



Universidad de Valladolid



PROGRAMA DE DOCTORADO EN INVESTIGACIÓN
BIOMÉDICA

TESIS DOCTORAL:

**Alteración de los patrones de *splicing* en genes de
susceptibilidad a cáncer de mama: análisis
funcional mediante minigenes híbridos**

Presentada por Lara Sanoguera Miralles para optar al
grado de Doctora por la Universidad de Valladolid

Dirigida por:
Dr. Eladio Andrés Velasco Sampedro

En este párrafo se agradece a los organismos e instituciones implicadas en la financiación de este trabajo. En concreto, al proyecto europeo BRIDGES (ID del proyecto 634935), al instituto de Salud Carlos III, Ministerio de Economía y Competitividad (ID de los proyectos PI17/00227 y PI20/00225), y a la Consejería de Educación de la Junta de Castilla y León (ID del proyecto CSI242P18). Además, para la realización de esta tesis doctoral, la doctoranda, Lara Sanoguera Miralles, ha contado con la financiación de la Ayuda Predoctoral de la Asociación Española contra el Cáncer (2019-2023), la ayuda ERASMUS+ Prácticas (curso académico 2021/2022) que convoca anualmente el Vicerrectorado de Internacionalización de la Universidad de Valladolid, y la ayuda para estancias breves en el desarrollo de tesis doctorales de la Universidad de Valladolid 2022.

PRODUCTIVIDAD CIENTÍFICA

Gailite, L., Valenzuela-Palomo, A., **Sanoguera-Miralles, L.**, Rots, D., Kreile, M., & Velasco, E. A. (2020). UGT1A1 Variants c.864+5G<T and c.996+2_996+5del of a Crigler-Najjar Patient Induce Aberrant Splicing in Minigene Assays. *Frontiers in Genetics*, 11, 169.

Sanoguera-Miralles, L., Valenzuela-Palomo, A., Bueno-Martínez, E., Llovet, P., Díez-Gómez, B., Caloca, M. J., Pérez-Segura, P., Fraile-Bethencourt, E., Colmena, M., Carvalho, S., Allen, J., Easton, D. F., Devilee, P., Vreeswijk, M. P. G., de la Hoya, M., & Velasco, E. A. (2020). Comprehensive Functional Characterization and Clinical Interpretation of 20 Splice-Site Variants of the RAD51C Gene. *Cancers*, 12(12), 3771.

Bueno-Martínez, E., **Sanoguera-Miralles, L.**, Valenzuela-Palomo, A., Lorca, V., Gómez-Sanz, A., Carvalho, S., Allen, J., Infante, M., Pérez-Segura, P., Lázaro, C., Easton, D. F., Devilee, P., Vreeswijk, M. P. G., de la Hoya, M., & Velasco, E. A. (2021). RAD51D Aberrant Splicing in Breast Cancer: Identification of Splicing Regulatory Elements and Minigene-Based Evaluation of 53 DNA Variants. *Cancers*, 13(11), 2845.

Valenzuela-Palomo, A., Bueno-Martínez, E., **Sanoguera-Miralles, L.**, Lorca, V., Fraile-Bethencourt, E., Esteban-Sánchez, A., Gómez-Barrero, S., Carvalho, S., Allen, J., García-Álvarez, A., Pérez-Segura, P., Dorling, L., Easton, D. F., Devilee, P., Vreeswijk, M. P. G., de la Hoya, M., & Velasco, E. A. (2022). Splicing predictions, minigene analyses, and ACMG-AMP clinical classification of 42 germline PALB2 splice-site variants. *The Journal of Pathology*, 256(3), 321–334.

Zhu, L., Miao, B., Dymerska, D., Kuswik, M., Bueno-Martínez, E., **Sanoguera-Miralles, L.**, Velasco, E. A., Paramasivam, N., Schlesner, M., Kumar, A., Yuan, Y., Lubinski, J., Bandapalli, O. R., Hemminki, K., & Försti, A. (2022). Germline Variants of CYBA and TRPM4 Predispose to Familial Colorectal Cancer. *Cancers*, 14(3), 670.

Sanoguera-Miralles, L., Bueno-Martínez, E., Valenzuela-Palomo, A., Esteban-Sánchez, A., Llinares-Burguet, I., Pérez-Segura, P., García-Álvarez, A., Hoya, M. de la, & Velasco-Sampedro, E. A. (2022). Minigene Splicing Assays Identify 20 Spliceogenic Variants of the Breast/Ovarian Cancer Susceptibility Gene RAD51C. *Cancers*, 14(12), 2960.

Bueno-Martínez, E., **Sanoguera-Miralles, L.**, Valenzuela-Palomo, A., Esteban-Sánchez, A., Lorca, V., Llinares-Burguet, I., Allen, J., García-Álvarez, A., Pérez-Segura, P., Durán, M.,

Easton, D. F., Devilee, P., Vreeswijk, M. P. G., de la Hoya, M., & Velasco-Sampedro, E. A. (2022). Minigene-based splicing analysis and ACMG/AMP-based tentative classification of 56 ATM variants. *The Journal of Pathology*, 258(1), 83–101.

Valenzuela-Palomo, A., **Sanoguera-Miralles, L.**, Bueno-Martínez, E., Esteban-Sánchez, A., Llinares-Burguet, I., García-Álvarez, A., Pérez-Segura, P., Gómez-Barrero, S., de la Hoya, M., & Velasco-Sampedro, E. A. (2022). Splicing Analysis of 16 PALB2 ClinVar Variants by Minigene Assays: Identification of Six Likely Pathogenic Variants. *Cancers*, 14(18), 4541.

Sanoguera-Miralles, L., Valenzuela-Palomo, A., Bueno-Martínez, E., Esteban-Sánchez, A., Lorca, V., Llinares-Burguet, I., García-Álvarez, A., Pérez-Segura, P., Infante, M., Easton, D. F., Devilee, P., Vreeswijk, M. P. G., de la Hoya, M., & Velasco-Sampedro, E. A. (2023). Systematic minigene-based splicing analysis and tentative clinical classification of 52 CHEK2 splice-site variants. *Clinical Chemistry*.

ÍNDICE

ÍNDICE DE FIGURAS.....	IX
ÍNDICE DE TABLAS	XI
ÍNDICE DEL ANEXO	XIII
LISTADO DE ABREVIATURAS.....	XV
LISTADO DE GENES	XVII
RESUMEN	XXI
INTRODUCCIÓN.....	1
<i>SPLICING</i>	3
1.1. ESPLICEOSOMA Y REACCIÓN DE <i>SPLICING</i>	6
1.2. SITIOS DE <i>SPLICING</i> CANÓNICOS Y ATÍPICOS	9
1.3. REGULACIÓN DEL <i>SPLICING</i>	11
1.4. <i>SPLICING</i> ALTERNATIVO	15
1.5. ALTERACIONES DEL PROCESO DE <i>SPLICING</i> Y ENFERMEDAD	17
1.6. MÉTODOS USADOS PARA EL ANÁLISIS DEL <i>SPLICING</i> ABERRANTE.....	19
CÁNCER DE MAMA.....	21
2. GENERALIDADES DEL CÁNCER DE MAMA.....	23
2.1. CÁNCER DE MAMA HEREDITARIO.....	26
2.2. MANEJO CLÍNICO DE PACIENTES CON CÁNCER DE MAMA HEREDITARIO	29
2.3. GENES DE SUSCEPTIBILIDAD A CÁNCER DE MAMA/OVARIO INVESTIGADOS	
.....	32
▪ <i>RAD51C</i>	33
▪ <i>CHEK2</i>	36
▪ <i>BARD1</i>	38
▪ <i>BRIP1</i>	39
2.4. <i>SPLICING</i> ABERRANTE EN CÁNCER DE MAMA HEREDITARIO.....	41
HIPÓTESIS Y OBJETIVOS.....	43
HIPÓTESIS.....	45
OBJETIVOS.....	47

RESULTADOS	49
<i>RAD51C</i>	51
1. ARTÍCULO 1.....	51
2. ARTÍCULO 2.....	53
<i>CHEK2</i>	55
1. ARTÍCULO 3.....	55
2. ESTUDIO DE REGULACIÓN DE LOS EXONES 8 Y 10 DE <i>CHEK2</i>	59
2.1. MAPEO FUNCIONAL DE INTERVALOS RICOS EN ESE.....	59
2.2. ESTUDIO FUNCIONAL DE VARIANTES SRE.....	63
2.3. INTERPRETACIÓN CLÍNICA DE VARIANTES SIGUIENDO LAS NORMAS ACMG/AMP.....	71
<i>BARD1</i>	75
1. CONSTRUCCIÓN Y VALIDACIÓN DEL MINIGÉN <i>WILD TYPE</i>	75
2. MAPEO FUNCIONAL DE LOS MOTIVOS REGULADORES.....	77
3. MOTIVOS SILENCIADORES PUTATIVOS PRÓXIMOS A LOS DONADORES CRÍPTICOS.....	80
<i>BRIP1</i>	83
1. CONSTRUCCIÓN Y VALIDACIÓN DEL MINIGÉN <i>WILD TYPE</i>	83
2. MAPEO FUNCIONAL DE LOS MOTIVOS REGULADORES.....	84
DISCUSIÓN	87
1. VARIANTES DE SITIOS DE <i>SPLICING</i>	93
2. ELEMENTOS REGULADORES DEL <i>SPLICING</i> Y VARIANTES SRE.....	109
CONCLUSIONES	117
BIBLIOGRAFÍA	121
ANEXO	145

ÍNDICE DE FIGURAS

Figura 1. Micrografía electrónica de la hibridación del ARNm procedente del gen <i>hexón</i> con un fragmento de ADN de adenovirus	4
Figura 2. Secuencias consenso de <i>splicing</i> de un intrón U2-dependiente	5
Figura 3. La reacción de <i>splicing</i>	8
Figura 4. Secuencias consenso de <i>splicing</i> de un intrón U12-dependiente	10
Figura 5. Sitios de <i>splicing</i> no-canónicos.....	11
Figura 6. Secuencias en <i>cis</i> y factores en <i>trans</i> que participan en la regulación del proceso de <i>splicing</i>	12
Figura 7. Estructura de las proteínas SR y hnRNP humanas.....	14
Figura 8. Eventos alternativos de <i>splicing</i>	17
Figura 9. Protocolo del estudio funcional de variantes candidatas de alterar el <i>splicing</i> mediante minigenes híbridos.....	20
Figura 10. Recortes de pinturas renacentistas que proporcionan representaciones de patologías mamarias malignas	22
Figura 11. Gráficos de los datos de A) incidencia y B) mortalidad por tipo de cáncer a nivel mundial del año 2020	23
Figura 12. Gráficos de los datos de A) incidencia y B) mortalidad por tipo de cáncer en la población femenina a nivel mundial del año 2020	24
Figura 13. Árbol genealógico de Broca	27
Figura 14. Representación esquemática de los genes de susceptibilidad a cáncer de mama que participan en la vía de reparación de DSB por HR	29
Figura 15. Resumen de las principales rutas moleculares de reparación de daños en el ADN.....	35
Figura 16. Mapeo de los intervalos ricos en SRE de los exones 8 y 10 de <i>CHEK2</i>	61
Figura 17. Ensayos de <i>splicing</i> de las variantes espliceogénicas del exón 8 de <i>CHEK2</i>	64
Figura 18. Ensayos de <i>splicing</i> de las variantes espliceogénicas del exón 10 de <i>CHEK2</i>	65
Figura 19. Correlación entre las predicciones SRE de HEXplorer y HExoSplice y los resultados funcionales	70
Figura 20. Estructura y validación funcional del minigén final de <i>BARD1</i>	76
Figura 21. Ensayos funcionales de <i>splicing</i> de las μ del de <i>BARD1</i> espliceogénicas.....	78

Figura 22. Electroforesis en gel de agarosa de los transcritos generados tras la destrucción de los motivos hnRNPA1 predichos próximos a los donadores críticos del exón 4 de <i>BARD1</i>	81
Figura 23. Estructura y validación funcional del minigén final de <i>BRIP1</i>	84
Figura 24. Electroforesis en gel de agarosa de los resultados del ensayo funcional de las μ del incorporadas en el minigén mgBRIP1_ex1-2	85
Figura 25. Perfiles de HEXplorer de las regiones A) 5' y B) 3' del exón 4 de <i>BARD1</i>	112

ÍNDICE DE TABLAS

Tabla 1. Criterios de selección para el testado germinal de mutaciones asociadas al cáncer de mama establecidos por la SEOM.....	30
Tabla 2. Resultados de <i>splicing</i> de las μ del incorporadas en los exones 8 y 10 de <i>CHEK2</i>	62
Tabla 3. Resultados de <i>splicing</i> de las variantes exónicas espliceogénicas de los exones 8 y 10 de <i>CHEK2</i>	66
Tabla 4. Sensibilidad, especificidad y precisión de las herramientas de predicción de ESE/ESS	70
Tabla 5. Clasificación ACMG/AMP de las 38 variantes espliceogénicas testadas en los exones 8 y 10 de <i>CHEK2</i>	72
Tabla 6. Resultados de <i>splicing</i> de las microdeleciones espliceogénicas ensayadas en mgBARD1_ex3-5	79
Tabla 7. Variantes P/LP y VUS registradas en la base de datos ClinVar en los genes que conforman el núcleo central de la susceptibilidad a CM	90
Tabla 8. Resumen del análisis bioinformático, resultados de <i>splicing</i> y clasificación clínica siguiendo las normas ACMG/AMP de las 92 variantes de sitios de <i>splicing</i> testadas en esta tesis.....	94
Tabla 9. Secuencias WT y mutantes de las 10 variantes SRE de <i>CHEK2</i> que exhiben impactos muy fuertes, y motivo putativo de interacción del factor de <i>splicing</i> Tra2 β	113

