colon Cancer LoVo and Leukemia HL60 Cells: Common Features, Different Underlying Mechanisms." *Current Issues in Molecular Biology* 43 (1): 163–75.
- Logan, Catriona Y, and Roel Nusse. 2004. "The Wnt Signaling Pathway in Development and Disease." *Annu. Rev. Cell Dev. Biol.* 20: 781–810.
- Long, J Scott, and Laurie H Ervin. 2000. "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." *The American Statistician* 54 (3): 217–24.
- Lopez, Jose J, Letizia Albarran, Luis J Gómez, Tarik Smani, Gines M Salido, and Juan A Rosado.

-
2016. “Molecular Modulators of Store-Operated Calcium Entry.” *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* 1863 (8): 2037–43.
- Lortie, Mark J, William F Novotny, Orjan W Peterson, Volker Vallon, Kirsten Malvey, Margarida Mendonca, Joseph Satriano, et al. 1996. “Agmatine, a Bioactive Metabolite of Arginine. Production, Degradation, and Functional Effects in the Kidney of the Rat.” *The Journal of Clinical Investigation* 97 (2): 413–20.
- Love, Michael I, Wolfgang Huber, and Simon Anders. 2014a. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15: 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Love, Michael I, Wolfgang Huber, and Simon Anders. 2014b. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 1–21.
- Ma, Huan, Samuel Cohen, Boxing Li, and Richard W Tsien. 2012. “Exploring the Dominant Role of Cav1 Channels in Signalling to the Nucleus.” *Bioscience Reports* 33 (1): e00009.
- Ma, Xin, Yanfei Cai, Dongxu He, Chang Zou, Peng Zhang, Chun Yin Lo, Zhenyu Xu, et al. 2012. “Transient Receptor Potential Channel Trpc5 Is Essential for p-Glycoprotein Induction in Drug-Resistant Cancer Cells.” *Proceedings of the National Academy of Sciences* 109 (40): 16282–87.
- Ma, Xin, Kwong-Tai Cheng, Ching-On Wong, Roger G O’Neil, Lutz Birnbaumer, Indu S Ambudkar, and Xiaoqiang Yao. 2011. “Heteromeric Trpv4-C1 Channels Contribute to Store-Operated Ca²⁺ Entry in Vascular Endothelial Cells.” *Cell Calcium* 50 (6): 502–9.
- MacDonald, James W. 2020. *Affycoretools: Functions Useful for Those Doing Repetitive Analyses with Affymetrix GeneChips*.
- Mammano, Fabio, and Mario Bortolozzi. 2010. “Ca²⁺ Imaging: Principles of Analysis and Enhancement.” In *Calcium Measurement Methods*, 57–80. Springer.
- Mann, Shivtanj, Manreet Sidhu, and Krisstina Gowin. 2020. “Understanding the Mechanisms of Diet and Outcomes in Colon, Prostate, and Breast Cancer; Malignant Gliomas; and Cancer Patients on Immunotherapy.” *Nutrients* 12 (8): 2226.
- Mansvelder, Huibert D, J Russel Keath, and Daniel S McGehee. 2002. “Synaptic Mechanisms Underlie Nicotine-Induced Excitability of Brain Reward Areas.” *Neuron* 33 (6): 905–19.
- Mansvelder, Huibert D, and Daniel S McGehee. 2000. “Long-Term Potentiation of Excitatory Inputs to Brain Reward Areas by Nicotine.” *Neuron* 27 (2): 349–57.
- Manterola, Carlos, Guissella Quiroz, Paulina Salazar, and Nayeli García. 2019. “Metodología de Los Tipos y Diseños de Estudio más Frecuentemente Utilizados En Investigación Clínica.” *Revista Médica Clínica Las Condes* 30 (1): 36–49.
- Marchi, Saverio, Laura Lupini, Simone Patergnani, Alessandro Rimessi, Sonia Missiroli, Massimo Bonora, Angela Bononi, et al. 2013. “Downregulation of the Mitochondrial Calcium Uniporter by Cancer-Related miR-25.” *Current Biology* 23 (1): 58–63.

-
- Mark, Leighton P, Robert W Prost, John L Ulmer, Michelle M Smith, David L Daniels, James M Strottmann, W Douglas Brown, and Lotfi Hacein-Bey. 2001. "Pictorial Review of Glutamate Excitotoxicity: Fundamental Concepts for Neuroimaging." *American Journal of Neuroradiology* 22 (10): 1813–24.
- Marot, Guillemette, Jean-Louis Foulley, Claus-Dieter Mayer, and Florence Jaffrézic. 2009. "Moderated Effect Size and p-Value Combinations for Microarray Meta-Analyses." *Bioinformatics* 25 (20): 2692–99.
- Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. 2017. "Balancing Type I Error and Power in Linear Mixed Models." *Journal of Memory and Language* 94: 305–15.
- McCarthy, Davis J, Yunshun Chen, and Gordon K Smyth. 2012. "Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation." *Nucleic Acids Research* 40 (10): 4288–97.
- McKenna, Neil J, and Bert W O'Malley. 2001. "Nuclear Receptors, Coregulators, Ligands, and Selective Receptor Modulators: Making Sense of the Patchwork Quilt." *Annals of the New York Academy of Sciences* 949 (1): 3–5.
- Meissner, Gerhard. 2017. "The Structural Basis of Ryanodine Receptor Ion Channel Function." *Journal of General Physiology* 149 (12): 1065–89.
- Mercer, Jason C, Wayne I DeHaven, Jeremy T Smyth, Barbara Wedel, Rebecca R Boyles, Gary S Bird, and James W Putney Jr. 2006. "Large Store-Operated Calcium Selective Currents Due to Co-Expression of Orai1 or Orai2 with the Intracellular Calcium Sensor, Stim1." *Journal of Biological Chemistry* 281 (34): 24979–90.
- Mesquita, Gonçalo, Natalia Prevarskaya, Albrecht Schwab, and others. 2021. "Role of the TRP Channels in Pancreatic Ductal Adenocarcinoma Development and Progression." *Cells* 10 (5): 1021.
- Messina, Mark J, Victoria Persky, Kenneth DR Setchell, and Stephen Barnes. 1994. "Soy Intake and Cancer Risk: A Review of the in Vitro and in Vivo Data." *Nutrition and Cancer* 21 (2): 113–31.
- Metzker, Michael L. 2010. "Sequencing Technologies—the Next Generation." *Nature Reviews Genetics* 11 (1): 31–46.
- Meyskens, Frank L, Christine E McLaren, Daniel Pelot, Sharon Fujikawa-Brooks, Philip M Carpenter, Ernest Hawk, Gary Kelloff, et al. 2008. "Difluoromethylornithine Plus Sulindac for the Prevention of Sporadic Colorectal Adenomas: A Randomized Placebo-Controlled, Double-Blind Trial." *Cancer Prevention Research* 1 (1): 32–38.
- Miederer, Anna-Maria, Dalia Alansary, Gertrud Schwär, Po-Hsien Lee, Martin Jung, Volkhard Helms, and Barbara A Niemeyer. 2015. "A Stim2 Splice Variant Negatively Regulates Store-Operated Calcium Entry." *Nature Communications* 6 (1): 1–12.

-
- Mignen, Olivier, Jill L Thompson, and Trevor J Shuttleworth. 2007. "Stim1 Regulates Ca²⁺ Entry via Arachidonate-Regulated Ca²⁺-Selective (ARC) Channels Without Store Depletion or Translocation to the Plasma Membrane." *The Journal of Physiology* 579 (3): 703–15.
- . 2008. "Both Orai1 and Orai3 Are Essential Components of the Arachidonate-Regulated Ca²⁺-Selective (ARC) Channels." *The Journal of Physiology* 586 (1): 185–95.
- . 2009. "The Molecular Architecture of the Arachidonate-Regulated Ca²⁺-Selective ARC Channel Is a Pentameric Assembly of Orai1 and Orai3 Subunits." *The Journal of Physiology* 587 (17): 4181–97.
- Miller, Melissa B, and Yi-Wei Tang. 2009. "Basic Concepts of Microarrays and Potential Applications in Clinical Microbiology." *Clinical Microbiology Reviews* 22 (4): 611–33.
- Minois, Nadège. 2014. "Molecular Basis of the 'Anti-Aging' effect of Spermidine and Other Natural Polyamines—a Mini-Review." *Gerontology* 60 (4): 319–26.
- Minois, Nadège, Didac Carmona-Gutierrez, and Frank Madeo. 2011. "Polyamines in Aging and Disease." *Aging (Albany NY)* 3 (8): 716.
- Missiaen, Ludwig, Leonard Dode, Jo Vanoevelen, Luc Raeymaekers, and Frank Wuytack. 2007. "Calcium in the Golgi Apparatus." *Cell Calcium* 41 (5): 405–16.
- Mizuno, Hideki, Yoshiro Suzuki, Masaki Watanabe, Takaaki Sokabe, Tokunori Yamamoto, Ryohei Hattori, Momokazu Gotoh, and Makoto Tominaga. 2014. "Potential Role of Transient Receptor Potential (TRP) Channels in Bladder Cancer Cells." *The Journal of Physiological Sciences* 64 (4): 305–14.
- Moinard, Christophe, Luc Cynober, and Jean-Pascal de Bandt. 2005. "Polyamines: Metabolism and Implications in Human Diseases." *Clinical Nutrition* 24 (2): 184–97.
- Monteith, Gregory R, Felicity M Davis, and Sarah J Roberts-Thomson. 2012. "Calcium Channels and Pumps in Cancer: Changes and Consequences." *Journal of Biological Chemistry* 287 (38): 31666–73.
- Monteith, Gregory R, Damara McAndrew, Helen M Faddy, and Sarah J Roberts-Thomson. 2007. "Calcium and Cancer: Targeting Ca²⁺ Transport." *Nature Reviews Cancer* 7 (7): 519–30.
- Montero, Mayte, J Garcia-Sancho, and Javier Alvarez. 1993a. "Inhibition of the Calcium Store-Operated Calcium Entry Pathway by Chemotactic Peptide and by Phorbol Ester Develops Gradually and Independently Along Differentiation of HL60 Cells." *Journal of Biological Chemistry* 268 (36): 26911–19.
- . 1993b. "Transient Inhibition by Chemotactic Peptide of a Store-Operated Ca²⁺ Entry Pathway in Human Neutrophils." *Journal of Biological Chemistry* 268 (18): 13055–61.
- Mootha, Vamsi K, Cecilia M Lindgren, Karl-Fredrik Eriksson, Aravind Subramanian, Smita Sihag, Joseph Lehar, Pere Puigserver, et al. 2003. "PGC-1 α -Responsive Genes Involved in Oxidative