ÍNDICE DEL ANEXO

1. TABLAS	147
Tabla Suplementaria 1. Primers para incorporar en los exones 8 y 10 del mgChk2_ex6-10 las 12 μ del de 10, 11, 15 y 30 nt.....	147
Tabla Suplementaria 2. Primers para incorporar las 87 variantes exónicas candidatas de los exones 8 y 10 de <i>CHEK2</i> en el mgChk2_ex6-10	148
Tabla Suplementaria 3. Resultados de <i>splicing</i> de las variantes exónicas no espliceogénicas de los exones 8 y 10 de <i>CHEK2</i>	152
Tabla Suplementaria 4. Análisis bioinformático de las variantes exónicas de <i>CHEK2</i> candidatas empleando distintos <i>software</i> de predicción de SRE.....	154
Tabla Suplementaria 5. Análisis comparativo de las predicciones de SpliceAI y las lecturas de los minigenes mgChk2_ex6-10 mutantes.....	157
Tabla Suplementaria 6. Sensibilidad, especificidad y precisión de SpliceAI	159
Tabla Suplementaria 7. Primers empleados para el clonaje consecutivo y la estabilización del minigén mgBARD1_ex3-5	160
Tabla Suplementaria 8. Primers para incorporar en el exón 4 del mgBARD1_ex3-5 las μ del de 30 y 12 nt	161
Tabla Suplementaria 9. Sitios de unión del <i>enhancer</i> Tra2 β predichos por la herramienta HSF	162
Tabla Suplementaria 10. Primers para destruir los motivos silenciadores predichos alrededor de los donadores crípticos del exón 4 de <i>BARD1</i>	163
Tabla Suplementaria 11. Primers empleados para el clonaje del minigén mgBRIP1_ex1-2 y su estabilización.....	164
Tabla Suplementaria 12. Primers para incorporar en los exones 1 y 2 del mgBRIP1_ex1-2 las μ del de 16 y 30 nt.....	165
2. FIGURAS	166
Figura Suplementaria 1. Estructura y validación funcional del minigén mgChk2_ex6 10	166
Figura Suplementaria 2. Secuencia del inserto del minigén mgChk2_ex6-10	167

Figura Suplementaria 3. Análisis de fragmentos fluorescentes de las variantes no espliceogénicas de los exones A) 8 y B) 10 de <i>CHEK2</i>	169
Figura Suplementaria 4. Análisis bioinformático de las variantes espliceogénicas de los exones 8 y 10 de <i>CHEK2</i>	172
Figura Suplementaria 5. Reproducibilidad de los resultados de <i>splicing</i> del minigén mgChk2_ex6-10.....	173
Figura Suplementaria 6. Conservación de la secuencia proteica de CHK2 en vertebrados	175
Figura Suplementaria 7A. Fortalezas de código patogénico/benigno aplicables a transcritos individuales del gen <i>CHEK2</i>	176
Figura Suplementaria 7B. Árbol de decisión propuesto, asignando un código PVS1_0/BP7_0 a las lecturas de los minigenes mutantes de <i>CHEK2</i>	177
Figura Suplementaria 8. Secuencia de los insertos empleados para construir los distintos minigenes de <i>BARD1</i>	179
Figura Suplementaria 9. Estructura y validación funcional de los minigenes intermediarios de <i>BARD1</i>	180
Figura Suplementaria 10. Electroforesis en gel de agarosa de los transcritos generados por las μ del incorporadas en mgBARD1_ex3-5.....	181
Figura Suplementaria 11. Representación de los donadores crípticos del exón 4 de <i>BARD1</i> y los sitios hnRNPA1 predichos que los rodean.....	182
Figura Suplementaria 12. Secuencia de los insertos empleados para construir los distintos minigenes de <i>BRIP1</i>	183
Figura Suplementaria 13. Estructura y validación funcional de los minigenes intermediarios de <i>BRIP1</i>	184
Figuras Suplementarias Excel en el siguiente enlace: https://doi.org/10.6084/m9.figshare.24223114.v2	

LISTADO DE ABREVIATURAS

μdel	Microdelección
3'ss	Sitio aceptor o sitio 3' de <i>splicing</i>
5'ss	Sitio donador o sitio 5' de <i>splicing</i>
5'UTR	<i>5' Untranslated Region</i>
ACMG	<i>American College of Medical Genetics</i>
AG	<i>Acceptor Gain</i>
AL	<i>Acceptor Loss</i>
AMP	<i>Association for Molecular Pathology</i>
AS	<i>Alternative Splicing</i>
B	<i>Benign</i>
BER	<i>Base Excision Repair</i>
BP	<i>Branch Point</i>
BRIDGES	<i>Breast Cancer Risk after Diagnostic Gene Sequencing</i>
CAP	<i>College of American Pathologists</i>
CERES	<i>Composite Exon Regulator Element of Splicing</i>
CM	Cáncer de Mama
CMH	Cáncer de Mama Hereditario
Cola poli(A)	Cola de poliadenina
DG	<i>Donor Gain</i>
DL	<i>Donor Loss</i>
DSB	<i>Double-Strand Break</i>
ER	<i>Estrogen Receptor</i>
ESE	<i>Exonic Splicing Enhancer</i>
ESS	<i>Exonic Splicing Silencer</i>
Ex	Exón
FHA domain	<i>Forkhead-Associated Domain</i>
GLOBOCAN	<i>The Global Cancer Observatory</i>
hnRNP	<i>Heterogeneous Nuclear Ribonucleoprotein</i>
HR	<i>Homologous recombination</i>
HSF	<i>Human Splicing Finder</i>
IBGM	Instituto de Biomedicina y Genética Molecular
ICL	<i>Interstrand Cross-link</i>
InDel	Inserción/delección
ISE	<i>Intronic Splicing Enhancer</i>
ISS	<i>Intronic Splicing Silencer</i>
Ivs	Intrón
kb	Kilobase
KD domain	<i>Kinase Domain</i>
kDa	Kilodalton
LB	<i>Likely Benign</i>
LP	<i>Likely Pathogenic</i>
MES	MaxEntScan
MMR	<i>Mismatch Repair</i>
NCCN	<i>National Comprehensive Cancer Network</i>
NGS	<i>Next-Generation Sequencing</i>
NLS	<i>Nuclear Localization Signal</i>
NMD	<i>Nonsense-Mediated mRNA decay</i>
nt	Nucleótido

ORF	<i>Open Reading Frame</i>
P	<i>Pathogenic</i>
p. ej.	Por ejemplo
pb	Par de bases
PPT	<i>Polypyrimidine Tract</i>
PR	<i>Progesterone Receptor</i>
Pre-ARNm	ARN mensajero precursor
PTC	<i>Premature Termination Codon</i>
RBP	<i>RNA-Binding Protein</i>
RFU	<i>Relative Fluorescence Units</i>
RNA pol II	<i>RNA polymerase II</i>
SC2	Sitio críptico 2 de <i>BARD1</i>
SC4	Sitio críptico 4 de <i>BARD1</i>
SC6	Sitio críptico 6 de <i>BARD1</i>
SCD domain	<i>SQ/TQ Cluster Domain</i>
SEOM	Sociedad Española de Oncología Médica
snRNP	<i>Small Nuclear Ribonucleoprotein</i>
Splicing ClinGen	<i>Clinical Genome Resource Sequence Variant Interpretation</i>
SVI	<i>Working Group</i>
SR protein	<i>Serine/Arginine-rich protein</i>
SRE	<i>Splicing Regulatory Element</i>
Transcrito mgFL	Transcrito <i>Full-Length</i> del Minigén
U12DB	<i>U12 Intron Database</i>
V1	Exón 1 del vector
V2	Exón 2 del vector
VUS	<i>Variant of Uncertain Significance</i>
WT	<i>Wild Type</i>

LISTADO DE GENES

9G8	<i>Splicing Factor 9G8</i>
AmpR	<i>Transcriptional regulator AmpR</i>
APE1	<i>Apurinic/Apyrimidinic Endodeoxyribonuclease</i>
ASF/SF2	<i>Alternative splicing factor/splicing factor 2</i>
ATM	<i>ATM Serine/Threonine Kinase</i>
ATR	<i>ATR Serine/Threonine Kinase</i>
BACH1	<i>BTB Domain and CNC Homolog 1</i>
BARD1	<i>BRCA1-Associated Ring Domain 1</i>
BCL3	<i>BCL3 Transcription Coactivator</i>
BIRC5	<i>Baculoviral IAP Repeat Containing 5</i>
BLM	<i>BLM RecQ Like Helicase</i>
BRCA1	<i>Breast Cancer Type 1</i>
BRCA2	<i>Breast Cancer Type 2</i>
BRIP1	<i>BRCA1-Interacting Protein 1</i>
BTK	<i>Bruton Tyrosine Kinase</i>
CDH1	<i>Cadherin 1</i>
CELF1	<i>CUGBP Elav-Like Family Member-1</i>
CEN2	<i>Centrin 2</i>
CHD7	<i>Chromodomain Helicase DNA Binding Protein 7</i>
CHEK1	<i>Checkpoint Kinase 1</i>
CHEK2	<i>Checkpoint Kinase 2</i>
COL1A1	<i>Collagen Type I Alpha 1 Chain</i>
CSA	<i>ERCC Excision Repair 8, CSA Ubiquitin Ligase Complex Subunit</i>
CSB	<i>ERCC Excision Repair 6, Chromatin Remodeling Factor</i>
CtIP	<i>RB Binding Protein 8 Endonuclease</i>
DDX23	<i>DEAD-Box Helicase 23</i>
DDX39B	<i>DEAD-Box Helicase 39B</i>
DDX46	<i>DEAD-Box Helicase 46</i>
DHX8	<i>DEAH-Box Helicase 8</i>
DHX15	<i>DEAH-Box Helicase 15</i>
DHX16	<i>DEAH-Box Helicase 16</i>
DHX38	<i>DEAH-Box Helicase 38</i>
DMTF1	<i>Cyclin D Binding Myb Like Transcription Factor 1</i>
EFTUD2	<i>Elongation Factor Tu GTP Binding Domain Containing 2</i>
ERBB2	<i>Erb-B2 Receptor Tyrosine Kinase 2</i>
ERCC1	<i>ERCC Excision Repair 1, Endonuclease non-Catalytic Subunit</i>
EXO1	<i>Exonuclease 1</i>
FAAP24	<i>FA Core Complex Associated Protein 24</i>
FANCD1	<i>Fanconi Anemia Complementation Group D1</i>
FANCD2	<i>Fanconi Anemia Complementation Group I</i>
FANCI	<i>Fanconi Anemia Complementation Group D2</i>
FANCF	<i>Fanconi Anemia Complementation Group J</i>
FANCM	<i>Fanconi Anemia Complementation Group M</i>
FANCN	<i>Fanconi Anemia Complementation Group N</i>
FANCO	<i>Fanconi Anemia Complementation Group O</i>
FANCP	<i>Fanconi Anemia Complementation Group P</i>
FANCF	<i>Fanconi Anemia Complementation Group Q</i>
FEN1	<i>Flap Structure-Specific Endonuclease 1</i>

FGD1	<i>FYVE, RhoGEF and PH Domain Containing 1</i>
FGFR1	<i>Fibroblast Growth Factor Receptor 1</i>
FGFR2	<i>Fibroblast Growth Factor Receptor 2</i>
FOX-M1	<i>Forkhead Box M1</i>
GLA	<i>Galactosidase Alpha</i>
GRN	<i>Granulin Precursor</i>
HER2	<i>Human Epidermal Growth Factor Receptor 2</i>
Hexon	<i>Adenovirus Hexon Gene</i>
HNRNPA0	<i>Heterogeneous Nuclear Ribonucleoprotein A0</i>
HNRNPA1	<i>Heterogeneous Nuclear Ribonucleoprotein A1</i>
HNRNPA2B1	<i>Heterogeneous Nuclear Ribonucleoprotein A2/B1</i>
HNRNPA3	<i>Heterogeneous Nuclear Ribonucleoprotein A3</i>
HNRNPAB	<i>Heterogeneous Nuclear Ribonucleoprotein AB</i>
HNRNPC	<i>Heterogeneous Nuclear Ribonucleoprotein C</i>
HNRNPD	<i>Heterogeneous Nuclear Ribonucleoprotein D</i>
HNRNPE1	<i>Heterogeneous Nuclear Ribonucleoprotein E1</i>
HNRNPE2	<i>Heterogeneous Nuclear Ribonucleoprotein E2</i>
HNRNPF	<i>Heterogeneous Nuclear Ribonucleoprotein F</i>
HNRNPG	<i>Heterogeneous Nuclear Ribonucleoprotein G</i>
HNRNPH1	<i>Heterogeneous Nuclear Ribonucleoprotein H1</i>
HNRNPH2	<i>Heterogeneous Nuclear Ribonucleoprotein H2</i>
HNRNPI	<i>Heterogeneous Nuclear Ribonucleoprotein I</i>
HNRNPK	<i>Heterogeneous Nuclear Ribonucleoprotein K</i>
HNRNPL	<i>Heterogeneous Nuclear Ribonucleoprotein L</i>
HNRNPM	<i>Heterogeneous Nuclear Ribonucleoprotein M</i>
HNRNPQ	<i>Heterogeneous Nuclear Ribonucleoprotein Q</i>
HNRNPR	<i>Heterogeneous Nuclear Ribonucleoprotein R</i>
HNRNPU	<i>Heterogeneous Nuclear Ribonucleoprotein U</i>
HR23B	<i>RAD23 Homolog B, Nucleotide Excision Repair Protein</i>
KAP1	<i>KRAB-Associated Protein 1</i>
KHDRBS3	<i>KH RNA Binding Domain Containing, Signal Transduction Associated 3</i>
KLF6	<i>KLF Transcription Factor 6</i>
KU70	<i>X-Ray Repair Cross Complementing 6</i>
KU80	<i>X-Ray Repair Cross Complementing 5</i>
LacZ	<i>Beta-Galactosidase</i>
LIG1	<i>DNA Ligase 1</i>
LIG3	<i>DNA Ligase 3</i>
LIG4	<i>DNA Ligase 4</i>
MBNL1	<i>Muscleblind-Like protein1</i>
MFH1	<i>Forkhead Box C2</i>
MKI67	<i>Marker of Proliferation Ki-67</i>
MLH1	<i>MutL Homolog 1</i>
MRE11	<i>MRE11 Homolog, Double Strand Break Repair Nuclease</i>
MSH2	<i>MutS Homolog 2</i>
MSH3	<i>MutS Homolog 3</i>
MSH6	<i>MutS Homolog 6</i>
NBN	<i>Nibrin</i>
NF1	<i>Neurofibromin 1</i>
NFκB	<i>Nuclear Factor Kappa B Subunit 1</i>
NOVA1	<i>NOVA Alternative Splicing Regulator 1</i>
NOVA2	<i>NOVA Alternative Splicing Regulator 2</i>

PALB2	<i>Partner and Localizer of BRCA2,</i>
PARP	<i>Poly (ADP-Ribose) Polymerase</i>
PCNA	<i>Proliferating Cell Nuclear Antigen</i>
PMS2	<i>PMS1 Homolog 2, Mismatch Repair System Component</i>
PTBP1	<i>Polypyrimidine Tract-Binding Protein 1</i>
PTBP2	<i>Polypyrimidine Tract-Binding Protein 2</i>
PTEN	<i>Phosphatase and Tensin Homolog</i>
RAD51	<i>RAD51 Recombinase</i>
RAD51B	<i>RAD51 Paralog B</i>
RAD51C	<i>RAD51 Paralog C</i>
RAD51D	<i>RAD51 Paralog D</i>
RBFOX1	<i>RNA binding Fox-1 Homolog 1</i>
RBFOX2	<i>RNA Binding Fox-1 Homolog 2</i>
RBM24	<i>RNA Binding Motif Protein 24</i>
REV1	<i>REV1 DNA Directed Polymerase</i>
RFC	<i>Replication Factor C Subunit</i>
RPA	<i>Replication Protein A</i>
SC35	<i>Splicing Component, 35 KDa</i>
SERPINA1	<i>Serpin Family A Member 1</i>
SF1	<i>Splicing Factor 1</i>
SF3B1	<i>Splicing Factor 3b Subunit 1</i>
SGMS2	<i>Sphingomyelin Synthase 2</i>
SLX4	<i>SLX4 Structure-Specific Endonuclease Subunit</i>
Sm-B/B'	<i>Small Nuclear Ribonucleoprotein Polypeptides B and B1</i>
Sm-D1	<i>Small Nuclear Ribonucleoprotein D1 Polypeptide</i>
Sm-D2	<i>Small Nuclear Ribonucleoprotein D2 Polypeptide</i>
Sm-D3	<i>Small Nuclear Ribonucleoprotein D3 Polypeptide</i>
Sm-E	<i>Small Nuclear Ribonucleoprotein Polypeptide E</i>
Sm-F	<i>Small Nuclear Ribonucleoprotein Polypeptide F</i>
Sm-G	<i>Small Nuclear Ribonucleoprotein Polypeptide G</i>
SNRNP200	<i>Small Nuclear Ribonucleoprotein U5 Subunit 200</i>
SPINK5	<i>Serine Peptidase Inhibitor Kazal Type 5</i>
SRP30C	<i>Pre-mRNA-Splicing Factor SRp30C</i>
SRP40	<i>Pre-mRNA-Splicing Factor SRP40</i>
SRP55	<i>Splicing Factor, Arginine/Serine-Rich, 55 KDa</i>
SRRM4	<i>Serine/Arginine Repetitive Matrix Protein 4</i>
SRrp35	<i>SR-Repressor Protein 35</i>
SRrp40	<i>SR-Repressor Protein 40</i>
SRSF1	<i>Serine and Arginine Rich Splicing Factor 1</i>
SRSF10	<i>Serine and Arginine Rich Splicing Factor 10</i>
SRSF11	<i>Serine and Arginine Rich Splicing Factor 11</i>
SRSF12	<i>Serine and Arginine Rich Splicing Factor 12</i>
SRSF2	<i>Serine and Arginine Rich Splicing Factor 2</i>
SRSF3	<i>Serine and Arginine Rich Splicing Factor 3</i>
SRSF4	<i>Serine and Arginine Rich Splicing Factor 4</i>
SRSF5	<i>Serine and Arginine Rich Splicing Factor 5</i>
SRSF6	<i>Serine and Arginine Rich Splicing Factor 6</i>
SRSF7	<i>Serine and Arginine Rich Splicing Factor 7</i>
SRSF8	<i>Serine and Arginine Rich Splicing Factor 8</i>
SRSF9	<i>Serine and Arginine Rich Splicing Factor 9</i>
STK11	<i>Serine/Threonine Kinase 11</i>

SWSAP1	<i>SWIM-Type Zinc Finger 7 Associated Protein 1</i>
TDP43	<i>Transactive-Response DNA-Binding Protein 43</i>
TOP2A	<i>DNA Topoisomerase II Alpha</i>
TOPBP1	<i>DNA Topoisomerase II Binding Protein 1</i>
TP53	<i>Tumor Protein P53</i>
TRA2B	<i>Transformer 2 Beta Homolog</i>
TRPM4	<i>Transient Receptor Potential Cation Channel Subfamily M Member 4</i>
U2AF1	<i>U2 Small Nuclear RNA Auxiliary Factor 1</i>
U2AF2	<i>U2 Small Nuclear RNA Auxiliary Factor 2</i>
UBE2T	<i>Ubiquitin Conjugating Enzyme E2 T</i>
UGT1A1	<i>UDP Glucuronosyltransferase Family 1 Member A1</i>
XLF	<i>Non-Homologous End Joining Factor 1</i>
XPA	<i>XPA, DNA Damage Recognition and Repair Factor</i>
XPB	<i>ERCC Excision Repair 3, TFIIH Core Complex Helicase Subunit</i>
XPC	<i>XPC Complex Subunit, DNA Damage Recognition and Repair Factor</i>
XPD	<i>ERCC Excision Repair 2, TFIIH Core Complex Helicase Subunit</i>
XPF	<i>ERCC Excision Repair 4, Endonuclease Catalytic Subunit</i>
XPG	<i>ERCC Excision Repair 5, Endonuclease 2</i>
XRCC1	<i>X-Ray Repair Cross Complementing 1</i>
XRCC2	<i>X-Ray Repair Cross Complementing 2</i>
XRCC3	<i>X-Ray Repair Cross Complementing 3</i>
XRCC4	<i>X-Ray Repair Cross Complementing 4</i>
ZRSR2	<i>Zinc Finger CCCH-Type, RNA Binding Motif and Serine/Arginine Rich 2</i>

RESUMEN

RESUMEN

Introducción: De un 5-10% de todos los cánceres de mama se consideran de tipo hereditario y, por tanto, atribuibles a mutaciones germinales deletéreas en genes de predisposición a esta enfermedad. Recientemente, dos grandes estudios caso/control han evidenciado que son 8 genes los que muestran una asociación significativa con el cáncer de mama: *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *BARD1*, *RAD51C* y *RAD51D*. El rastreo de estos genes empleando técnicas de secuenciación masiva ha permitido la detección de miles de variantes, muchas de ellas clasificadas como variantes de significado clínico desconocido (VUS). Interesantemente, la desregulación del *splicing* es un mecanismo etiopatogénico prevalente en genes de susceptibilidad a esta enfermedad. En este contexto, los ensayos funcionales de *splicing* de variantes potencialmente espliceogénicas permitirán la reclasificación de una amplia fracción de VUS. El objetivo de esta tesis doctoral es la clasificación funcional y clínica de variantes de sitios de *splicing* identificadas en los genes de predisposición a cáncer de mama *CHEK2* y *RAD51C*, así como la caracterización de los mecanismos reguladores implicados en la inclusión de exones sujetos a *splicing* alternativo, como los exones 8 y 10 de *CHEK2*, y de exones especiales, como el exón 4 de *BARD1* (macroexón) y el exón 1 de *BRIP1* (donador GC).

Materiales y métodos: Inicialmente, para la selección de variantes de sitios de *splicing* potencialmente espliceogénicas de los genes *CHEK2* y *RAD51C*, las mutaciones reportadas tanto en el proyecto europeo BRIDGES (*Breast Cancer Risk after Diagnostic Gene Sequencing*; <https://cordis.europa.eu/project/id/634935>) como en la base de datos pública ClinVar fueron analizadas *in silico* mediante los programas MaxEntScan (http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html) y SpliceAI (<https://spliceailookup.broadinstitute.org/>). A continuación, para su estudio funcional, se construyeron 4 minigenes reporteros de *splicing*: mgR51C_ex2-8, mgChk2_ex1-7, mgChk2_ex6-10 y mgChk2_ex11-15. Finalmente, las variantes seleccionadas se incorporaron en los minigenes *wild type* mediante mutagénesis dirigida y se ensayaron funcionalmente en varias líneas celulares. Por otra parte, para los estudios de regulación, se usó el minigén mgChk2_ex6-10 y se construyeron otros 2 minigenes híbridos (mgBARD1_ex3-5 y mgBRIP1_ex1-2). Con el fin de identificar intervalos ricos en *enhancers*, se siguió la estrategia de microdeleciones solapantes. Además, en los exones 8 y 10 de *CHEK2* se incorporaron por mutagénesis dirigida 87 variantes localizadas en los intervalos críticos.

Resultados: Todos los minigenes empleados produjeron un transcrito estable con el tamaño, secuencia y estructura esperados. Por un lado, un total de 92 variantes de sitios de *splicing* de los genes *CHEK2* y *RAD51C* fueron seleccionadas tras el filtrado bioinformático. Mediante ensayos en minigenes, observamos que la mayor parte de ellas alteraban el *splicing* (85/92; 92,4%), de las cuales 69 (75%) provocaban alteraciones graves (<5% de transcrito *full-length*). Según las normas ACMG/AMP (*American College of Medical Genetics/Association for Molecular Pathology*), 58 variantes fueron clasificadas como patogénicas/probablemente patogénicas. Curiosamente, el análisis funcional de la variante c.684-2A>G de *CHEK2* reveló el uso de un sitio aceptor TG atípico altamente conservado (17/23 posiciones), cuyo reconocimiento depende de secuencias críticas ubicadas en el exón 6 (c.685_698, c.709_722 y c.707_743). Por otro lado, el mapeo funcional de elementos reguladores del *splicing* reveló varias regiones esenciales para el correcto reconocimiento de los exones 8 y 10 de *CHEK2* (c.863_877, c.1073_1083 y c.1083_1092) y del exón 4 de *BARD1* (c.1232_1261, c.1257_1286 y c.1282_1311). Asimismo, treinta y ocho de las 87 variantes testadas (44%) en los intervalos diana de *CHEK2* alteraron el *splicing*, cuatro de las cuales (c.883G>A, c.883G>T, c.884A>T y c.1080G>T) indujeron un impacto casi total (<5% de transcrito *full-length*) y seis mostraron impactos fuertes (20-50% de transcrito *full-length*). De acuerdo con las normas ACMG/AMP, 3 variantes fueron clasificadas como patogénicas/probablemente patogénicas.

Conclusiones: La desregulación del *splicing* de los genes *RAD51C* y *CHEK2* es un mecanismo etiopatogénico prevalente en la predisposición genética al cáncer de mama. En dichos genes, se han clasificado 61 variantes como patogénicas/probablemente patogénicas, de modo que pacientes y familias portadoras se pueden beneficiar de medidas personalizadas preventivas e incluso terapéuticas. Asimismo, los exones 8 y 10 de *CHEK2*, sujetos a eventos de *splicing* alternativo, y el macroexón 4 de *BARD1* presentan regiones ricas en elementos reguladores del *splicing* que son nichos potenciales de variantes espliceogénicas. Por último, cabe destacar que los minigenes son una herramienta idónea para los estudios de regulación del *splicing* y la caracterización inicial del impacto sobre el *splicing* de las variantes de interés.

ABSTRACT

Introduction: 5-10% of all breast cancers are hereditary and, therefore, attributable to deleterious germline mutations in predisposition genes. Recently, two large case/control studies have revealed that 8 genes are significantly associated with breast cancer: *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *BARD1*, *RAD51C* and *RAD51D*. Screening of these genes using high-throughput sequencing has allowed the detection of thousands of variants, many of them classified as variants of unknown significance (VUS). Interestingly, splicing dysregulation is a prevalent etiopathogenic mechanism in breast cancer susceptibility genes. In this context, functional splicing assays of potentially spliceogenic variants reported in breast cancer susceptibility genes will allow the reclassification of a large fraction of VUS. The aim of this PhD thesis is to functionally and clinically classify splice site variants identified in the breast cancer predisposition genes *CHEK2* and *RAD51C*, as well as to characterize the regulatory mechanisms involved in the inclusion of exons subject to alternative splicing, such as exons 8 and 10 of *CHEK2*, and special exons, such as *BARD1* exon 4 (macroexon) and *BRIP1* exon 1 (GC donor).

Materials and methods: Initially, to select potentially spliceogenic splice site variants of the *CHEK2* and *RAD51C* genes, the mutations reported both in the European BRIDGES project (Breast Cancer Risk after Diagnostic Gene Sequencing; <https://cordis.europa.eu/project/id/634935>) and in the public database ClinVar were analyzed *in silico* using the programs MaxEntScan (http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html) and SpliceAI (<https://spliceailookup.broadinstitute.org/>). Then, for functional testing, 4 splicing reporter minigenes were constructed: mgR51C_ex2-8, mgChk2_ex1-7, mgChk2_ex6-10, and mgChk2_ex11-15. Finally, the selected variants were incorporated into the wild type minigenes by site-directed mutagenesis and functionally assayed in several cell lines. On the other side, for regulatory studies, minigene mgChk2_ex6-10 was employed and 2 other hybrid minigenes (mgBARD1_ex3-5 and mgBRIP1_ex1-2) were constructed. In order to identify enhancer-rich intervals, an overlapping microdeletion strategy was envisioned. In addition, 87 variants located in critical intervals were incorporated into *CHEK2* exons 8 and 10 by site-directed mutagenesis.

Results: All minigenes used produced a stable transcript with the expected size, sequence and structure. On the one hand, a total of 92 splice site variants of *CHEK2* and *RAD51C* genes were selected after bioinformatic filtering. Using minigene assays, we observed that most of them altered splicing (85/92; 92.4%), 69 of which (75%) caused severe alterations (<5% full-length transcript). According to ACMG/AMP (American College of Medical Genetics/Association for Molecular Pathology)-based standards, 58 variants were classified as pathogenic/likely pathogenic. Interestingly, functional analysis of the c.684-2A>G variant of *CHEK2* revealed the use of a highly conserved atypical TG acceptor site (17/23 positions), whose recognition depends on critical sequences located in exon 6 (c.685_698, c.709_722 and c.707_743). On the other hand, functional mapping of splicing regulatory elements revealed several regions essential for recognition of exons 8 and 10 of *CHEK2* (c.863_877, c.1073_1083 and c.1083_1092) and exon 4 of *BARD1* (c.1232_1261, c.1257_1286 and c.1282_1311). Also, thirty-eight out of the 87 variants tested (44%) in *CHEK2* critical intervals altered splicing, four of which (c.883G>A, c.883G>T, c.884A>T and c.1080G>T) induced a near-total impact (<5% of full-length transcript) and six showed strong impacts (20-50% of full-length transcript). According to ACMG/AMP standards, 3 variants were classified as pathogenic/likely pathogenic.