-
- Phosphorylation Are Coordinately Downregulated in Human Diabetes.” *Nature Genetics* 34 (3): 267–73.
- Morris, Jordan L, Germain Gillet, Julien Prudent, and Nikolay Popgeorgiev. 2021. “Bcl-2 Family of Proteins in the Control of Mitochondrial Calcium Signalling: An Old Chap with New Roles.” *International Journal of Molecular Sciences* 22 (7): 3730.
- Motiani, Rajender K, Xuexin Zhang, Kelly E Harmon, Rebecca S Keller, Khalid Matrougui, James A Bennett, and Mohamed Trebak. 2013. “Orai3 Is an Estrogen Receptor α -Regulated Ca²⁺ Channel That Promotes Tumorigenesis.” *The FASEB Journal* 27 (1): 63–75.
- Mousses, Spyro. 2011. “Microarray (cDNA) Technology.” In *Encyclopedia of Cancer*, edited by Manfred Schwab, 2289–92. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-16483-5_3710.
- Moyer, Mary Pat, Lawrence A Manzano, Ronald L Merriman, Jay S Stauffer, and Lee R Tanzer. 1996. “Ncm460, a Normal Human Colon Mucosal Epithelial Cell Line.” *In Vitro Cellular & Developmental Biology-Animal* 32 (6): 315–17.
- Mulholland, David J, Shoukat Dedhar, Gerhard A Coetzee, and Colleen C Nelson. 2005. “Interaction of Nuclear Receptors with the Wnt/ β -Catenin/Tcf Signaling Axis: Wnt You Like to Know?” *Endocrine Reviews* 26 (7): 898–915.
- Mullins, Franklin M, and Richard S Lewis. 2016. “The Inactivation Domain of Stim1 Is Functionally Coupled with the Orai1 Pore to Enable Ca²⁺-Dependent Inactivation.” *Journal of General Physiology* 147 (2): 153–64.
- Mullins, Franklin M, Michelle Yen, and Richard S Lewis. 2016. “Orai1 Pore Residues Control CRAC Channel Inactivation Independently of Calmodulin.” *Journal of General Physiology* 147 (2): 137–52.
- Mundade, Rasika, Thomas F Imperiale, Lakshmi Prabhu, Patrick J Loehrer, and Tao Lu. 2014. “Genetic Pathways, Prevention, and Treatment of Sporadic Colorectal Cancer.” *Oncoscience* 1 (6): 400.
- Nakamura, Eliane Shizuka, Keiichi Koizumi, Mitsuo Kobayashi, Yurika Saitoh, Yoshihisa Arita, Takashi Nakayama, Hiroaki Sakurai, Osamu Yoshie, and Ikuo Saiki. 2006. “RANKL-Induced Ccl22/Macrophage-Derived Chemokine Produced from Osteoclasts Potentially Promotes the Bone Metastasis of Lung Cancer Expressing Its Receptor Ccr4.” *Clinical & Experimental Metastasis* 23 (1): 9–18.
- Natarajan, Gayathri K, Jyotsna Mishra, Amadou KS Camara, and Wai-Meng Kwok. 2021. “Letm1: A Single Entity with Diverse Impact on Mitochondrial Metabolism and Cellular Signaling.” *Frontiers in Physiology* 12: 321.
- “National Human Genome Research Institute (NHGRI) Web Site Genome.gov Is in the Public Domain.” 2014. <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>.

-
- Nedergaard, Maiken, Takahiro Takano, and Anker J Hansen. 2002. "Beyond the Role of Glutamate as a Neurotransmitter." *Nature Reviews Neuroscience* 3 (9): 748–55.
- Nelson, David L, and Michael M Lehninger Cox. 2016. *Lehninger Principios de Bioquímica, 6 Edición*. OMEGA.
- Neophytou, Christiana M, Theodora-Christina Kyriakou, and Panagiotis Papageorgis. 2019. "Mechanisms of Metastatic Tumor Dormancy and Implications for Cancer Therapy." *International Journal of Molecular Sciences* 20 (24): 6158.
- Network, Cancer Genome Atlas, and others. 2012. "Comprehensive Molecular Characterization of Human Colon and Rectal Cancer." *Nature* 487 (7407): 330.
- Neubauer, Katharina, and Barbara Zieger. 2017. "The Mammalian Septin Interactome." *Frontiers in Cell and Developmental Biology* 5: 3.
- Newmark, Harold L. 1996. "Plant Phenolics as Potential Cancer Prevention Agents." *Dietary Phytochemicals in Cancer Prevention and Treatment*, 25–34.
- Nilius, Bernd, Jean Prenen, Thomas Voets, and Guy Droogmans. 2004. "Intracellular Nucleotides and Polyamines Inhibit the Ca²⁺-Activated Cation Channel TRPM4b." *Pflügers Archiv* 448 (1): 70–75.
- Niu, Liling, Fuyun Wu, Kaili Li, Jing Li, Shenyuan L Zhang, Junjie Hu, and Qian Wang. 2020. "Stim1 Interacts with Termini of Orai Channels in a Sequential Manner." *Journal of Cell Science* 133 (8).
- Novello, Matthew J, Jinhui Zhu, MengQi Zhang, Qingping Feng, and Peter B Stathopoulos. 2020. "Synergistic Stabilization by Nitrosoglutathione-Induced Thiol Modifications in the Stromal Interaction Molecule-2 Luminal Domain Suppresses Basal and Store Operated Calcium Entry." *Scientific Reports* 10 (1): 1–17.
- Nowotarski, Shannon L, Patrick M Woster, and Robert A Casero. 2013. "Polyamines and Cancer: Implications for Chemotherapy and Chemoprevention." *Expert Reviews in Molecular Medicine* 15.
- Núñez, Lucía, Ruth A Valero, Laura Senovilla, Sara Sanz-Blasco, Javier García-Sancho, and Carlos Villalobos. 2006. "Cell Proliferation Depends on Mitochondrial Ca²⁺ Uptake: Inhibition by Salicylate." *The Journal of Physiology* 571 (1): 57–73.
- Ochoa-Hernández, Alejandra Berenice, Clara Ibet Juárez-Vázquez, Mónica Alejandra Rosales-Reynoso, and Patricio Barros-Núñez. 2012. "La vía de Señalización Wnt-b-Catenina y Su Relación Con cáncer." *Cirugía y Cirujanos* 80 (4): 389–98.
- Oehlert, Gary W. 2010. *A First Course in Design and Analysis of Experiments*.
- Ong, Hwei Ling, Kwong Tai Cheng, Xibao Liu, Bidhan C Bandyopadhyay, Biman C Paria, Jonathan Soboloff, Biswaranjan Pani, et al. 2007. "Dynamic Assembly of Trpc1-Stim1-Orai1 Ternary Complex Is Involved in Store-Operated Calcium Influx: Evidence for Similarities in