Conclusions: Splicing dysregulation of *RAD51C* and *CHEK2* is a prevalent ethiopathogenic mechanism in the genetic predisposition to breast cancer. In these genes, 61 variants have been classified as pathogenic/likely pathogenic, so that carrier patients and families may benefit from personalized preventive and even therapeutic measures. Likewise, *CHEK2* exons 8 and 10, which undergo alternative splicing events, as well as *BARD1* macroexon 4, present splicing regulatory elements-enriched regions that are potential clusters of spliceogenic variants. Lastly, minigenes are an ideal tool for splicing regulatory studies and the initial characterization of the impact on splicing of candidate variants.

INTRODUCCIÓN

SPLICING

Parte

1

Durante los años 60 y 70, la biología molecular se estaba enfrentando a un gran enigma con respecto a la síntesis del ARNm en células animales. A lo largo de esos años, se había demostrado que una gran fracción del ARN nuclear se degradaba muy rápidamente después de su síntesis, mientras que solo una pequeña porción del ARN inicialmente sintetizado se exportaba al citoplasma, donde funcionaba como un ARN mucho más estable. Varias líneas de investigación habían evidenciado que dichos ARN presentaban una relación precursor-producto, algo que solamente podía explicarse a través de la existencia de un proceso encargado de la escisión del ARN nuclear [1]. Sin embargo, surgían muchas dudas al respecto, entre ellas cómo se procesa este ARN nuclear para que ambos ARN mantengan en sus extremos 5' y 3' las mismas estructuras, en concreto, la caperuza y la cola de poliadenina (poli(A)), respectivamente.

En el año 1977, Phillip Sharp y Richard Roberts dieron respuesta a esta y otras preguntas descubriendo, de manera independiente, el proceso de *splicing*. Ambos visualizaron por microscopía electrónica que una única molécula de ARNm hibridaba de forma discontinua con el ADN que plagiaba (**Figura 1**) [2,3], poniendo de manifiesto que la secuencia génica está interrumpida por secuencias que no se traducen a proteínas. Se postuló que este ARNm maduro derivaba de un ARN nuclear precursor continuo que era sometido a un mecanismo de corte y empalme (*splicing*). Muy pronto se intuyó que este mecanismo de maduración del ARNm que desprende los «intrones» y anuda los «exones» entre sí, términos acuñados por el bioquímico estadounidense Walter Gilbert [4], era más intrincado de lo que se pensaba en un principio, pues podía tomar múltiples caminos y, de este modo, engendrar varios productos a partir del mismo precursor común [5,6]; la preclara de lo que ahora conocemos como *splicing* alternativo (AS).

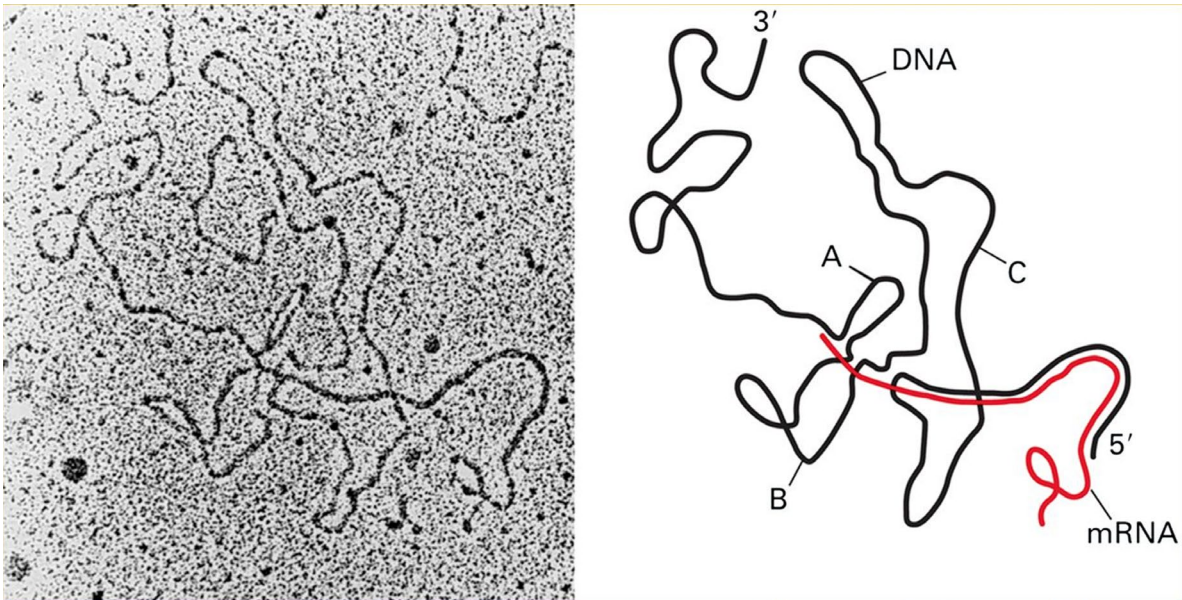


Figura 1. Micrografía electrónica de la hibridación del ARNm procedente del gen *hexón* con un fragmento de ADN de adenovirus. En esta imagen vemos una adaptación de parte de la figura 4 del artículo de Berget y colaboradores publicado el año 1977. En la imagen de la derecha se representa el ADN viral en negro y el ARNm en color rojo. Las regiones en las que el ARN rojo es paralelo a la cadena de ADN negra representan el híbrido ARN-ADN. Las regiones donde las dos cadenas de ADN negras son paralelas representan el dúplex ADN-ADN.

1. DEFINICIÓN DE *SPLICING*

En eucariotas, la mayoría de los genes contienen sus secuencias codificantes (exones) en segmentos discontinuos, interrumpidas por otras secuencias (intrones), generalmente más largas, que no formarán parte del ARNm maduro. Tras el proceso de transcripción de nuestros genes, reacción catalizada por la ARN polimerasa II (RNA pol II), se genera un ARN mensajero precursor (pre-ARNm) que deberá de sufrir una serie de modificaciones antes de convertirse en el ARNm maduro funcional que viajará al citoplasma celular y se traducirá en la proteína en cuestión. Además de la adición de la caperuza en el extremo 5' y la poliadenilación del extremo 3', otra de las modificaciones postranscripcionales a la que serán sometidos estos pre-ARNm es el proceso de *splicing* [7].

El *splicing* es un proceso clave de la expresión génica que elimina los intrones del pre-ARNm y enlaza ordenadamente los exones. Esta reacción es catalizada por un complejo macromolecular altamente dinámico tanto en conformación como en composición llamado espliceosoma. El correcto funcionamiento de esta maquinaria depende, por un lado, de una serie de secuencias conservadas que definen las fronteras

exón/intrón: el sitio donador o sitio 5' de *splicing* (5'ss), el punto de ramificación (BP), el tracto de polipirimidina (PPT) y el sitio aceptor o sitio 3' de *splicing* (3'ss) (**Figura 2**). Por otro lado, la actividad de esta maquinaria también está sujeta a la presencia de un conjunto de motivos reguladores en *cis* (*Splicing Regulatory Elements*, SRE) a los cuales se unen en *trans* factores de *splicing* activadores o represores que van a participar en el reconocimiento del exón [8].

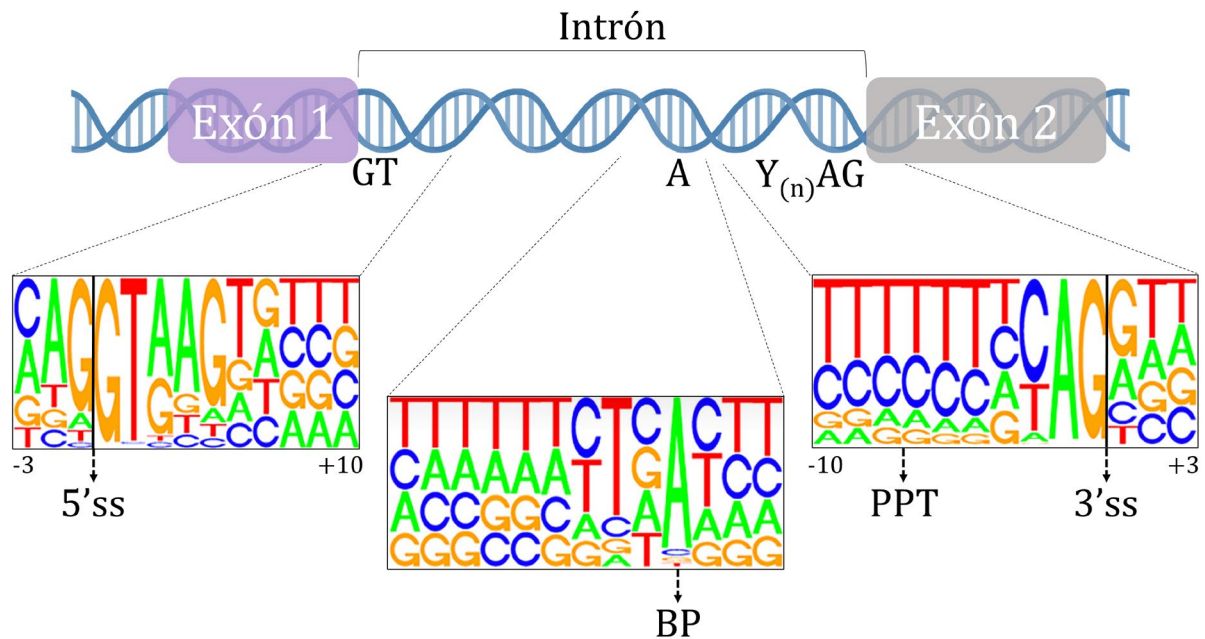


Figura 2. Secuencias consenso de *splicing* de un intrón U2-dependiente. En esta figura están representadas todas las secuencias consenso clave para el reconocimiento de un intrón por parte del espliceosoma mayor: sitio donador o sitio 5' de *splicing* (5'ss), punto de ramificación (BP), tracto de polipirimidina (PPT) y sitio aceptor o sitio 3' de *splicing* (3'ss) [8].

En eucariotas superiores, existen diversos factores que conceden al proceso de *splicing* una enorme plasticidad en la selección de los puntos de corte, lo que se conoce como *splicing* alternativo. Este permite que los exones de nuestros genes puedan combinarse entre sí de formas muy variopintas, originando distintos ARNm maduros que en algunos casos serán capaces de codificar proteínas funcionalmente diferentes (isoformas) [9]. Esta regulación flexible del proceso de *splicing* se produce en la mayoría de los genes de los organismos multicelulares y contribuye a la diversificación del transcriptoma y el proteoma, y a la regulación de los niveles de expresión génica, principalmente a través de la degradación de los transcritos que contienen codones de terminación prematura (PTC). El AS permite que las células se adapten rápidamente a los cambios en el entorno circundante, manteniendo la homeostasis tisular, promueve la diferenciación celular, el desarrollo de tejidos y la organogénesis durante el desarrollo embrionario, y preserva la identidad y la función del tejido. Por ello, su desregulación

conlleva graves consecuencias que pueden suponer el origen de una gran variedad de enfermedades humanas, como el cáncer [10,11].

Por otra parte, existen varios tipos de *splicing* no canónico que también actúan sobre nuestros genes, como es el caso del *back-splicing*, capaz de generar transcritos circulares, o el *trans-splicing*, que origina transcritos quiméricos al unir exones de dos moléculas de pre-ARNm independientes, derivadas del mismo gen (intragénico) o de genes diferentes (intergénico). El *splicing* no canónico es también responsable de la incorporación de microexones (<30 nt) en los ARNm maduros y la eliminación de intrones extremadamente largos de nuestros pre-ARNm. En el caso de los microexones, estos requerirán para su correcta inclusión de intrones flanqueantes más cortos y particularmente enriquecidos en motivos *enhancer* reconocidos por factores de *splicing* activadores muy específicos (p. ej., SRRM4, PTBP1 y RBFOX1) y de sitios 3'ss y 5'ss más fuertes. Por otro lado, la escisión de los intrones largos va a depender en gran medida de la presencia de los conocidos como sitios recursivos de *splicing* o «exones de longitud cero», que combinan los motivos 3'ss y 5'ss para que dicho intrón sea eliminado en varios pasos consecutivos [12,13]. Estos tipos de *splicing* no canónico contribuyen a la existencia de una gran diversidad de mecanismos moleculares y celulares y funciones biológicas, y desempeñan un papel fundamental en la regulación de la expresión génica.

1.1. ESPLICEOSOMA Y REACCIÓN DE *SPLICING*

El espliceosoma es un complejo de ribonucleoproteínas (snRNP) encargado de catalizar la reacción de *splicing*. Su naturaleza compleja y dinámica le confiere precisión a través de interacciones ARN-ARN, ARN-proteína y proteína-proteína que aseguran que los sitios reactivos del pre-ARNm están correctamente posicionados para la catálisis. Estas características del espliceosoma también le confieren una extraordinaria flexibilidad para responder rápidamente a cambios en el entorno celular [14].

En nuestras células, encontramos dos tipos de espliceosoma: el espliceosoma mayor o U2-dependiente, responsable de eliminar los intrones de tipo U2 (como el mostrado en la **Figura 2**), y el menos abundante, el espliceosoma menor o U12-dependiente, encargado de la escisión de una clase rara de intrones, los intrones de tipo U12, que constituyen solo 881 intrones humanos según *U12 intron database* (U12DB; <https://genome.crg.eu/cgi-bin/u12db/u12db.cgi>; último acceso el 12 de septiembre de 2023) [15]. La función y el significado del espliceosoma U12-dependiente aún están lejos

de comprenderse por completo, pero hay varias líneas de evidencia que indican que el *splicing* de este tipo de intrones tiene funciones esenciales en nuestro organismo [16]. La principal diferencia entre estos dos tipos de espliceosomas son las snRNP que los componen. El espliceosoma mayor está formado por las snRNP U1, U2, U4/U6 y U5, mientras que el espliceosoma menor solo comparte la snRNP U5 y el resto son análogos funcionales de las anteriores, denominados U11, U12 y U4atac/U6atac. En ambos casos, cada snRNP consta de un ARN nuclear pequeño (o dos en el caso de U4/U6 o U4atac/U6atac) y un conjunto común de siete proteínas Sm (B/B', D3, D2, D1, E, F y G). Además de las snRNP, en la reacción de *splicing* participan un número variable de proteínas asociadas, entre las que destacan una gran variedad de ARN helicasas dependientes de ATP (DDX46, DDX39B, DDX23, SNRNP200, EFTUD2, DHX16, DHX38, DHX8 y DHX15, entre otras) [17]. Otra diferencia entre estas dos maquinarias de *splicing* es el tipo de secuencias consenso que reconocen y, por tanto, el tipo de intrones que escinden, tipo U2- o U12-dependientes, como hemos explicado anteriormente. A pesar de estas diferencias, el proceso de *splicing* requiere de los mismos pasos: ensamblaje inicial del espliceosoma y formación del complejo catalítico, seguido por la ejecución de dos reacciones de transesterificación consecutivas y la disociación del complejo [18].

Inicialmente, el 5'ss y el BP son reconocidos por la snRNP U1 y por SF1, respectivamente (espliceosoma mayor), o bien por el complejo constituido por las snRNP U11/U12 (espliceosoma menor). Prácticamente al mismo tiempo, el 3'ss y el PPT son detectados por el heterodímero proteico U2AF1/U2AF2 en los intrones U2-dependientes, que servirá de guía, junto con SF1, para la unión de la snRNP U2 al BP. Mientras que, en el caso de los intrones U12-dependientes, el sitio 3' de *splicing* es localizado por la proteína ZRSR2. Hasta aquí se habría formado lo que se conoce como el complejo A (pre-espliceosoma). Tras el reconocimiento de los sitios de *splicing*, entran en escena las tri-snRNP U4/U6.U5 (espliceosoma mayor) o U4atac/U6atac.U5 (espliceosoma menor), formando el complejo B (espliceosoma pre-catalítico). La activación del complejo B dependerá de la liberación de U1 y U4 (intrones U2-dependientes) o U11 y U4atac (intrones U12-dependientes) y será la responsable de desencadenar el primer paso catalítico del proceso de *splicing*. En esta primera reacción de transesterificación, el grupo 2'OH de la adenina principal del BP desencadena un ataque nucleofílico sobre el 5'ss, el cual corta la cadena de ADN en este punto y enlaza el sitio 5'ss con esta adenina del BP, formando una estructura llamada «lazo intrónico» o *lariat*. A continuación, este complejo B se transforma en el complejo C (espliceosoma catalítico), que es el responsable de la

segunda reacción de transesterificación. En concreto, esta reacción es catalizada por el grupo 3'OH libre del 5'ss que ataca el sitio 3'ss, ligando los dos extremos exónicos y liberando el complejo P (post-espliceosoma), el cual incluye el intrón eliminado. Finalmente, se disocia el complejo y se liberan las snRNP y las proteínas asociadas (**Figura 3**) [19,20].

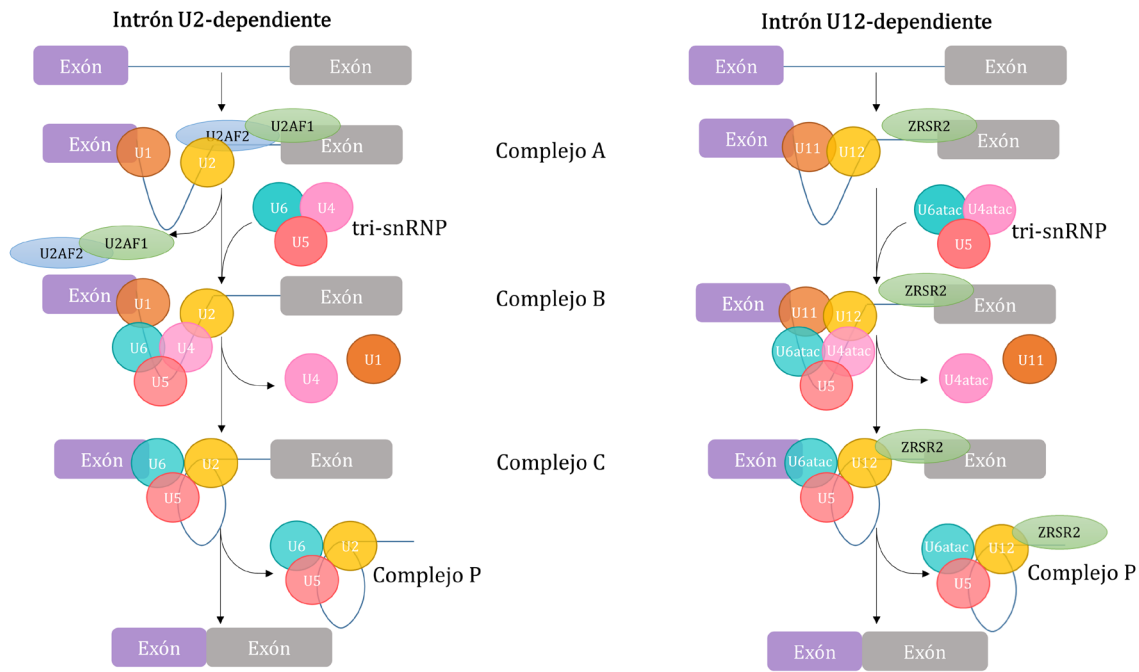


Figura 3. La reacción de *splicing*. Representación esquemática del ensamblaje del espliceosoma mayor (izquierda) y menor (derecha), formación de los distintos complejos espliceosómicos y reacción de *splicing* en intrones U2- (izquierda) y U12-dependientes (derecha).

Una cuestión que puede surgir después de entender cómo funciona el proceso de *splicing* es si el espliceosoma primero identifica los exones o si, por el contrario, el reconocimiento sucede a través del intrón. Hay evidencias que indican que esto va a depender de la arquitectura del genoma, en concreto, del tamaño de los exones e intrones implicados en la reacción de *splicing*. En eucariotas superiores, que generalmente muestran pre-ARNm con exones cortos separados por intrones mucho más largos, parece que predomina el modelo de «definición de exón», propuesto en el año 1990 por Susan Berget y, según el cual, la maquinaria de *splicing* se conforma a través del exón [21]. Sin embargo, en los eucariotas inferiores, los cuales presentan una arquitectura génica caracterizada por intrones pequeños y exones grandes, probablemente predomine el modelo de «definición de intrón», donde los sitios de *splicing* se emparejan entre intrones en lugar de entre exones [22].

Interesantemente, tanto el primer como el último exón van a requerir mecanismos especiales para su definición. En el caso del primer exón, que carece de señal de *splicing* 3'ss, la caperuza 5' es esencial para el reconocimiento del sitio 5' de *splicing* de este exón y la eliminación del primer intrón. Respecto al último exón, este comienza con un sitio aceptor y termina con una cola poli(A), la cual es crucial, junto con la maquinaria proteica encargada del procesamiento del extremo 3', en la escisión del último intrón [23].

1.2. SITIOS DE *SPLICING* CANÓNICOS Y ATÍPICOS

El mantenimiento de la integridad del marco abierto de lectura (ORF) requiere una definición precisa de los límites entre las regiones intrónicas y exónicas. Como se ha indicado en apartados anteriores, estas señales básicas en el ADN para que se efectúe la reacción de *splicing* engloban el sitio donador o sitio 5' de *splicing*, que está a caballo entre el final del exón y el principio del intrón, y tres secuencias consenso localizadas en el extremo opuesto: el sitio aceptor o sitio 3' de *splicing*, que incluye los últimos nucleótidos (nt) del intrón y los primeros del exón; el tracto de polipirimidina, ubicado entre las posiciones intrónicas -5 y -15, y el punto de ramificación, localizado unos 40 nt *upstream* del sitio 3'ss (**Figura 2**). Las posiciones altamente conservadas de estas secuencias clásicas de *splicing* son el dinucleótido GT (GU en el ARN), localizado en las posiciones intrónicas +1 y +2 (5'ss), y las posiciones intrónicas -1 y -2 que exhiben el dinucleótido AG (3'ss). Estas posiciones fuertemente conservadas están rodeadas por una serie de nucleótidos conservados en mayor o menor medida y todos ellos conjuntamente constituyen las secuencias consenso de *splicing*: 5'ss→MAG|GTRAGT, BP→YTNAY, PPT+3'ss→Y_(n)NYCAG|G (Exón|Intrón; M=C/A; R=A/G; Y=C/T; Y_n=PPT; N=A/C/G/T) [8]. Estas secuencias consenso son posibles dianas de variantes espliceogénicas, las cuales pueden producir transcritos deletéreos vinculados a una enfermedad [24].

Más del 99% de los intrones son eliminados a través del espliceosoma mayor, guiándose por los motivos de *splicing* consenso que acabamos de indicar y reflejados en la **Figura 2**. Sin embargo, como hemos visto en el apartado previo, existen otro tipo de intrones, los U12-dependientes (≈0,4%), cuyos sitios de *splicing* son ligeramente distintos y entre los cuales se incluyen los conocidos como intrones AT-AC (≈0,09%) (**Figura 4**). Algunas características diferenciales de los intrones U12 son: i) una conservación más fuerte de las secuencias de *splicing*, incluido el punto de ramificación, lo cual hace que sean más sensibles a las mutaciones puntuales, presenten una menor flexibilidad en la elección del sitio de corte y exhiban menores niveles de AS, y ii) la ausencia de PPT [16].

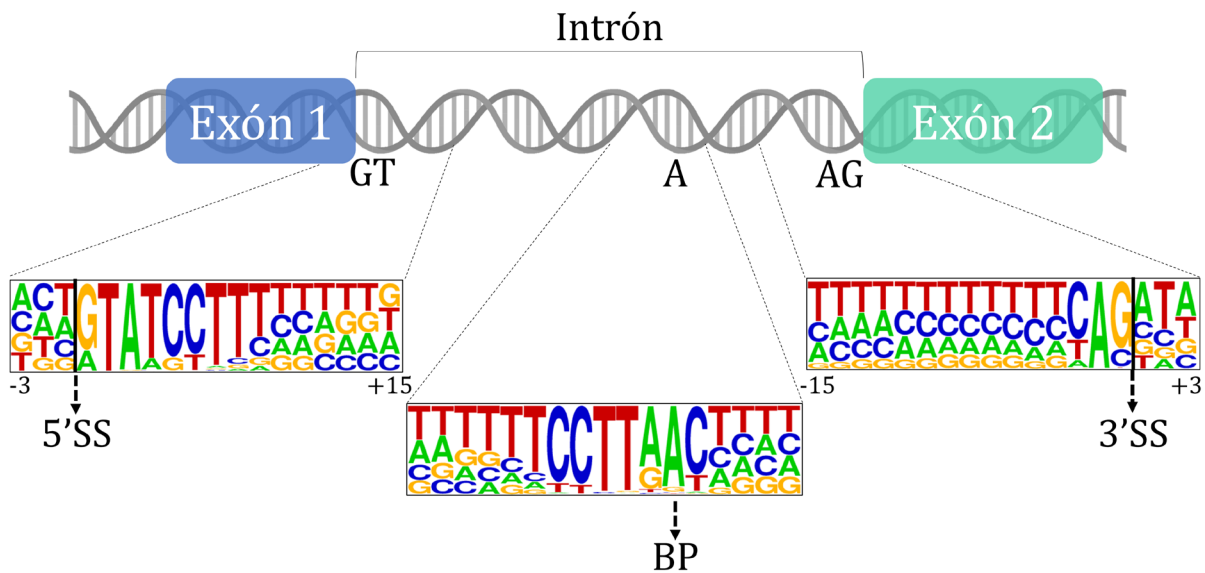


Figura 4. Secuencias consenso de *splicing* de un intrón U12-dependiente. En esta figura están representadas todas las secuencias consenso clave para el reconocimiento de un intrón por parte del espliceosoma menor (sitio donador o sitio 5' de *splicing* [5'ss], punto de ramificación [BP], y sitio aceptor o sitio 3' de *splicing* [3'ss]) obtenidas empleando la herramienta *online* WebLogo (<https://weblogo.berkeley.edu/logo.cgi>) y usando los datos publicados en la U12DB (último acceso el 12 de septiembre de 2023).

Por otro lado, existen una amplia gama de sitios atípicos, entre los que se incluyen los donadores no-canónicos más comunes, los GC ($\approx 0,9\%$) (**Figura 5**) [12]. En la mayoría de estos sitios de *splicing* no-canónicos, solo un nucleótido de las posiciones $\pm 1,2$ (GT-AG) está alterado. Generalmente, se trata de sitios de *splicing* con una fortaleza inferior a los canónicos, y es por ello por lo que suelen estar enriquecidos con motivos SRE y mostrar mayores niveles de *splicing* alternativo. Curiosamente, el análisis de algunos de estos sitios de *splicing* atípicos ha revelado su transformación en sitios canónicos a través de un proceso de edición del ARN que facilita su selección [25].

La existencia de esta gran variedad de sitios de *splicing* pone sobre la mesa lo intrincada que puede ser la regulación de este mecanismo. Además, la selección precisa de los sitios de *splicing* se complica más aún debido a la presencia de una gran cantidad de pseudositos, también denominados sitios crípticos, con secuencia similar o incluso idéntica a la consenso (**Figura 2**) que deben ser correctamente reprimidos [26].

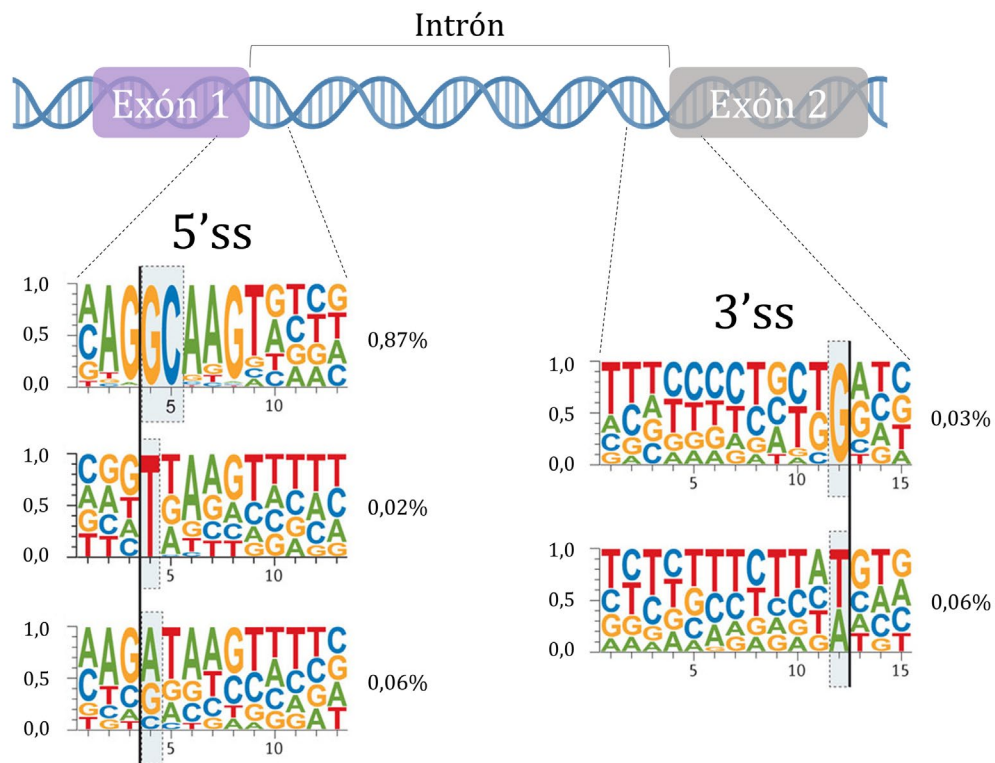


Figura 5. Sitios de *splicing* no-canónicos. Secuencias consenso de sitios de *splicing* 5'ss y 3'ss atípicos, y el porcentaje de aparición de cada uno de ellos en los intrones humanos [12].