-
- Store-Operated and Calcium Release-Activated Calcium Channel Components.” *Journal of Biological Chemistry* 282 (12): 9105–16.
- Ortega-Gutiérrez, Silvia. 2013. “Avances En El Estudio de Receptores Acoplados a Proteínas g,” no. 4: 276–84.
- Otto, Warburg. 1956. “On the Origin of Cancer Cells.” *Science* 123 (3191): 309–14.
- Pacifico, F, L Ulianich, S De Micheli, S Treglia, A Leonardi, P Vito, S Formisano, E Consiglio, and B Di Jeso. 2003. “The Expression of the Sarco/Endoplasmic Reticulum Ca²⁺-ATPases in Thyroid and Its down-Regulation Following Neoplastic Transformation.” *Journal of Molecular Endocrinology* 30 (3): 399–409.
- Pagès, Hervé, Marc Carlson, Seth Falcon, and Nianhua Li. 2020. *AnnotationDbi: Manipulation of SQLite-Based Annotations in Bioconductor*. <https://bioconductor.org/packages/AnnotationDbi>.
- Palty, Raz, William F Silverman, Michal Hershinkel, Teresa Caporale, Stefano L Sensi, Julia Parnis, Christiane Nolte, et al. 2010. “NCLX Is an Essential Component of Mitochondrial Na⁺/Ca²⁺ Exchange.” *Proceedings of the National Academy of Sciences* 107 (1): 436–41.
- Pan, Pengchao, Zhenyu Cai, Chunlin Zhuang, Xiaofei Chen, and Yifeng Chai. 2019. “Methodology of Drug Screening and Target Identification for New Necroptosis Inhibitors.” *Journal of Pharmaceutical Analysis* 9 (2): 71–76.
- Pang, Ji-Jing, John Alexander, Bo Lei, Wentao Deng, Keqing Zhang, Chang Li Qihong, and William W Hauswirth. 2010. “Achromatopsia as a Potential Candidate for Gene Therapy.” In *Retinal Degenerative Diseases*, 639–46. Springer.
- Papp, Béla, and Jean-Philippe Brouland. 2011. “Altered Endoplasmic Reticulum Calcium Pump Expression During Breast Tumorigenesis.” *Breast Cancer: Basic and Clinical Research* 5: BCBCR–S7481.
- Parekh, Anant B, and James W Putney Jr. 2005. “Store-Operated Calcium Channels.” *Physiological Reviews* 85 (2): 757–810.
- Park, Jae-Il, Andrew S Venteicher, Ji Yeon Hong, Jinkuk Choi, Sohee Jun, Marina Shkreli, Woody Chang, et al. 2009. “Telomerase Modulates Wnt Signalling by Association with Target Gene Chromatin.” *Nature* 460 (7251): 66–72.
- Parvez, Suhel, Andreas Beck, Christine Peinelt, Jonathan Soboloff, Annette Lis, Mahealani Monteilh-Zoller, Donald L Gill, Andrea Fleig, and Reinhold Penner. 2008. “Stim2 Protein Mediates Distinct Store-Dependent and Store-Independent Modes of CRAC Channel Activation.” *The FASEB Journal* 22 (3): 752–61.
- Pathak, Trayambak, Maxime Gueguinou, Vonn Walter, Celine Delierneux, Martin T Johnson, Xuexin Zhang, Ping Xin, et al. 2020. “Dichotomous Role of the Human Mitochondrial Na⁺/Ca²⁺/Li⁺ Exchanger NCLX in Colorectal Cancer Growth and Metastasis.” *Elife* 9: e59686.

-
- Patil, Mayur J, Nathaniel A Jeske, and Armen N Akopian. 2010. “Transient Receptor Potential V1 Regulates Activation and Modulation of Transient Receptor Potential A1 by Ca²⁺.” *Neuroscience* 171 (4): 1109–19.
- Patirupanusara, Pantida, and Jackrit Suthakorn. 2012. “Introduction of an Active DNA Microarray Fabrication for Medical Applications.” In *International Conference on Advances in Electrical and Electronics Engineering*, 75–79.
- Patron, Maria, Veronica Granatiero, Javier Espino, Rosario Rizzuto, and Diego De Stefani. 2019. “Micu3 Is a Tissue-Specific Enhancer of Mitochondrial Calcium Uptake.” *Cell Death & Differentiation* 26 (1): 179–95.
- Paz, Maria F, Susan Wei, Juan C Cigudosa, Sandra Rodriguez-Perales, Miguel A Peinado, Tim Hui-Ming Huang, and Manel Esteller. 2003. “Genetic Unmasking of Epigenetically Silenced Tumor Suppressor Genes in Colon Cancer Cells Deficient in DNA Methyltransferases.” *Human Molecular Genetics* 12 (17): 2209–19.
- Penner, Reinhold, Gary Matthews, and Erwin Neher. 1988. “Regulation of Calcium Influx by Second Messengers in Rat Mast Cells.” *Nature* 334 (6182): 499–504.
- Peña, Daniel. 2002a. *Análisis de Datos Multivariantes*.
- . 2002b. *Regresión y Diseño de Experimentos*.
- Peters, Amelia A, Michael JG Milevskiy, Wei C Lee, Merrill C Curry, Chanel E Smart, Jodi M Saunus, Lynne Reid, et al. 2016. “The Calcium Pump Plasma Membrane Ca²⁺-ATPase 2 (Pmca2) Regulates Breast Cancer Cell Proliferation and Sensitivity to Doxorubicin.” *Scientific Reports* 6 (1): 1–13.
- Phan, Nam Nhut, Chih-Yang Wang, Chien-Fu Chen, Zhengda Sun, Ming-Derg Lai, and Yen-Chang Lin. 2017. “Voltage-Gated Calcium Channels: Novel Targets for Cancer Therapy.” *Oncology Letters* 14 (2): 2059–74.
- Piao, Longzhen, Yuwen Li, Soung Jung Kim, Hee Sun Byun, Song Mei Huang, Soon-Kyung Hwang, Keum-Jin Yang, et al. 2009. “Association of Letm1 and Mrpl36 Contributes to the Regulation of Mitochondrial ATP Production and Necrotic Cell Death.” *Cancer Research* 69 (8): 3397–3404.
- Pierro, Cristina, Simon J Cook, Thomas CF Foets, Martin D Bootman, and H Llewelyn Roderick. 2014. “Oncogenic k-Ras Suppresses Ip3-Dependent Ca²⁺ Release Through Remodelling of the Isoform Composition of IP3Rs and ER Luminal Ca²⁺ Levels in Colorectal Cancer Cell Lines.” *Journal of Cell Science* 127 (7): 1607–19.
- Pinheiro, Jose, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. 2021. *nlme: Linear and Nonlinear Mixed Effects Models*. <https://CRAN.R-project.org/package=nlme>.
- Pirinen, Eija, Teemu Kuulasmaa, Marko Pietilä, Sami Heikkinen, Maija Tusa, Paula Itkonen, Susanna Boman, et al. 2007. “Enhanced Polyamine Catabolism Alters Homeostatic Control of

White Adipose Tissue Mass, Energy Expenditure, and Glucose Metabolism.” *Molecular and Cellular Biology* 27 (13): 4953–67.

- Pla, A Fiorio, HL Ong, KT Cheng, Alessia Brossa, Benedetta Bussolati, T Lockwich, B Paria, L Munaron, and IS Ambudkar. 2012. “Trpv4 Mediates Tumor-Derived Endothelial Cell Migration via Arachidonic Acid-Activated Actin Remodeling.” *Oncogene* 31 (2): 200–212.
- Pla, Alessandra Fiorio, Cristina Grange, Susanna Antoniotti, Cristiana Tomatis, Annalisa Merlino, Benedetta Bussolati, and Luca Munaron. 2008. “Arachidonic Acid-Induced Ca²⁺ Entry Is Involved in Early Steps of Tumor Angiogenesis.” *Molecular Cancer Research* 6 (4): 535–45.
- Poff, Angela, Andrew P Koutnik, Kathleen M Egan, Solmaz Sahebjam, Dominic D’Agostino, and Nagi B Kumar. 2019. “Targeting the Warburg Effect for Cancer Treatment: Ketogenic Diets for Management of Glioma.” In *Seminars in Cancer Biology*, 56:135–48. Elsevier.
- Polakis, Paul. 1999. “The Oncogenic Activation of β -Catenin.” *Current Opinion in Genetics & Development* 9 (1): 15–21.
- Potier, Marie, and Mohamed Trebak. 2008. “New Developments in the Signaling Mechanisms of the Store-Operated Calcium Entry Pathway.” *Pflügers Archiv-European Journal of Physiology* 457 (2): 405–15.
- Poturnajova, Martina, Tatiana Furielova, Sona Balintova, Silvia Schmidtova, Lucia Kucerova, and Miroslava Matuskova. 2021. “Molecular Features and Gene Expression Signature of Metastatic Colorectal Cancer.” *Oncology Reports* 45 (4): 1–1.
- Prakriya, Murali, Stefan Feske, Yousang Gwack, Sonal Srikanth, Anjana Rao, and Patrick G Hogan. 2006. “Orai1 Is an Essential Pore Subunit of the CRAC Channel.” *Nature* 443 (7108): 230–33.
- Prakriya, Murali, and Richard S Lewis. 2015. “Store-Operated Calcium Channels.” *Physiological Reviews* 95 (4): 1383–1436.
- Prevarskaya, Natalia, Halima Ouadid-Ahidouch, Roman Skryma, and Yaroslav Shuba. 2014. “Remodelling of Ca²⁺ Transport in Cancer: How It Contributes to Cancer Hallmarks?” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1638): 20130097.
- Prevarskaya, Natalia, Roman Skryma, Gabriel Bidaux, Matthieu Flourakis, and Yaroslav Shuba. 2007. “Ion Channels in Death and Differentiation of Prostate Cancer Cells.” *Cell Death & Differentiation* 14 (7): 1295–1304.
- Putney Jr, James W. 1986. “A Model for Receptor-Regulated Calcium Entry.” *Cell Calcium* 7 (1): 1–12.
- . 2004. “The Enigmatic TRPCs: Multifunctional Cation Channels.” *Trends in Cell Biology* 14 (6): 282–86.
- Qi, Lisha, Wangzhao Song, Lingmei Li, Lu Cao, Yue Yu, Chunmin Song, Yalei Wang, et al. 2016. “Fgf4 Induces Epithelial-Mesenchymal Transition by Inducing Store-Operated Calcium Entry in Lung Adenocarcinoma.” *Oncotarget* 7 (45): 74015.

-
- R Diez-Bello, GM Salido, I Jardin. 2012. "Orai1 and Orai2 Mediate Store-Operated Calcium Entry That Regulates HL60 Cell Migration and FAK Phosphorylation." *Biochimica Et Biophysica Acta* 6 (1864): 1064–70.
- R., Steele. 2013. *Model Selection for Multilevel Models*. Scott, Marc a and Simonoff, Jeffrey s and Marx and Brian d. In the *SAGE Handbook of Multilevel Modeling*. Sage.
- Raisinghani, Manish, Linlin Zhong, Joseph A Jeffrey, Mahendra Bishnoi, Reddy M Pabbidi, Fátima Pimentel, De-Shou Cao, M Steven Evans, and Louis S Premkumar. 2011. "Activation Characteristics of Transient Receptor Potential Ankyrin 1 and Its Role in Nociception." *American Journal of Physiology-Cell Physiology* 301 (3): C587–600.
- Rajagopal, Senthilkumar, and Murugavel Ponnusamy. 2017. *Calcium Signaling: From Physiology to Diseases*. Springer.
- Rana, Anshul, Michelle Yen, Amir Masoud Sadaghiani, Seth Malmersjö, Chan Young Park, Ricardo E Dolmetsch, and Richard S Lewis. 2015. "Alternative Splicing Converts Stim2 from an Activator to an Inhibitor of Store-Operated Calcium Channels." *Journal of Cell Biology* 209 (5): 653–70.
- Rathner, Petr, Marc Fahrner, Linda Cerofolini, Herwig Grabmayr, Ferdinand Horvath, Heinrich Krobath, Agrim Gupta, et al. 2021. "Interhelical Interactions Within the Stim1 Cc1 Domain Modulate CRAC Channel Activation." *Nature Chemical Biology* 17 (2): 196–204.
- Raymond, Clarke R, and Stephen J Redman. 2006. "Spatial Segregation of Neuronal Calcium Signals Encodes Different Forms of LTP in Rat Hippocampus." *The Journal of Physiology* 570 (1): 97–111.
- Reff, Mitchell E, Kristin Carner, Karen S Chambers, Paul C Chinn, John E Leonard, Ron Raab, Roland A Newman, Nabil Hanna, and Darrell R Anderson. 1994. "Depletion of b Cells in Vivo by a Chimeric Mouse Human Monoclonal Antibody to Cd20."
- Reya, Tannishtha, and Hans Clevers. 2005. "Wnt Signalling in Stem Cells and Cancer." *Nature* 434 (7035): 843–50.
- Richardson, Alessandra M, Lauren S Havel, Allyson E Koyen, Jessica M Konen, John Shupe, 4th Wiles WG, W David Martin, et al. 2018. "Vimentin Is Required for Lung Adenocarcinoma Metastasis via Heterotypic Tumor Cell–Cancer-Associated Fibroblast Interactions During Collective Invasion." *Clinical Cancer Research* 24 (2): 420–32.
- Ritchie, Matthew E, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. 2015. "limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7): e47. <https://doi.org/10.1093/nar/gkv007>.
- Rizopoulos, Theodoros, and Martha Assimakopoulou. 2021. "Transient Receptor Potential (TRP) Channels in Human Colorectal Cancer: Evidence and Perspectives." *Histology and Histopathology*, 18308.

-
- Rizzuto, Rosario, Diego De Stefani, Anna Raffaello, and Cristina Mammucari. 2012. "Mitochondria as Sensors and Regulators of Calcium Signalling." *Nature Reviews Molecular Cell Biology* 13 (9): 566–78.
- Robinson, Mark D, Davis J McCarthy, and Gordon K Smyth. 2010. "edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26 (1): 139–40.
- Robinson, Mark D, and Alicia Oshlack. 2010. "A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data." *Genome Biology* 11 (3): 1–9.
- Rocke, David M. 1996. "Robustness Properties of s-Estimators of Multivariate Location and Shape in High Dimension." *The Annals of Statistics*, 1327–45.
- Rodrigues, Tiago, Gabriela Nohemi Nunez Estevez, and Ivarne Luis dos Santos Tersariol. 2019. "Na⁺/Ca²⁺ Exchangers: Unexploited Opportunities for Cancer Therapy?" *Biochemical Pharmacology* 163: 357–61.
- Rong, Yi-Ping, Geert Bultynck, Ademuyiwa S Aromolaran, Fei Zhong, Jan B Parys, Humbert De Smedt, Gregory A Mignery, H Llewelyn Roderick, Martin D Bootman, and Clark W Distelhorst. 2009. "The Bh4 Domain of Bcl-2 Inhibits ER Calcium Release and Apoptosis by Binding the Regulatory and Coupling Domain of the Ip3 Receptor." *Proceedings of the National Academy of Sciences* 106 (34): 14397–402.
- Ruppert, David. 1992. "Computing s Estimators for Regression and Multivariate Location/Dispersion." *Journal of Computational and Graphical Statistics* 1 (3): 253–70.
- Russell, Diane, and Solomon H Snyder. 1968. "Amine Synthesis in Rapidly Growing Tissues: Ornithine Decarboxylase Activity in Regenerating Rat Liver, Chick Embryo, and Various Tumors." *Proceedings of the National Academy of Sciences of the United States of America* 60 (4): 1420.
- Saefken, Benjamin, David Ruegamer, Thomas Kneib, and Sonja Greven. 2018. "Conditional Model Selection in Mixed-Effects Models with cAIC4." *ArXiv e-Prints*.
- Saito, Hajime, Yoshihiro Minamiya, Hiroyuki Watanabe, Naoko Takahashi, Manabu Ito, Hiroshi Toda, Hayato Konno, Masafumi Mitsui, Satoru Motoyama, and Jun-ichi Ogawa. 2011. "Expression of the Transient Receptor Potential Channel C3 Correlates with a Favorable Prognosis in Patients with Adenocarcinoma of the Lung." *Annals of Surgical Oncology* 18 (12): 3377–83.
- Saito, Kengo, Katsuhiko Uzawa, Yosuke Endo, Yoshikuni Kato, Dai Nakashima, Katsunori Ogawara, Masashi Shiba, Hiroki Bukawa, Hidetaka Yokoe, and Hideki Tanzawa. 2006. "Plasma Membrane Ca²⁺ ATPase Isoform 1 down-Regulated in Human Oral Cancer." *Oncology Reports* 15 (1): 49–55.
- Sakakura, Chouhei, Akeo Hagiwara, Kenichiro Fukuda, Katsumi Shimomura, Tsuyoshi Takagi, Shuichi Kin, Yuenn Nakase, et al. 2003. "Possible Involvement of Inositol 1, 4, 5-Trisphosphate

-
- Receptor Type 3 (Ip3r3) in the Peritoneal Dissemination of Gastric Cancers.” *Anticancer Research* 23 (5A): 3691–97.
- Sakamoto, Kathleen M, and David A Frank. 2009. “CREB in the Pathophysiology of Cancer: Implications for Targeting Transcription Factors for Cancer Therapy.” *Clinical Cancer Research* 15 (8): 2583–87.
- Sakhno, LA, VV Sarnatskaya, LA Yushko, EA Snezhkova, KI Bardakhivskaya, OO Shevchuk, AS Sidorenko, and VG Nikolaev. 2019. “Adsorptive Therapy as a Modifier for Tumor-Host Interaction.” *Experimental Oncology* 41 (3): 254–57.
- Salibian-Barrera, Matías, and Víctor J Yohai. 2006. “A Fast Algorithm for s-Regression Estimates.” *Journal of Computational and Graphical Statistics* 15 (2): 414–27.
- Sallinger, Matthias, Sascha Berlansky, and Irene Frischauf. 2020. “Orai Channels: Key Players in Ca²⁺ Homeostasis.” *Current Opinion in Physiology*.
- Samanta, Amrita, Taylor ET Hughes, and Vera Y Moiseenkova-Bell. 2018. “Transient Receptor Potential (TRP) Channels.” *Membrane Protein Complexes: Structure and Function*, 141–65.
- Sandler, Robert S, Susan Halabi, John A Baron, Susan Budinger, Electra Paskett, Roger Keresztes, Nicholas Petrelli, et al. 2003. “A Randomized Trial of Aspirin to Prevent Colorectal Adenomas in Patients with Previous Colorectal Cancer.” *New England Journal of Medicine* 348 (10): 883–90.
- Sas, Leen, Filip Lardon, Peter B Vermeulen, Jan Hauspy, Peter Van Dam, Patrick Pauwels, Luc Y Dirix, and Steven J Van Laere. 2012. “The Interaction Between ER and NFκb in Resistance to Endocrine Therapy.” *Breast Cancer Research* 14 (4): 1–14.
- Sánchez, García-Cuevas, Julio César. 2016. “Función Del Intercambiador Sodio Calcio (NCX) En La Hipoxia Neuronal y La Neuroprotección.” *Revista Mexicana de Neurociencia* 17 (5): 60–69.
- Sánchez, Luciana, Marcos Dellamea, Lorena Moreno, Carlos Osorio, Rodrigo González Toranzo, and Mariano Sosa. 2015. “Destinos Inesperados: Sitios Infrecuentes de Metástasis En El cáncer Colorectal.”
- Schena, Mark, Dari Shalon, Ronald W Davis, and Patrick O Brown. 1995. “Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray.” *Science* 270 (5235): 467–70.
- Schindl, Rainer, Reinhard Fritsch, Isaac Jardin, Irene Frischauf, Heike Kahr, Martin Muik, Maria Christine Riedl, Klaus Groschner, and Christoph Romanin. 2012. “Canonical Transient Receptor Potential (TRPC) 1 Acts as a Negative Regulator for Vanilloid Trpv6-Mediated Ca²⁺ Influx.” *Journal of Biological Chemistry* 287 (42): 35612–20.
- Schlingemann, Jörg. 2005. “Expression Profiling by DNA Microarrays: Development of Amplification Methods for the Analysis of Minimal Tumor Samples.” PhD thesis.
- Schwarz, Gideon. 1978. “Estimating the Dimension of a Model.” *The Annals of Statistics*, 461–64.