1.3. REGULACIÓN DEL *SPLICING*

El reconocimiento de un exón es un proceso altamente complejo que depende de un control combinatorio único de múltiples parámetros, principalmente: i) la fortaleza de los sitios de *splicing*; ii) una amplia gama de motivos potenciadores o silenciadores que actúan *en cis*, los cuales reclutan factores reguladores del *splicing* que ejercen su función *en trans*; iii) la arquitectura genómica, incluidos los tamaños de exón e intrón; iv) la estructura secundaria del pre-ARNm, que puede exponer u ocultar tanto los sitios de *splicing* como motivos de unión para proteínas reguladoras de este proceso; v) la cinética de la transcripción, ya que este proceso se produce mayoritariamente de forma cotranscripcional, y vi) la presencia de nucleótidos modificados postranscripcionalmente (epitranscriptoma), que a su vez pueden modular la unión de ciertas proteínas o la formación de estructuras secundarias en el ARN [27]. De todos ellos, un factor clave en el reconocimiento de exones y en el *splicing* alternativo es la presencia de SRE. Los motivos SRE son secuencias degeneradas a las que pueden unirse diversas proteínas reguladoras que van a modular la reacción de *splicing* [10,26]. Básicamente, los SRE se pueden clasificar en *enhancers*, que favorecen la inclusión de los exones en el ARNm maduro, y silenciadores, cuyo efecto es bloquear el reconocimiento de los exones diana por parte de

la maquinaria de *splicing*. Estos motivos pueden localizarse tanto en regiones intrónicas como exónicas, distinguiéndose 4 tipos de SRE: *Exonic Splicing Enhancers* (ESE), *Exonic Splicing Silencers* (ESS), *Intronic Splicing Enhancers* (ISE) e *Intronic Splicing Silencers* (ISS).

Las dos familias clásicas de proteínas de unión al ARN (RBP) que actúan en *trans* regulando el proceso de *splicing* son la familia de proteínas SR (proteínas ricas en serina/arginina), que suelen actuar como factores potenciadores del proceso de *splicing*, y la familia de ribonucleoproteínas nucleares heterogéneas (hnRNP), las cuales generalmente ejercen un papel silenciador (**Figura 6**) [28]. Sin embargo, esta clasificación es muy simplista, puesto que a menudo el mismo SRE y sus factores en *trans* afines pueden mostrar efectos contrapuestos de acuerdo con su posición, ostentando un papel tanto represor como activador en función del contexto genético y su ubicación a lo largo del gen. De hecho, se han detectado diferencias en el funcionamiento de algunos de estos factores reguladores del *splicing* según se unan a motivos ubicados *upstream* o *downstream* del exón regulado (p. ej., los factores NOVA) o a SRE intrónicos o exónicos (como muchas hnRNP) [29]. Por otra parte, los conocidos como CERES (*Composite Exon Regulator Element of Splicing*) son otro ejemplo clásico de SRE cuya actividad potenciadora o silenciadora está fuertemente asociada al contexto genómico [30]. Adicionalmente, cabe destacar que el reconocimiento eficiente de un exón depende de efectos combinatorios o cooperativos/competitivos entre distintos factores reguladores[31].

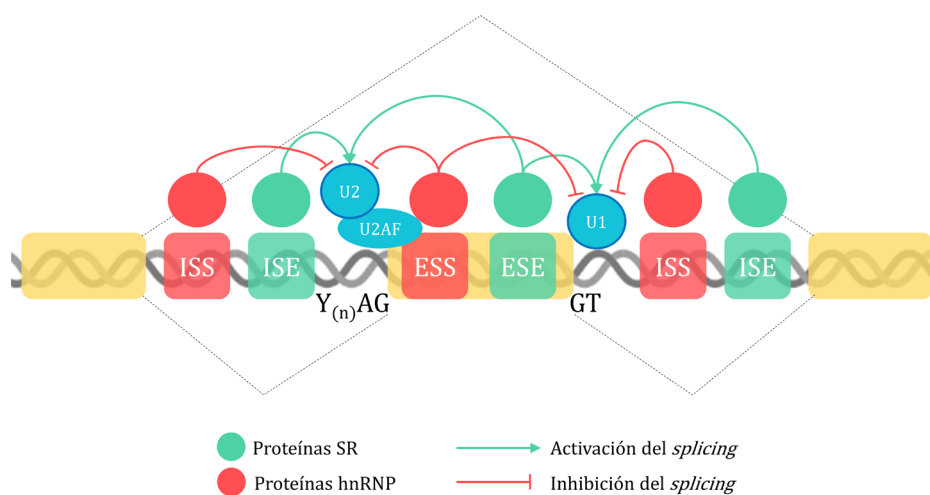


Figura 6. Secuencias en *cis* y factores en *trans* que participan en la regulación del proceso de *splicing*. Representación esquemática de los principales motivos SRE (ESE, ESS, ISE e ISS) y factores de *splicing* (proteínas SR y hnRNP) implicados en la regulación de este mecanismo.

Por un lado, la familia proteica SR comprende una serie de proteínas filogenéticamente conservadas y estructuralmente relacionadas, caracterizadas por poseer un dominio rico en arginina y serina en la región C-terminal, denominado dominio RS. El estado de fosforilación de dicho dominio cambia rápidamente en respuesta al estrés y a la señalización celular, y es el encargado de regular la actividad y el transporte nucleocitoplasmático de las proteínas SR. En particular, la familia de proteínas SR «básicas» consta de 12 miembros relativamente bien caracterizados, presentados en la **Figura 7** [32,33]. En 1990, fue descubierto de forma paralela por dos grupos distintos el primer miembro de esta familia proteica, SRSF1, al que denominaron SF2 y ASF, y desde entonces las funciones en el *splicing* de estas proteínas SR han sido ampliamente investigadas [34,35]. Las proteínas SR ejercen funciones esenciales en el *splicing* tanto constitutivo como alternativo, pues se encargan de promover el ensamblaje del espliceosoma a los ARN nacientes y favorecen la transición hacia los distintos complejos espliceosómicos. En concreto, estas proteínas son claves en el proceso de definición de exón, participando en el reclutamiento de las snRNP U1, U11, U2 y U12 a los sitios de *splicing*, y están implicadas tanto en la formación del complejo B, colaborando en el ensamblaje de las tri-snRNP (U4/U6.U5 y U4atac/U6atac.U5), como en la formación del complejo C [36]. Curiosamente, se ha demostrado que algunas de estas proteínas pueden ejercer también un papel antagónico en el *splicing* cuando se unen a regiones intrónicas, pues probablemente constituyan un impedimento estérico para la formación del espliceosoma [32]. Otras directamente funcionan habitualmente como factores de *splicing* supresores como SRSF10 o SRSF12, que de hecho inicialmente fueron nombradas como SRrp40 y SRrp35, respectivamente, siglas que derivan de *SR-repressor proteins* [37]. Además, esta familia de proteínas interviene también en otros aspectos cruciales del metabolismo del ARNm, como la poliadenilación, transcripción, empaquetamiento y compactación de las ribonucleoproteínas mensajeras, exportación y transporte al citoplasma de transcritos, regulación de la estabilidad del ARNm a través de la vía *Nonsense-Mediated mRNA decay* (NMD) y traducción. [33,38]. Esta flexibilidad funcional de la familia de proteínas SR es crucial para la plasticidad, solidez y adaptabilidad de los programas de expresión génica.

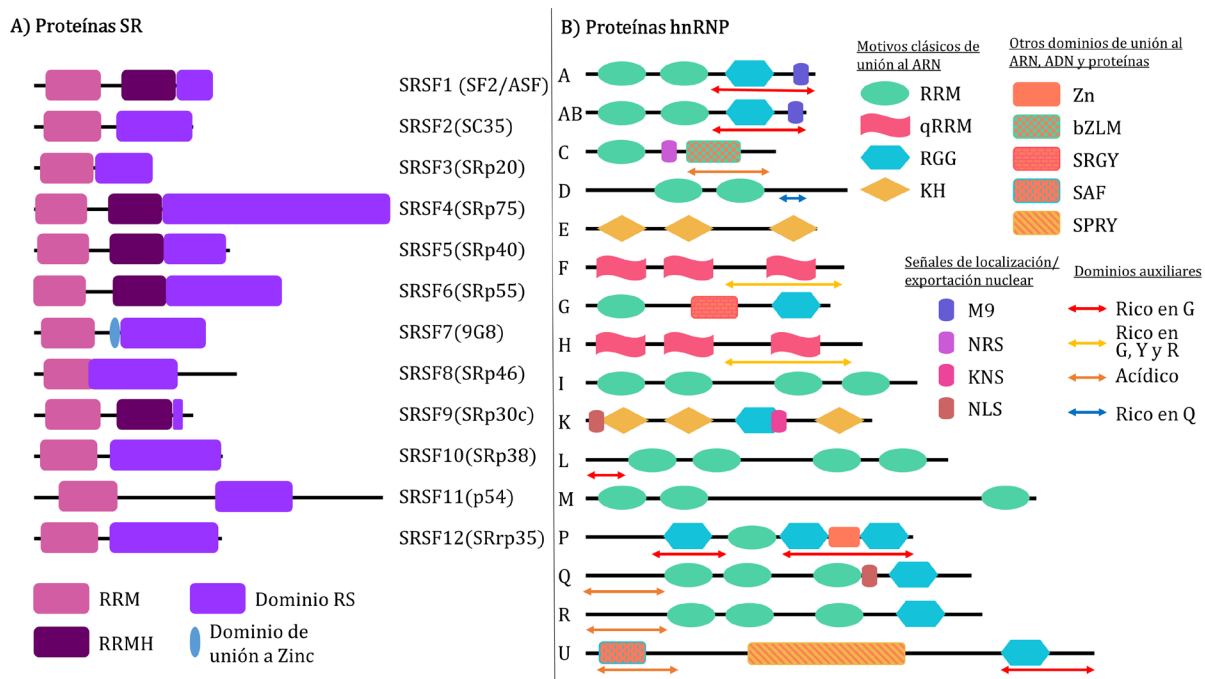


Figura 7. Estructura de las proteínas SR y hnRNP humanas. A) Familia de proteínas SR. Se muestra la organización estructural de las proteínas SR clásicas. Además de su nomenclatura actual, se ha incluido entre paréntesis sus nombres alternativos. RS, dominio rico en Arg/Ser; RRM, *RNA recognition motif*; RRMH, *RNA Recognition Motif Homologous*; Zn, *Zinc Binding Domain*. **B) Familia de proteínas hnRNP.** Se muestra la organización estructural de las proteínas hnRNP clásicas, desde la A hasta la U. RRM, *RNA recognition motif*; qRRM, *quasi-RNA recognition motif*; KH, *K-homology domain*; la caja RGG, que consta de repeticiones de Arg-Gly-Gly; Zn, dedos de zinc; bZLM, *basic leucine Zipper-like Motif*; SRGY, motivo rico en serina, arginina, glicina y tirosina; SAF, *scaffold attachment factor domain*; SPRY, *spore lysis A and ryanodine receptor homology domain*; M9, señal de exportación nucleocitoplasmática; NRS, señal de retención nuclear; KNS dominio K de transporte nuclear, y NLS, señal de localización nuclear; G, glicina; Y, tirosina; R, arginina; Q glutamina.

Por otro lado, las hnRNPs son una amplia familia de RBP, estructural y funcionalmente distintas, que comprende aproximadamente 20 proteínas principales, cuyos pesos moleculares van desde 34 a 120 kilodaltons (kDa). Estructuralmente, las hnRNP presentan una organización modular que consta de uno o más dominios de unión al ARN y otros motivos relacionados con la redistribución subcelular o la interacción con el ADN o con otras proteínas (**Figura 7**) [39,40]. Además de la modularidad creada por la combinación de múltiples dominios, la complejidad de las hnRNP aumenta aún más a través de distintas variantes alternativas de *splicing* y modificaciones postraduccionales, las cuales les confieren una gran versatilidad y una enorme diversidad funcional [41]. Las hnRNP desempeñan un papel importante en todas las facetas del metabolismo del ARN, especialmente en el proceso de *splicing* constitutivo y en la regulación del alternativo. La unión de proteínas hnRNP a posiciones próximas a los sitios de *splicing* puede antagonizar el efecto producido por las proteínas SR, dificultando el ensamblaje del espliceosoma por

impedimento estérico. Además, se puede producir una agregación cooperativa de hnRNP desde un sitio de alta afinidad y en dirección 5' (como ocurre con la hnRNPA1) que directamente obstaculiza la entrada de cualquier otro factor de *splicing* en la región en cuestión. Por otra parte, una hnRNP puede inhibir la definición de exón cuando se une a un exón o la definición de intrón cuando se enlaza a un intrón. Finalmente, las interacciones entre proteínas hnRNP unidas a distintas porciones del pre-ARNm pueden generar bucles, los cuales tienen dos efectos principales: promover el *skipping* de un exón que queda dentro de este bucle o estimular la definición de intrón al aproximar los sitios de *splicing* 5' y 3' [42]. Las hnRNP, además de regular el *splicing*, son cruciales en todas las etapas de maduración del ARNm, incluyendo la transcripción, adición de la caperuza, poliadenilación, exportación, localización, traducción y estabilidad del ARNm [39]. Asimismo, las hnRNP también participan en la biogénesis del ADN, puesto que están involucradas en la replicación de este ácido nucleico, la reparación de daños, la remodelación de la cromatina y el funcionamiento de los telómeros [43].