-
- Scrideli, Carlos A, Carlos G Carlotti, Oswaldo K Okamoto, Vanessa S Andrade, Maria AA Cortez, Fábio JN Motta, Agda K Lucio-Eterovic, et al. 2008. "Gene Expression Profile Analysis of Primary Glioblastomas and Non-Neoplastic Brain Tissue: Identification of Potential Target Genes by Oligonucleotide Microarray and Real-Time Quantitative PCR." *Journal of Neuro-Oncology* 88 (3): 281–91.
- Shaalan, Abeer Kamal, Guy Carpenter, and Gordon Proctor. 2017. "Measurement of Intracellular Calcium of Submandibular Glands Using a High Throughput Plate Reader." *Journal of Biological Methods* 4 (3).
- Shaked, Yuval, Sandra McAllister, Ofer Fainaru, and Nava Almog. 2014. "Tumor Dormancy and the Angiogenic Switch: Possible Implications of Bone Marrow-Derived Cells." *Current Pharmaceutical Design* 20 (30): 4920–33.
- Shantz, LM, and VA Levin. 2007. "Regulation of Ornithine Decarboxylase During Oncogenic Transformation: Mechanisms and Therapeutic Potential." *Amino Acids* 33 (2): 213–23.
- Sharma, Divya, Gautam Jamra, Uma M Singh, Salej Sood, and Anil Kumar. 2017. "Calcium Biofortification: Three Pronged Molecular Approaches for Dissecting Complex Trait of Calcium Nutrition in Finger Millet (*Eleusine Coracana*) for Devising Strategies of Enrichment of Food Crops." *Frontiers in Plant Science* 7: 2028.
- Shay, Jerry W, and Woodring E Wright. 2011. "Role of Telomeres and Telomerase in Cancer." In *Seminars in Cancer Biology*, 21:349–53. 6. Elsevier.
- Shi, Yu, Xia Ding, ZH He, KC Zhou, Qun Wang, and YZ Wang. 2009. "Critical Role of Trpc6 Channels in G2 Phase Transition and the Development of Human Oesophageal Cancer." *Gut* 58 (11): 1443–50.
- Shibao, Kazunori, Michael J Fiedler, Jun Nagata, Noritaka Minagawa, Keiji Hirata, Yoshifumi Nakayama, Yasuko Iwakiri, Michael H Nathanson, and Koji Yamaguchi. 2010. "The Type III Inositol 1, 4, 5-Trisphosphate Receptor Is Associated with Aggressiveness of Colorectal Carcinoma." *Cell Calcium* 48 (6): 315–23.
- Shieh, Albert D, and Yeung Sam Hung. 2009. "Detecting Outlier Samples in Microarray Data." *Statistical Applications in Genetics and Molecular Biology* 8 (1).
- Shimizu, Shunichi, Nobuaki Takahashi, and Yasuo Mori. 2014. "TRPs as Chemosensors (ROS, RNS, RCS, Gasotransmitters)." *Mammalian Transient Receptor Potential (TRP) Cation Channels*, 767–94.
- Shuttleworth, Trevor J, Jill L Thompson, and Olivier Mignen. 2007. "Stim1 and the Noncapacitative ARC Channels." *Cell Calcium* 42 (2): 183–91.
- Simeone, Timothy A, Russell M Sanchez, and Jong M Rho. 2004. "Molecular Biology and Ontogeny of Glutamate Receptors in the Mammalian Central Nervous System." *Journal of Child Neurology* 19 (5): 343–60.

-
- Smyth, Gordon K. 2004. "Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments." *Statistical Applications in Genetics and Molecular Biology* 3 (1).
- Soboloff, Jonathan, Maria A Spassova, Xiang D Tang, Tamara Hewavitharana, Wen Xu, and Donald L Gill. 2006. "Orai1 and STIM Reconstitute Store-Operated Calcium Channel Function." *Journal of Biological Chemistry* 281 (30): 20661–65.
- Sobradillo, Diego, Miriam Hernández-Morales, Daniel Ubierna, Mary P Moyer, Lucía Núñez, and Carlos Villalobos. 2014. "A Reciprocal Shift in Transient Receptor Potential Channel 1 (Trpc1) and Stromal Interaction Molecule 2 (Stim2) Contributes to Ca²⁺ Remodeling and Cancer Hallmarks in Colorectal Carcinoma Cells." *Journal of Biological Chemistry* 289 (42): 28765–82.
- Sokal, Robert R. 1995. "Biometry: The Principles and Practice of Statistics." *Biological Research*.
- Sozucan, Y, ME Kalender, I Sari, A Suner, S Oztuzcu, K Arman, O Yumrutas, et al. 2015. "TRP Genes Family Expression in Colorectal Cancer." *Experimental Oncology*.
- Srikanth, Sonal, Hea-Jin Jung, Kyun-Do Kim, Puneet Souda, Julian Whitelegge, and Yousang Gwack. 2010. "A Novel EF-Hand Protein, Cracr2a, Is a Cytosolic Ca²⁺ Sensor That Stabilizes CRAC Channels in t Cells." *Nature Cell Biology* 12 (5): 436–46.
- Sritangos, Pishyaporn, Eduardo Pena Alarcon, Andrew D James, Ahlam Sultan, Daniel A Richardson, and Jason IE Bruce. 2020. "Plasma Membrane Ca²⁺ Atpase Isoform 4 (Pmca4) Has an Important Role in Numerous Hallmarks of Pancreatic Cancer." *Cancers* 12 (1): 218.
- Staal, FJT, Mirjam van der Burg, LFA Wessels, BH Barendregt, MRM Baert, CMM Van Den Burg, C Van Huffel, et al. 2003. "DNA Microarrays for Comparison of Gene Expression Profiles Between Diagnosis and Relapse in Precursor-b Acute Lymphoblastic Leukemia: Choice of Technique and Purification Influence the Identification of Potential Diagnostic Markers." *Leukemia* 17 (7): 1324–32.
- Stacker, Steven A, Steven P Williams, Tara Karnezis, Ramin Shayan, Stephen B Fox, and Marc G Achen. 2014. "Lymphangiogenesis and Lymphatic Vessel Remodelling in Cancer." *Nature Reviews Cancer* 14 (3): 159–72.
- Stanisz, Hedwig, Stephanie Saul, Cornelia SL Müller, Reinhard Kappl, Barbara A Niemeyer, Thomas Vogt, Markus Hoth, Alexander Roesch, and Ivan Bogeski. 2014. "Inverse Regulation of Melanoma Growth and Migration by o Rai1/STIM 2-Dependent Calcium Entry." *Pigment Cell & Melanoma Research* 27 (3): 442–53.
- Stathopoulos, Peter B, Le Zheng, Guang-Yao Li, Michael J Plevin, and Mitsuhiko Ikura. 2008. "Structural and Mechanistic Insights into Stim1-Mediated Initiation of Store-Operated Calcium Entry." *Cell* 135 (1): 110–22.
- Stevens, Stanley Smith, and others. 1946. "On the Theory of Scales of Measurement."

-
- Strehler, Emanuel E. 2013. "Plasma Membrane Calcium ATPases as Novel Candidates for Therapeutic Agent Development." *Journal of Pharmacy & Pharmaceutical Sciences: A Publication of the Canadian Society for Pharmaceutical Sciences, Societe Canadienne Des Sciences Pharmaceutiques* 16 (2): 190.
- Strehler, Emanuel E, and David A Zacharias. 2001. "Role of Alternative Splicing in Generating Isoform Diversity Among Plasma Membrane Calcium Pumps." *Physiological Reviews*.
- Stuelten, Christina H, Carole A Parent, and Denise J Montell. 2018. "Cell Motility in Cancer Invasion and Metastasis: Insights from Simple Model Organisms." *Nature Reviews Cancer* 18 (5): 296–312.
- Subramanian, Aravind, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, et al. 2005. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences* 102 (43): 15545–50.
- Sudarsanam, Sucha, and Dale E Johnson. 2010. "Functional Consequences of mTOR Inhibition." *Curr Opin Drug Discov Devel* 13 (1): 31–40.
- Sulk, Mathias, and Martin Steinhoff. 2015. "Role of TRP Channels in Skin Diseases." In *TRP Channels as Therapeutic Targets*, 293–323. Elsevier.
- Sumoza-Toledo, Adriana, Mario Iván Espinoza-Gabriel, and Dvorak Montiel-Condado. 2016. "Evaluation of the Trpm2 Channel as a Biomarker in Breast Cancer Using Public Databases Analysis." *Boletín Médico Del Hospital Infantil de México* 73 (6): 397–404.
- Sun, Jianwei, Fujian Lu, Huifang He, Junling Shen, Jane Messina, Rahel Mathew, Dapeng Wang, et al. 2014. "Stim1-and Orai1-Mediated Ca²⁺ Oscillation Orchestrates Invadopodium Formation and Melanoma Invasion." *Journal of Cell Biology* 207 (4): 535–48.
- Sun, Lixin, Frank L Margolis, Michael T Shipley, and Michael S Lidow. 1998. "Identification of a Long Variant of mRNA Encoding the Nr3 Subunit of the NMDA Receptor: Its Regional Distribution and Developmental Expression in the Rat Brain." *FEBS Letters* 441 (3): 392–96.
- Sun, Suxia, Wenjun Li, He Zhang, Longying Zha, Yong Xue, Xianbo Wu, and Fei Zou. 2012. "Requirement for Store-Operated Calcium Entry in Sodium Butyrate-Induced Apoptosis in Human Colon Cancer Cells." *Bioscience Reports* 32 (1): 83–90.
- Sung, Hyuna, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 2021. "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians* 71 (3): 209–49.
- Suresh, Pavithra, Sarayut Phasuk, and Ingrid Y Liu. 2021. "Modulation of Microglia Activation and Alzheimer's Disease: Cx3 Chemokine Ligand 1/Cx3cr and P2x7r Signaling." *Tzu-Chi Medical Journal* 33 (1): 1.

-
- Tajada, Sendoa, and Carlos Villalobos. 2020. "Calcium Permeable Channels in Cancer Hallmarks." *Frontiers in Pharmacology* 11: 968.
- Takemura, Haruo, AR Hughes, O Thastrup, and JW Putney Jr. 1989. "Activation of Calcium Entry by the Tumor Promoter Thapsigargin in Parotid Acinar Cells: Evidence That an Intracellular Calcium Pool, and Not an Inositol Phosphate, Regulates Calcium Fluxes at the Plasma Membrane." *Journal of Biological Chemistry* 264 (21): 12266–71.
- Tamai, Shoko, Hiroshi Iida, Sadaki Yokota, Tomoko Sayano, Shoko Kiguchiya, Naotada Ishihara, Jun-Ichi Hayashi, Katsuyoshi Mihara, and Toshihiko Oka. 2008. "Characterization of the Mitochondrial Protein Letm1, Which Maintains the Mitochondrial Tubular Shapes and Interacts with the AAA-ATPase Bcs11." *Journal of Cell Science* 121 (15): 2588–2600.
- Tanwar, Jyoti, Samridhhi Arora, and Rajender K Motiani. 2020. "Orai3: Oncochannel with Therapeutic Potential." *Cell Calcium* 90: 102247.
- Tassetto, Matteo, Anna Scialdone, Anna Solini, and Francesco Di Virgilio. 2021. "The P2x7 Receptor: A Promising Pharmacological Target in Diabetic Retinopathy." *International Journal of Molecular Sciences* 22 (13): 7110.
- Taylor, James T, Xiang-Bin Zeng, Jonathan E Pottle, Kevin Lee, Alun R Wang, Stephenie G Yi, Jennifer AS Scruggs, Suresh S Sikka, and Ming Li. 2008. "Calcium Signaling and t-Type Calcium Channels in Cancer Cell Cycling." *World Journal of Gastroenterology: WJG* 14 (32): 4984.
- Territo, Paul R, and Hamideh Zarrinmayeh. 2021. "P2x7 Receptors in Neurodegeneration: Potential Therapeutic Applications from Basic to Clinical Approaches." *Frontiers in Cellular Neuroscience* 15: 83.
- Themistocleous, Sophia C, Andreas Yiallouris, Constantinos Tsioutis, Apostolos Zaravinos, Elizabeth O Johnson, and Ioannis Patrikios. 2021. "Clinical Significance of p-Class Pumps in Cancer." *Oncology Letters* 22 (3): 1–18.
- Thompson, Jill L, and Trevor J Shuttleworth. 2013. "Exploring the Unique Features of the ARC Channel, a Store-Independent Orai Channel." *Channels* 7 (5): 364–73.
- Tian, Lu, Steven A Greenberg, Sek Won Kong, Josiah Altschuler, Isaac S Kohane, and Peter J Park. 2005. "Discovering Statistically Significant Pathways in Expression Profiling Studies." *Proceedings of the National Academy of Sciences* 102 (38): 13544–49.
- Tian, Wenjuan, Shanshan Liu, and Burong Li. 2019. "Potential Role of Exosomes in Cancer Metastasis." *BioMed Research International* 2019.
- Tiffner, Adéla, Lena Maltan, Sarah Weiß, and Isabella Derler. 2021. "The Orai Pore Opening Mechanism." *International Journal of Molecular Sciences* 22 (2): 533.
- Toaldo, Cristina, Stefania Pizzimenti, Angelo Cerbone, Piergiorgio Pettazoni, Elisa Menegatti, Berardi Daniela, Rosalba Minelli, et al. 2010. "PPAR γ Ligands Inhibit Telomerase Activity

-
- and hTERT Expression Through Modulation of the Myc/Mad/Max Network in Colon Cancer Cells.” *Journal of Cellular and Molecular Medicine* 14 (6a): 1347–57.
- Todorov, Valentin, and Peter Filzmoser. 2009. “An Object-Oriented Framework for Robust Multivariate Analysis.” *Journal of Statistical Software* 32 (3): 1–47. <https://www.jstatsoft.org/article/view/v032i03>.
- Toyota, Minoru, Coty Ho, Mutsumi Ohe-Toyota, Stephen B Baylin, and Jean-Pierre J Issa. 1999. “Inactivation of Cacna1g, a t-Type Calcium Channel Gene, by Aberrant Methylation of Its 5' CpG Island in Human Tumors.” *Cancer Research* 59 (18): 4535–41.
- Trebak, Mohamed, Gary St J Bird, Richard R McKay, and James W Putney Jr. 2002. “Comparison of Human Trpc3 Channels in Receptor-Activated and Store-Operated Modes: Differential Sensitivity to Channel Blockers Suggests Fundamental Differences in Channel Composition.” *Journal of Biological Chemistry* 277 (24): 21617–23.
- Tsavalier, Larisa, Michael H Shapero, Stan Morkowski, and Reiner Laus. 2001. “Trp-P8, a Novel Prostate-Specific Gene, Is up-Regulated in Prostate Cancer and Other Malignancies and Shares High Homology with Transient Receptor Potential Calcium Channel Proteins.” *Cancer Research* 61 (9): 3760–69.
- Tsunoda, Toshiyuki, Hirofumi Koga, Akira Yokomizo, Katsunori Tatsugami, Masatoshi Eto, Junichi Inokuchi, Akira Hirata, Katsuaki Masuda, Koji Okumura, and Seiji Naito. 2005. “Inositol 1, 4, 5-Trisphosphate (IP 3) Receptor Type1 (IP 3 R1) Modulates the Acquisition of Cisplatin Resistance in Bladder Cancer Cell Lines.” *Oncogene* 24 (8): 1396–1402.
- Tyson, John R, and Terrance P Snutch. 2013. “Molecular Nature of Voltage-Gated Calcium Channels: Structure and Species Comparison.” *Wiley Interdisciplinary Reviews: Membrane Transport and Signaling* 2 (5): 181–206.
- Umemura, Masanari, Erdene Baljinnyam, Stefan Feske, Mariana S De Lorenzo, Lai-Hua Xie, Xianfeng Feng, Kayoko Oda, et al. 2014. “Store-Operated Ca²⁺ Entry (SOCE) Regulates Melanoma Proliferation and Cell Migration.” *PloS One* 9 (2): e89292.
- Vacher, Pierre, Anne-Marie Vacher, Raphael Pineau, Simon Latour, Isabelle Soubeyran, Celine Pangault, Karin Tarte, Pierre Soubeyran, Thomas Ducret, and Laurence Bresson-Bepoldin. 2015. “Localized Store-Operated Calcium Influx Represses Cd95-Dependent Apoptotic Effects of Rituximab in Non-Hodgkin b Lymphomas.” *The Journal of Immunology* 195 (5): 2207–15.
- Valero, Ruth A, Laura Senovilla, Lucía Núñez, and Carlos Villalobos. 2008. “The Role of Mitochondrial Potential in Control of Calcium Signals Involved in Cell Proliferation.” *Cell Calcium* 44 (3): 259–69.
- Van Blarigan, Erin L, Charles S Fuchs, Donna Niedzwiecki, Sui Zhang, Leonard B Saltz, Robert J Mayer, Rex B Mowat, et al. 2018. “Association of Survival with Adherence to the American Cancer Society Nutrition and Physical Activity Guidelines for Cancer Survivors After Colon Cancer Diagnosis: The CALGB 89803/Alliance Trial.” *JAMA Oncology* 4 (6): 783–90.