Ambas familias proteicas, SR y hnRNP, son reguladores clásicos del *splicing* en *trans*, sin embargo, existen otras RBP implicadas en este proceso, la mayoría de ellas específicas de tejido o contexto. Por ejemplo, en tejido neuronal nos encontramos con PTBP1, PTBP2, SRRM4, RBFOX, NOVA2, KHDRBS3 y TDP43, factores esenciales en la neurogénesis, mientras que en el desarrollo del tejido cardíaco participan CELF1, MBNL1, RBFOX1, RBFOX2, y RBM24, entre otros. No obstante, hoy en día carecemos aún de una lista exhaustiva de todos los factores implicados en la regulación del *splicing* y sus sitios de unión, de ahí que predecir los patrones de *splicing* basándonos únicamente en las secuencias del pre-ARNm sea actualmente muy difícil [10].

1.4. *SPLICING* ALTERNATIVO

La compleja regulación del proceso de *splicing* y la composición dinámica del espliceosoma confieren una gran flexibilidad y versatilidad a este mecanismo, permitiendo generar, a partir de un único transcrito primario, diferentes transcritos, algunos de ellos funcionales, en lo que se conoce como *splicing* alternativo (AS) [9,44]. El AS fue el que dio una explicación a la gran sorpresa con la que se encontró la comunidad científica una vez finalizado el Proyecto Genoma Humano en el año 2003, en concreto, esa discrepancia entre el número de genes codificantes de proteínas anotados y el número de polipéptidos humanos observados [45]. El AS es un mecanismo postranscripcional fundamental para reprogramar los perfiles de expresión génica y ampliar la diversidad

transcriptómica, proteómica y fenotípica de los organismos eucariotas [11,46]. Este fenómeno contribuye al desarrollo y la diferenciación celular, la determinación del linaje, la adquisición y el mantenimiento de la identidad del tejido y el desarrollo de órganos y permite a las células responder a su entorno, confiriéndoles plasticidad fenotípica [10]. En humanos, el análisis del transcriptoma a gran escala reveló que hasta el 95% de los genes multi-exónicos sufren *splicing* alternativo [47].

Los eventos de AS, tal y como hemos visto en el apartado anterior, van a depender de multitud de factores, como la longitud de los exones e intrones, el orden en el que se eliminan los intrones, la fortaleza de los sitios de *splicing*, la presencia y localización de los sitios críticos, la densidad de sitios SRE y los niveles de expresión de los factores en *trans* implicados, los cuales van a regular cuidadosamente este fenómeno por tipo de célula o tejido y en diferentes períodos del desarrollo. Otras características moleculares, como la estructura de la cromatina y del ARN, la tasa de elongación del transcrito, los promotores y los sitios de poliadenilación alternativos o la presencia de nucleótidos modificados, colaboran con los elementos anteriormente mencionados en la generación de esa diversidad proteómica debida al AS [27,48].

Como acabamos de mencionar, el término empleado para referirse a cada uno de los resultados del AS es «evento» y cada evento da lugar a distintas variantes de ARN o isoformas. Respecto a los eventos de *splicing* alternativo, estos pueden ser clasificados en cinco subgrupos distintos: (i) eliminación de exón/exones alternativo/s, fenómeno también conocido como *exon skipping* o *exon cassette*, es el evento de AS más común en células de eucariotas superiores y da como resultado la eliminación completa de uno o más exones; (ii) elección entre exones mutuamente excluyentes, es decir, exones que nunca aparecen juntos en el ARNm maduro final; (iii) uso de sitios 5'ss o (iv) 3'ss alternativos, y (v) retención intrónica, poco común en vertebrados y muy frecuente en plantas, hongos y protozoos (**Figura 8**) [46]. En cuanto a las isoformas de ARN generadas por el AS, cabe destacar que algunas son codificantes de proteínas, no obstante, lo más habitual es que contengan un codón de parada 50-55 nt *upstream* de la última unión exón-exón y sean degradadas por la vía NMD [49].

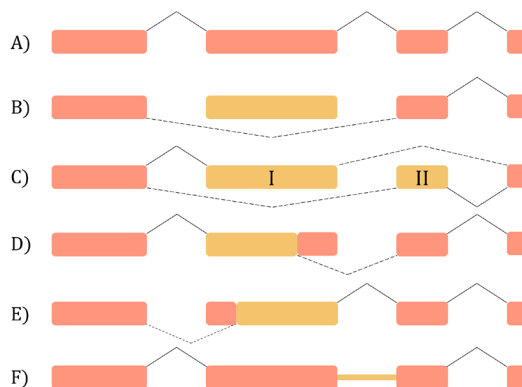


Figura 8. Eventos alternativos de *splicing*. A) *Splicing* constitutivo. B) *Exon skipping* o *exon cassette*. C) Exones mutuamente excluyentes. D) Sitio 5'ss alternativo. E) Sitio 3'ss alternativo. F) Retención intrónica.

1.5. ALTERACIONES DEL PROCESO DE *SPLICING* Y ENFERMEDAD

El *splicing* es un proceso clave de la expresión génica altamente regulado, de ahí que una pequeña alteración en cualquiera de sus protagonistas clave, el espliceosoma, los sitios de *splicing*, los SRE y los factores reguladores en *trans*, así como en otros personajes más secundarios, la estructura de la cromatina o del ARN o la tasa de transcripción, pueda desestabilizar este mecanismo. Esta desregulación del *splicing*, o bien ocasiona eventos de *splicing* aberrantes, o bien modifica la proporción de las diferentes isoformas de ARN, pudiendo desencadenar un trastorno genético como, por ejemplo, un cáncer [48]. Según *The Human Gene Mutation Database* (<https://www.hgmd.cf.ac.uk/ac/index.php>; último acceso 12 de septiembre del 2023), en torno al 9% de las variantes genéticas reportadas afectarían al *splicing*. No obstante, podría tratarse de una subestimación, teniendo en cuenta que, tradicionalmente, las mutaciones que afectan al proceso de *splicing* han sido ignoradas. En efecto, hay estudios que corroboran que muchas variantes en el ADN han sido erróneamente clasificadas como sinónimas, *frameshift*, *missense* o *nonsense*, cuando realmente están alterando los patrones de *splicing* del gen o, lo que es lo mismo, son variantes espliceogénicas [50–53]. De hecho, según algunas estimaciones, en torno al 60% de todas las mutaciones causantes de enfermedad están afectando al proceso de *splicing* de algún modo [54].

Las alteraciones en el *splicing* pueden causar la patología directamente, modificar la gravedad del fenotipo de la enfermedad o estar relacionadas con la susceptibilidad a esta. En general, estas alteraciones en el *splicing* del pre-ARNm se clasifican en dos tipos principales: las mutaciones de acción en *cis* y en *trans*. En cuanto a las mutaciones de acción en *cis*, se trata de cambios en el ADN puntuales que inducen anomalías de *splicing*

en un gen concreto. En cambio, las mutaciones en *trans* engloban variantes que afectan a la funcionalidad de las proteínas implicadas en el proceso de *splicing* y, por tanto, deberían de presentar efectos pleiotrópicos [55].

Sabemos que el *splicing* preciso del pre-ARNm depende de la presencia de una serie de secuencias en *cis* encargadas de definir los límites exón-intrón (5'ss, 3'ss, PPT y BP) y de regular este proceso (SRE). Las mutaciones puntuales localizadas en estos motivos, o aquellas que afectan a la estructura del ARNm, crean un sitio de *splicing de novo* o activan uno críptico, pueden producir transcritos deletéreos vinculados a una enfermedad [56]. Este tipo de variantes se han reportado en todos los genes asociados con enfermedades monogénicas, siendo las más habituales las que afectan a las señales de *splicing* y, más en concreto, las que modifican las posiciones intrónicas altamente conservadas, que mayoritariamente inducirán anomalías de *splicing* [23,56,57]. Aunque estas variantes ($\pm 1,2$) clásicamente se han clasificado como patogénicas, se sabe que pueden generar transcritos que mantienen el ORF original y la funcionalidad del gen [58]. Así, por ejemplo, la mutación c.68-2A>G en el gen *BRCA2* deleciona 2 aminoácidos, mientras que la variante c.108+2T>C del gen *PALB2* genera un donador atípico GC activo que produce proporciones elevadas del transcrito *full-length* [59,60]. También hay mutaciones, no tan comúnmente descritas, que afectan al PPT, como la variante c.7806-9T>G de *BRCA2* vinculada al cáncer de mama, y al BP, como es el caso de la deleción puntual c.2016-35del en el gen *FGD1* asociada al síndrome de Aarskog-Scott [61,62]. Respecto a las mutaciones que afectan a *enhancers*/silenciadores, hay múltiples ejemplos reportados en la literatura asociados a una patología. Este es el caso de la variante c.465-51T>C del gen *STK11*, la cual sirve también para entender como una alteración en el *splicing* puede modular la gravedad fenotípica de una patología. Dicha variante ubicada en el intrón 4 incrementa la unión del factor de *splicing* SRSF1, contrarrestando el efecto producido por la mutación c.597G>A (*skipping* del exón 4) y reduciendo así la severidad del síndrome de Peutz-Jeghers [63]. Otro ejemplo muy interesante es el de la mutación intrónica c.639+919G>A en el gen *GLA*, responsable de la enfermedad de Fabry, que activa un pseudoexón al destruir una secuencia silenciadora a la que se unen los factores de *splicing* hnRNPA1 y hnRNPA2B1 [64].

Respecto a las mutaciones en *trans*, este grupo incluye todas aquellas variantes que afectan a la actividad de las proteínas implicadas en el proceso de *splicing*, ya sean componentes del espliceosoma o factores reguladores de este mecanismo. Curiosamente,

aunque estos factores de *splicing* se requieren en todo el organismo, las enfermedades desencadenadas por mutaciones en los mismos suelen ser específicas de células o tejidos, afectando a tejidos tan diversos como la retina (retinitis pigmentaria), el linaje hematopoyético (síndromes mielodisplásicos o leucemias), el esqueleto (síndrome de Verheij, síndrome de Nager o síndrome de Burn-McKeown) y la médula espinal (atrofia muscular espinal o esclerosis lateral amiotrófica) [65]. También se han descubierto mutaciones en estos factores, aunque en menor medida, en varios tumores sólidos, incluido los cánceres de mama, páncreas, pulmón, y melanomas. Las mutaciones más comunes ocurren en *SF3B1*, *SRSF2*, *U2AF1* y *ZRSR2*, y suelen ser *missense*, somáticas, heterocigóticas y mutuamente excluyentes entre sí [66,67]. Es más, una expresión desregulada tanto de las hnRNP como de las proteínas SR, así como de otros factores de *splicing*, podría facilitar el inicio y la progresión tumoral, haciendo que estas proteínas actúen, por tanto, como oncoproteínas [68].

1.6. MÉTODOS USADOS PARA EL ANÁLISIS DEL *SPLICING* ABERRANTE

Las herramientas bioinformáticas encargadas de la detección de mutaciones espliceogénicas muestran una capacidad limitada para estimar el efecto exacto que ejercen los cambios en el ADN sobre el *splicing*. Es por ello por lo que los ensayos funcionales de las variantes candidatas continúan siendo esenciales para desvelar su impacto real [69]. El análisis funcional de variantes de ADN potencialmente espliceogénicas de genes responsables de una enfermedad se puede llevar a cabo, o bien empleando muestras de ARN de pacientes portadores (generalmente, ARN de linfocitos), o bien mediante minigenes híbridos [70]. Cuando la muestra de ARN del paciente es inaccesible, los minigenes híbridos representan una herramienta genética versátil, simple, robusta, de gran capacidad y de alta sensibilidad para la caracterización preliminar de posibles variantes espliceogénicas [71–73]. En concreto, los minigenes híbridos son vectores reporteros de *splicing* que incluyen un promotor fuerte, el primer (V1) y el último (V2) exón de la reacción de *splicing* y, entre ambos exones, un sitio de clonación múltiple donde se incorporan los exones del gen diana flanqueados por unos 200-300 nt intrónicos. Una vez construido y validado el minigén *wild type* (WT), se introducen por mutagénesis dirigida las mutaciones de interés, los constructos WT y mutantes se transfectan transitoriamente en células eucariotas y, tras extraer y purificar el ARN celular, su patrón de *splicing* se analiza mediante RT-PCR con primers específicos de los exones del vector. Es así como esta herramienta nos permite estudiar *ex vivo* cómo

funciona el *splicing* de un gen y, por tanto, determinar que efecto ejercen sobre este proceso las variantes de interés sin tener que recurrir a muestras de pacientes portadores (Figura 9).

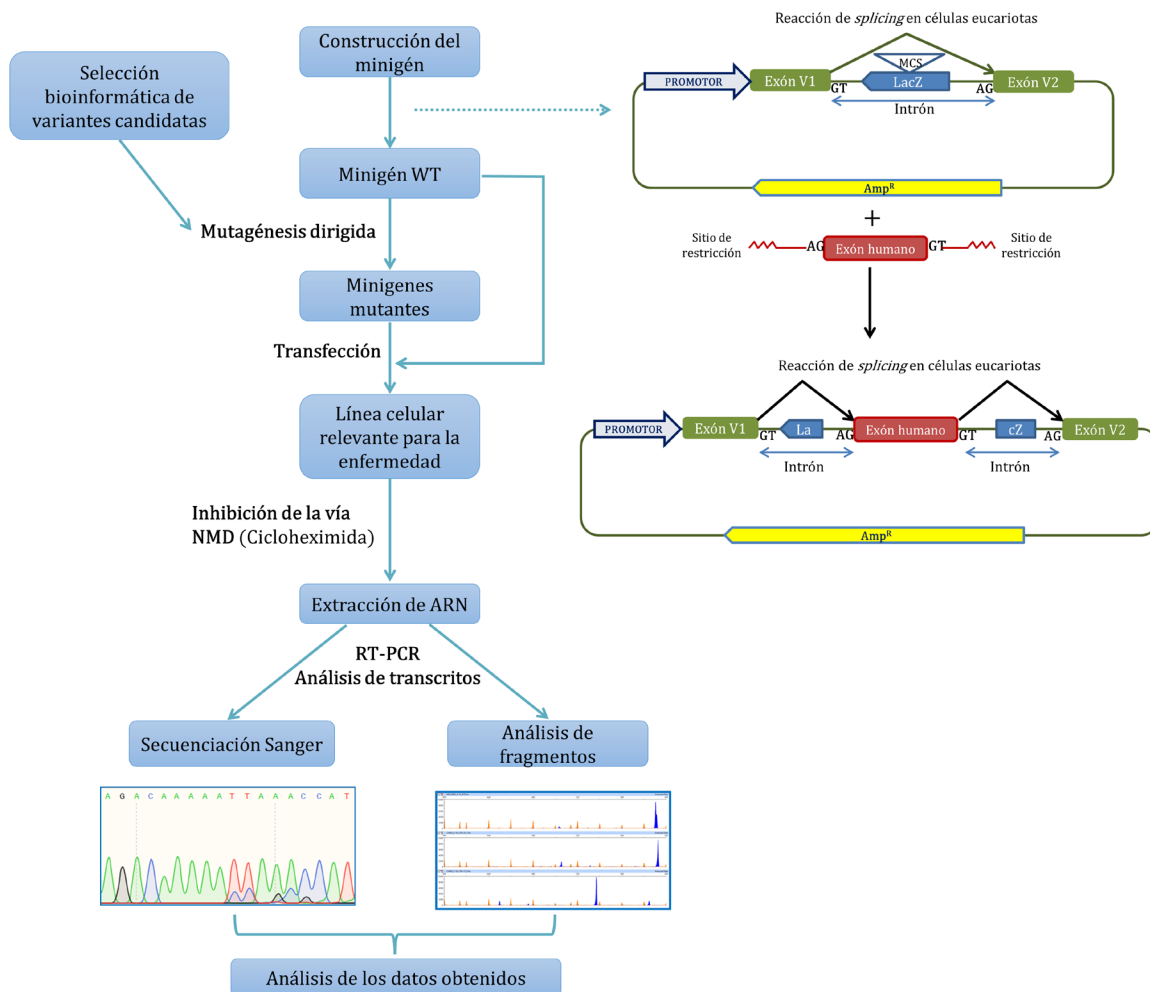


Figura 9. Protocolo del estudio funcional de variantes candidatas de alterar el *splicing* mediante minigenes híbridos. El ensayo básico incluye los siguientes pasos: (1) construcción del minigén WT; (2) recopilación de variantes por filtrado bioinformático; (3) mutagénesis dirigida; (4) transfección de los minigenes WT y mutantes; (5) inhibición del NMD y purificación del ARN; (6) secuenciación del transcrito y análisis de fragmentos mediante electroforesis capilar fluorescente, y (7) interpretación de los datos. MCS, sitio de clonación múltiple.