-
- Vander Ark, Alexandra, Jingchen Cao, and Xiaohong Li. 2018. "TGF- β Receptors: In and Beyond TGF- β Signaling." *Cellular Signalling* 52: 112–20.
- Varughese, Joyce T, Susan K Buchanan, and Ashley S Pitt. 2021. "The Role of Voltage-Dependent Anion Channel in Mitochondrial Dysfunction and Human Disease." *Cells* 10 (7): 1737.
- Vasen, Hans FA, Ian Tomlinson, and Antoni Castells. 2015. "Clinical Management of Hereditary Colorectal Cancer Syndromes." *Nature Reviews Gastroenterology & Hepatology* 12 (2): 88–97.
- Vazquez, Guillermo, Jean-Philippe Lievremont, Gary St J Bird, and James W Putney. 2001. "Human Trp3 Forms Both Inositol Trisphosphate Receptor-Dependent and Receptor-Independent Store-Operated Cation Channels in Dt40 Avian b Lymphocytes." *Proceedings of the National Academy of Sciences* 98 (20): 11777–82.
- Várnai, Péter, Balázs Tóth, Dániel J Tóth, László Hunyady, and Tamas Balla. 2007. "Visualization and Manipulation of Plasma Membrane-Endoplasmic Reticulum Contact Sites Indicates the Presence of Additional Molecular Components Within the Stim1-Orai1 Complex* \blacklozenge ." *Journal of Biological Chemistry* 282 (40): 29678–90.
- Vásquez, Karla Yohannessen, and Carlos Ubilla Pérez. 2016. "Diseños de Estudio Clínico-Epidemiológicos." *Rev Ped Elec* 13.
- Veliceasa, Dorina, Marina Ivanovic, Frank Thilo-Schulze Hoepfner, Praveen Thumbikat, Olga V Volpert, and Norm D Smith. 2007. "Transient Potential Receptor Channel 4 Controls Thrombospondin-1 Secretion and Angiogenesis in Renal Cell Carcinoma." *The FEBS Journal* 274 (24): 6365–77.
- Venkatachalam, Kartik, Hong-Tao Ma, Diana L Ford, and Donald L Gill. 2001. "Expression of Functional Receptor-Coupled Trpc3 Channels in Dt40 Triple Receptor InsP3 Knockout Cells." *Journal of Biological Chemistry* 276 (36): 33980–85.
- Venkatachalam, Kartik, and Craig Montell. 2007. "TRP Channels." *Annu. Rev. Biochem.* 76: 387–417.
- Verbe, Bruyere, M. A. P. 2016. "Expression of Store-Operated Channel Components in Prostate Cancer: The Prognostic Paradox." *Human Pathology*, no. 49: 77–82.
- Villalobos, Carlos, Lucía G Gutiérrez, Miriam Hernández-Morales, David Del Bosque, and Lucía Núñez. 2018. "Mitochondrial Control of Store-Operated Ca²⁺ Channels in Cancer: Pharmacological Implications." *Pharmacological Research* 135: 136–43.
- Villalobos, Carlos, Miriam Hernández-Morales, Lucía G Gutiérrez, and Lucía Núñez. 2019. "Trpc1 and Orai1 Channels in Colon Cancer." *Cell Calcium* 81: 59–66.
- Villalobos, Carlos, Diego Sobradillo, Miriam Hernández-Morales, and Lucía Núñez. 2017. "Calcium Remodeling in Colorectal Cancer." *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* 1864 (6): 843–49.

-
- Wallace, Heather M, Yi Huang, Laurence J Marton, Patrick M Woster, and Robert A Casero. 2009. "Polyamine Analogues Targeting Epigenetic Gene Regulation." *Essays in Biochemistry* 46: 95–110.
- Wang, Chih-Yang, Ming-Derg Lai, Nam Nhut Phan, Zhengda Sun, and Yen-Chang Lin. 2015. "Meta-Analysis of Public Microarray Datasets Reveals Voltage-Gated Calcium Gene Signatures in Clinical Cancer Patients." *PloS One* 10 (7): e0125766.
- Wang, Jaw-Yuan, Ben-Kuen Chen, Yu-Shiuan Wang, Yao-Ting Tsai, Wei-Chiao Chen, Wen-Chang Chang, Ming-Feng Hou, Yang-Chang Wu, and Wei-Chiao Chang. 2012. "Involvement of Store-Operated Calcium Signaling in EGF-Mediated COX-2 Gene Activation in Cancer Cells." *Cellular Signalling* 24 (1): 162–69.
- Wang, Jaw-Yuan, Jianwei Sun, Ming-Yii Huang, Yu-Shiuan Wang, Ming-Feng Hou, Yan Sun, Huifang He, et al. 2015. "Stim1 Overexpression Promotes Colorectal Cancer Progression, Cell Motility and COX-2 Expression." *Oncogene* 34 (33): 4358–67.
- Wang, Teng, Kuan Ning, Ting-Xun Lu, and Dong Hua. 2017. "Elevated Expression of TrpC5 and Glut1 Is Associated with Chemoresistance in Colorectal Cancer." *Oncology Reports* 37 (2): 1059–65.
- Wang, Xi-Tao, Yasushi Nagaba, Heide S Cross, Fritz Wrba, Lin Zhang, and Sandra E Guggino. 2000. "The mRNA of I-Type Calcium Channel Elevated in Colon Cancer: Protein Distribution in Normal and Cancerous Colon." *The American Journal of Pathology* 157 (5): 1549–62.
- Wang, Yifan, Jian Shi, Kequn Chai, Xuhua Ying, and Binhua P Zhou. 2013. "The Role of Snail in EMT and Tumorigenesis." *Current Cancer Drug Targets* 13 (9): 963–72.
- Wang, Youjun, Xiaoxiang Deng, Salvatore Mancarella, Eunan Hendron, Satoru Eguchi, Jonathan Soboloff, Xiang D Tang, and Donald L Gill. 2010. "The Calcium Store Sensor, Stim1, Reciprocally Controls Orai and CaV1. 2 Channels." *Science* 330 (6000): 105–9.
- Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. "RNA-Seq: A Revolutionary Tool for Transcriptomics." *Nature Reviews Genetics* 10 (1): 57–63.
- Warburg, Otto, Karl Posener, and E Negelein. 1924. "Über Den Stoffwechsel Der Carcinomzelle." *Naturwissenschaften* 12 (50): 1131–37.
- Weinberg, Robert A. 2013. *The Biology of Cancer*. Garland science.
- White, Eileen, and Robert S DiPaola. 2009. "The Double-Edged Sword of Autophagy Modulation in Cancer." *Clinical Cancer Research* 15 (17): 5308–16.
- White, Halbert, and others. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–38.
- Whitfield, James F. 1997. *Calcium: Cell Cycle Driver, Differentiator and Killer*. Landes Bioscience.
- Whitfield, James F, and Balu Chakravarthy. 2001. *Calcium: The Grand-Master Cell Signaler*. NRC Research Press.

-
- Willard, Stacey S, and Shahriar Koochekpour. 2013. "Glutamate, Glutamate Receptors, and Downstream Signaling Pathways." *International Journal of Biological Sciences* 9 (9): 948.
- Williams, Richard T, Shehnaaz SM Manji, Nigel J Parker, Manuela S Hancock, Leonie van Stekelenburg, Jean-Pierre Eid, Paul V Senior, et al. 2001. "Identification and Characterization of the STIM (Stromal Interaction Molecule) Gene Family: Coding for a Novel Class of Transmembrane Proteins." *Biochemical Journal* 357 (3): 673–85.
- Wilson, Lesley A, Lynn McKeown, Sarka Tumova, Jing Li, and David J Beech. 2015. "Expression of a Long Variant of Cracr2a That Belongs to the Rab GTPase Protein Family in Endothelial Cells." *Biochemical and Biophysical Research Communications* 456 (1): 398–402.
- Witsch, Esther, Michael Sela, and Yosef Yarden. 2010. "Roles for Growth Factors in Cancer Progression." *Physiology*.
- Woods, Joshua J, Neeharika Nemani, Santhanam Shanmughapriya, Akshay Kumar, MengQi Zhang, Sarah R Nathan, Manfred Thomas, et al. 2019. "A Selective and Cell-Permeable Mitochondrial Calcium Uniporter (MCU) Inhibitor Preserves Mitochondrial Bioenergetics After Hypoxia/Reoxygenation Injury." *ACS Central Science* 5 (1): 153–66.
- Woolf, Eric C, Nelofer Syed, and Adrienne C Scheck. 2016. "Tumor Metabolism, the Ketogenic Diet and β -Hydroxybutyrate: Novel Approaches to Adjuvant Brain Tumor Therapy." *Frontiers in Molecular Neuroscience* 9: 122.
- Wu, Dianqing, and Weijun Pan. 2010. "Gsk3: A Multifaceted Kinase in Wnt Signaling." *Trends in Biochemical Sciences* 35 (3): 161–68.
- Wu, Di, and Gordon K Smyth. 2012. "Camera: A Competitive Gene Set Test Accounting for Inter-Gene Correlation." *Nucleic Acids Research* 40 (17): e133–33.
- Xia, Longzheng, Shiming Tan, Yujuan Zhou, Jingguan Lin, Heran Wang, Linda Oyang, Yutong Tian, et al. 2018. "Role of the NF κ b-Signaling Pathway in Cancer." *Oncotargets and Therapy* 11: 2063.
- Xiao, Yan, Xiaoyan Lv, Ge Cao, Guohui Bian, Jingjing Duan, Jianzhong Ai, Huan Sun, et al. 2010. "Overexpression of Trpp5 Contributes to Cell Proliferation and Apoptosis Probably Through Involving Calcium Homeostasis." *Molecular and Cellular Biochemistry* 339 (1): 155–61.
- Xie, Jiansheng, Hongming Pan, Junlin Yao, Yubin Zhou, and Weidong Han. 2016. "SOCE and Cancer: Recent Progress and New Perspectives." *International Journal of Cancer* 138 (9): 2067–77.
- Xu, Dawei, Nikita Popov, Mi Hou, Qian Wang, Magnus Björkholm, Astrid Gruber, Annette R Menkel, and Marie Henriksson. 2001. "Switch from Myc/Max to Mad1/Max Binding and Decrease in Histone Acetylation at the Telomerase Reverse Transcriptase Promoter During Differentiation of H160 Cells." *Proceedings of the National Academy of Sciences* 98 (7): 3826–31.

-
- Xu, Jingyu, Bei Ji, Guorong Wen, Yuan Yang, Hai Jin, Xuemei Liu, Rui Xie, et al. 2016. “Na⁺/h⁺ Exchanger 1, Na⁺/Ca²⁺ Exchanger 1 and Calmodulin Complex Regulates Interleukin 6-Mediated Cellular Behavior of Human Hepatocellular Carcinoma.” *Carcinogenesis* 37 (3): 290–300.
- Xu, Jingyu, Yuan Yang, Rui Xie, Jilong Liu, Xubiao Nie, Jiaying An, Guorong Wen, Xuemei Liu, Hai Jin, and Biguang Tuo. 2018. “The Ncx1/Trpc6 Complex Mediates TGF β -Driven Migration and Invasion of Human Hepatocellular Carcinoma Cells.” *Cancer Research* 78 (10): 2564–76.
- Yamamoto, Hiroyuki, and Kohzoh Imai. 2015. “Microsatellite Instability: An Update.” *Archives of Toxicology* 89 (6): 899–921.
- Yamamoto, Taizo, Yutaka Seino, Hirofumi Fukumoto, Gyohan Koh, Hideki Yano, Nobuya Inagaki, Yuichiro Yamada, Kazutomo Inoue, Tadao Manabe, and Hiroo Imura. 1990. “Over-Expression of Facilitative Glucose Transporter Genes in Human Cancer.” *Biochemical and Biophysical Research Communications* 170 (1): 223–30.
- Yang, Dongki, and Jaehong Kim. 2020. “Emerging Role of Transient Receptor Potential (TRP) Channels in Cancer Progression.” *BMB Reports* 53 (3): 125.
- Yang, Ning, Ying Tang, Fang Wang, Haibin Zhang, Dan Xu, Yafeng Shen, Shuhan Sun, and Guangshun Yang. 2013. “Blockade of Store-Operated Ca²⁺ Entry Inhibits Hepatocarcinoma Cell Migration and Invasion by Regulating Focal Adhesion Turnover.” *Cancer Letters* 330 (2): 163–69.
- Yang, Shengyu, J Jillian Zhang, and Xin-Yun Huang. 2009. “Orai1 and Stim1 Are Critical for Breast Tumor Cell Migration and Metastasis.” *Cancer Cell* 15 (2): 124–34.
- Yang, SL, Q Cao, KC Zhou, YJ Feng, and YZ Wang. 2009. “Transient Receptor Potential Channel C3 Contributes to the Progression of Human Ovarian Cancer.” *Oncogene* 28 (10): 1320–28.
- Yee, Nelson S, Weiqiang Zhou, and Minsun Lee. 2010. “Transient Receptor Potential Channel Trpm8 Is over-Expressed and Required for Cellular Proliferation in Pancreatic Adenocarcinoma.” *Cancer Letters* 297 (1): 49–55.
- Yerushalmi, Hagit F, David G Besselsen, Natalia A Ignatenko, Karen A Blohm-Mangone, Jose L Padilla-Torres, David E Stringer, Jose M Guillen, Hana Holubec, Claire M Payne, and Eugene W Gerner. 2006. “Role of Polyamines in Arginine-Dependent Colon Carcinogenesis in ApcMin/+ Mice.” *Molecular Carcinogenesis: Published in Cooperation with the University of Texas MD Anderson Cancer Center* 45 (10): 764–73.
- Yeung, Priscilla S-W, Megumi Yamashita, and Murali Prakriya. 2017. “Pore Opening Mechanism of CRAC Channels.” *Cell Calcium* 63: 14–19.
- You, Dewen, Junping Xin, Andrew Volk, Wei Wei, Rachel Schmidt, Gina Scurti, Sucha Nand, et al. 2015. “FAK Mediates a Compensatory Survival Signal Parallel to Pi3k-AKT in PTEN-Null t-ALL Cells.” *Cell Reports* 10 (12): 2055–68.

-
- Yuan, Joseph P, Weizhong Zeng, Guo N Huang, Paul F Worley, and Shmuel Muallem. 2007. “Stim1 Heteromultimerizes TRPC Channels to Determine Their Function as Store-Operated Channels.” *Nature Cell Biology* 9 (6): 636–45.
- Yuan, Qing, Ramesh M Ray, Mary Jane Viar, and Leonard R Johnson. 2001. “Polyamine Regulation of Ornithine Decarboxylase and Its Antizyme in Intestinal Epithelial Cells.” *American Journal of Physiology-Gastrointestinal and Liver Physiology* 280 (1): G130–38.
- Zagranichnaya, Tatiana K, Xiaoyan Wu, and Mitchel L Villereal. 2005. “Endogenous Trpc1, Trpc3, and Trpc7 Proteins Combine to Form Native Store-Operated Channels in HEK-293 Cells.” *Journal of Biological Chemistry* 280 (33): 29559–69.
- Zeileis, Achim. 2004. “Econometric Computing with HC and HAC Covariance Matrix Estimators.”
- Zeki, Sebastian S, Trevor A Graham, and Nicholas A Wright. 2011. “Stem Cells and Their Implications for Colorectal Cancer.” *Nature Reviews Gastroenterology & Hepatology* 8 (2): 90–100.
- Zell, Jason A, Daniel Pelot, Wen-Pin Chen, Christine E McLaren, Eugene W Gerner, and Frank L Meyskens. 2009. “Risk of Cardiovascular Events in a Randomized Placebo-Controlled, Double-Blind Trial of Difluoromethylornithine Plus Sulindac for the Prevention of Sporadic Colorectal Adenomas.” *Cancer Prevention Research* 2 (3): 209–12.
- Zeng, Bo, Cunzhong Yuan, Xingsheng Yang, Stephen L Atkin, and Shang-Zhong Xu. 2013. “TRPC Channels and Their Splice Variants Are Essential for Promoting Human Ovarian Cancer Cell Proliferation and Tumorigenesis.” *Current Cancer Drug Targets* 13 (1): 103–16.
- Zeng, X, SC Sikka, L Huang, C Sun, C Xu, D Jia, AB Abdel-Mageed, JE Pottle, JT Taylor, and M Li. 2010. “Novel Role for the Transient Receptor Potential Channel Trpm2 in Prostate Cancer Cell Proliferation.” *Prostate Cancer and Prostatic Diseases* 13 (2): 195–201.
- Zhan, Zheng-Yu, Lu-Xing Zhong, Miao Feng, Jian-Feng Wang, Di-Bin Liu, and Jian-Ping Xiong. 2015. “Over-Expression of Orai1 Mediates Cell Proliferation and Associates with Poor Prognosis in Human Non-Small Cell Lung Carcinoma.” *International Journal of Clinical and Experimental Pathology* 8 (5): 5080.
- Zhang, Lin, Wei Zhou, Victor E Velculescu, Scott E Kern, Ralph H Hruban, Stanley R Hamilton, Bert Vogelstein, and Kenneth W Kinzler. 1997. “Gene Expression Profiles in Normal and Cancer Cells.” *Science* 276 (5316): 1268–72.
- Zhang, Shenyuan L, J Ashot Kozak, Weihua Jiang, Andriy V Yeromin, Jing Chen, Ying Yu, Aubin Penna, Wei Shen, Victor Chi, and Michael D Cahalan. 2008. “Store-Dependent and-Independent Modes Regulating Ca²⁺ Release-Activated Ca²⁺ Channel Activity of Human Orai1 and Orai3.” *Journal of Biological Chemistry* 283 (25): 17662–71.
- Zhang, Songbai, Nicolas Fritz, Cristian Ibarra, and Per Uhlén. 2011. “Inositol 1, 4, 5-Trisphosphate Receptor Subtype-Specific Regulation of Calcium Oscillations.” *Neurochemical Research* 36 (7): 1175–85.

-
- Zhang, Xuexin, Ping Xin, Ryan E Yoast, Scott M Emrich, Martin T Johnson, Trayambak Pathak, J Cory Benson, et al. 2020. “Distinct Pharmacological Profiles of Orai1, Orai2, and Orai3 Channels.” *Cell Calcium* 91: 102281.
- Zheng, Le, Peter B Stathopoulos, Rainer Schindl, Guang-Yao Li, Christoph Romanin, and Mitsuhiro Ikura. 2011. “Auto-Inhibitory Role of the EF-SAM Domain of STIM Proteins in Store-Operated Calcium Entry.” *Proceedings of the National Academy of Sciences* 108 (4): 1337–42.
- Zhou, Kun, Shui-Shen Zhang, Yan Yan, and Song Zhao. 2014. “Overexpression of Transient Receptor Potential Vanilloid 2 Is Associated with Poor Prognosis in Patients with Esophageal Squamous Cell Carcinoma.” *Medical Oncology* 31 (7): 17.
- Zhou, Yandong, Xizhuo Wang, Xianming Wang, Natalia A Loktionova, Xiangyu Cai, Robert M Nwokonko, Erin Vrana, Youjun Wang, Brad S Rothberg, and Donald L Gill. 2015. “Stim1 Dimers Undergo Unimolecular Coupling to Activate Orai1 Channels.” *Nature Communications* 6 (1): 1–10.
- Zhu, Mu, and Ali Ghodsi. 2006. “Automatic Dimensionality Selection from the Scree Plot via the Use of Profile Likelihood.” *Computational Statistics & Data Analysis* 51 (2): 918–30.
- Zhuang, Liyan, Ji-Bin Peng, Liqiang Tou, Hitomi Takanaga, Rosalyn M Adam, Matthias A Hediger, and Michael R Freeman. 2002. “Calcium-Selective Ion Channel, CaT1, Is Apically Localized in Gastrointestinal Tract Epithelia and Is Aberrantly Expressed in Human Malignancies.” *Laboratory Investigation* 82 (12): 1755–64.
- Zijl, Franziska van, Georg Krupitza, and Wolfgang Mikulits. 2011. “Initial Steps of Metastasis: Cell Invasion and Endothelial Transmigration.” *Mutation Research/Reviews in Mutation Research* 728 (1-2): 23–34.