CÁNCER DE MAMA

Parte 2

El cáncer de mama (CM) es conocido por la humanidad desde la antigüedad, mencionándose en casi todos los períodos de la historia registrada. Concretamente, el primer registro documentado que tenemos al respecto es el papiro egipcio de Edwin Smith, tratado médico que data del siglo XVII a.C. Entre las 48 historias clínicas relatadas en este tratado, describe ocho casos de tumores o úlceras de mama que fueron extirpados por ablación quirúrgica o cauterizados. Según el médico egipcio que redactó dicho papiro, no había ningún tratamiento para esta enfermedad [74,75].

Más de 1000 años después, Hipócrates de Cos (460-375 a.C.), considerado «padre de la medicina occidental», y sus discípulos fueron los primeros en separar la medicina del teísmo, la superstición y la demonología, y defendieron la idea de que las enfermedades se iniciaban por causas naturales. Según estos médicos griegos, el CM era una enfermedad sistémica que afloraba como consecuencia de un exceso de fluido o humor denominado «bilis negra». A Hipócrates se le otorga el mérito de ser el primero en acuñar el término *karkinos*, que en griego significa «cangrejo», para referirse a esas masas malignas. Los médicos hipocráticos también fueron los primeros en emplear la palabra *karkinoma*, así como el término *onkos*, que se traduce del griego antiguo como «masa» o «inflamación» y es raíz del vocablo oncología [76,77].

Posteriormente, en la época del Imperio romano, el médico griego Claudio Galeno de Pérgamo (131-203 d. C.) hizo una categorización detallada de los crecimientos anormales. Sobre todo, describió de forma minuciosa los tumores de seno, pues eran los más prevalentes. Al contrario que Hipócrates, Galeno abogó por la extirpación del tumor de mama, proponiendo una forma de lumpectomía en la cual se realizaba una incisión alrededor del tumor y se retiraba solamente la masa maligna [77,78].

Durante la Edad Media, el progreso médico estuvo inextricablemente entrelazado con las filosofías religiosas emergentes, las cuales respaldaban la idea de la curación a través de la fe y los milagros. Este periodo de letargo en el avance médico y científico

finalizó con la llegada del siglo XVI, que supuso el resurgimiento de la cirugía, dando un impulso a la investigación oncológica, en especial del CM [77,79,80]. No obstante, no fue hasta el siglo XVII cuando se empezaron a obtener los primeros resultados favorables, al poder relacionar la enfermedad con los nódulos linfáticos axilares y ser capaces de extraer el tumor junto con estos nódulos linfáticos y los músculos pectorales; aunque la mortalidad se mantuvo debido a las muertes por sepsis, abrió el camino a la mastectomía moderna. Por otra parte, el *boom* artístico de esta época renacentista, en la que los pintores representaban pechos femeninos con mucha frecuencia, dejó plasmado públicamente el CM en gran parte de sus obras artísticas (**Figura 10**) [81].



Figura 10. Recortes de pinturas renacentistas que proporcionan representaciones de patologías mamarias malignas. De izquierda a derecha: *Retrato de una joven (La Fornarina)*, Raffaello Sanzio 1518/19, óleo sobre tabla, Galería Borghese, Roma; *La Noche*, Michele di Rodolfo del Ghirlandaio 1555/65, óleo sobre tabla, Galería Colonna, Roma; *La alegoría de la fortaleza*, Maso da San Friano 1560/62, óleo sobre tabla, Galería de la Academia, Florencia; *Las tres gracias*, Peter Paul Rubens 1636/39, óleo sobre tabla, Museo del Prado, Madrid; *Betsabé con la carta de David*, Rembrandt Harmenszoon van Rijn 1645, óleo sobre lienzo, Museo del Louvre, París.

En el siglo XIX la cirugía alcanzó su máximo apogeo, momento en que logró altos niveles de seguridad y éxito gracias a la implementación de novedosas técnicas de anestesia y antisepsia. Durante este siglo y principios del siglo siguiente, la mastectomía radical se convirtió en la operación estándar para el CM en todo el mundo y en la principal forma de abordar esta enfermedad. A lo largo del siglo XX, se produjo un abandono progresivo de esta cirugía de mama radical en favor de un mayor minimalismo quirúrgico y se desarrollaron técnicas diagnósticas y tratamientos de gran valor en el mundo de la oncología, tales como la mamografía, la radioterapia, la quimioterapia y las terapias dirigidas. Este periodo también supuso una revolución a nivel genético y molecular, lo cual permitió la clasificación del CM desde otros puntos de vista y la demostración de la existencia de un componente hereditario en esta enfermedad [77].

2. GENERALIDADES DEL CÁNCER DE MAMA

El cáncer es un conjunto heterogéneo de enfermedades, todas ellas caracterizadas por tener un origen en común: la aparición de células anormales que crecen y se dividen a gran velocidad y de forma descontrolada, y que adquieren la capacidad de migrar, invadir y dañar tejidos y órganos. Durante los últimos años, se han ido desvelando las marcas de identidad de los tumores malignos, siendo las más relevantes: la exhibición de un potencial replicativo ilimitado (inmortalidad celular), la evasión de la muerte celular, la insensibilidad a las señales inhibitoras del crecimiento, la evasión de la vigilancia inmunitaria, la autosuficiencia en señales de proliferación, la activación de la capacidad de invasión y metástasis, y la inducción de la angiogénesis [82].

En cuanto a su origen, al tratarse de una enfermedad compleja de penetrancia variable, el riesgo de padecer cáncer va a depender de un gran número de factores, tanto genéticos como ambientales, que a su vez varían dependiendo del tipo de cáncer. En general, los cánceres suelen ser esporádicos debidos a la acumulación de mutaciones somáticas. Sin embargo, de un 5 a un 10% se considera que tienen un origen hereditario, es decir, que son causados por mutaciones de la línea germinal que aumentan el riesgo de padecer esta enfermedad. Desde un punto de vista epidemiológico, en el año 2020 se registraron cerca de 20 millones de nuevos diagnósticos y en torno a 10 millones de fallecimientos a causa del cáncer, según los datos publicados por *The Global Cancer Observatory* (GLOBOCAN; <https://gco.iarc.fr/>) (Figura 11).

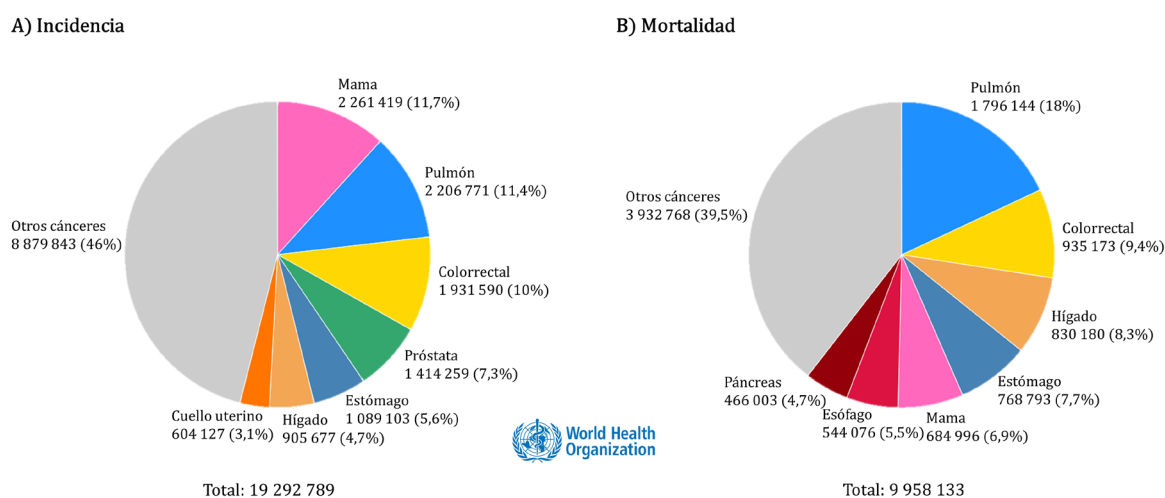


Figura 11. Gráficos de los datos de A) incidencia y B) mortalidad por tipo de cáncer a nivel mundial del año 2020. Estos gráficos han sido extraídos de GLOBOCAN.

Específicamente, el cáncer de mama es una neoplasia maligna que se origina en tejido mamario. Aunque generalmente nos referimos a él como carcinoma, pues el origen epitelial es el más habitual, también se detectan adenocarcinomas de origen glandular, sarcomas que derivan de tejido conectivo o linfomas de los ganglios linfáticos, siendo los dos últimos los diagnosticados con menor frecuencia. Esta neoplasia maligna se produce como consecuencia de la combinación de factores de riesgo no modificables, como la edad, pertenecer al sexo femenino y la presencia de ciertas mutaciones en el ADN, y modificables, como haber recibido una terapia de reemplazo hormonal, la falta de actividad física, la exhibición de índices de masa corporal muy elevados y el consumo de alcohol y tabaco [83]. El CM es el más común en mujeres en todo el mundo y, además, continúa siendo la primera causa de muerte por cáncer dentro de la población femenina a nivel mundial. Como ejemplo, en el año 2020 este tipo de cáncer fue el responsable del fallecimiento de más de 600 000 pacientes en todo el planeta (**Figura 12**). En nuestro país, en el año 2022 se diagnosticaron 34 740 nuevos casos de cáncer de mama y fallecieron 6696 pacientes a causa de esta enfermedad, de acuerdo con la información publicada por la Asociación Española Contra el Cáncer (<https://observatorio.contraelcancer.es/>). Cabe destacar que España es uno de los países de la Unión Europea que muestra una de las tasas de mortalidad más bajas para este tipo de cáncer [84].

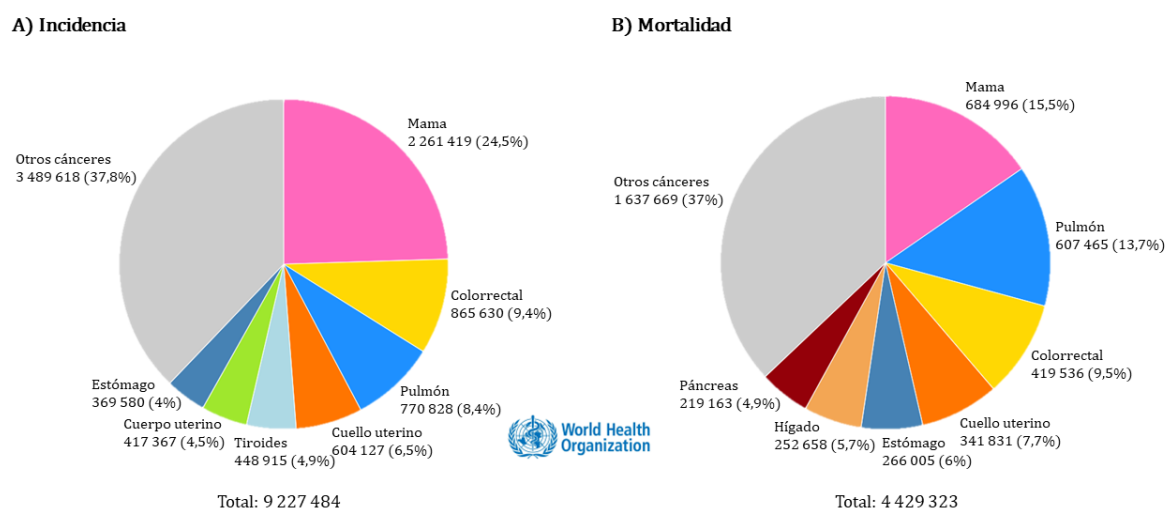


Figura 12. Gráficos de los datos de A) incidencia y B) mortalidad por tipo de cáncer en la población femenina a nivel mundial del año 2020. Estos gráficos han sido extraídos de GLOBOCAN.

La forma más directa y efectiva de plantarle cara a esta enfermedad es mediante el diagnóstico precoz, estableciendo programas de cribado destinados a mujeres que cumplen una serie de criterios de inclusión establecidos, y una vez diagnosticado, un punto clave es la correcta estratificación del tumor de seno, la cual va a guiar la forma de abordar esta patología. El cáncer de mama se puede clasificar de diversas formas en función de los rasgos del tumor que se tengan en cuenta. Así, se puede catalogar a nivel morfológico según su procedencia histológica, grado de diferenciación de sus células y estadio, o a nivel molecular, basándonos en los marcadores moleculares pronósticos (determinan el comportamiento del tumor) y predictivos (presagian la respuesta a la terapia específica) detectados. Describir de forma precisa un cáncer de mama desde todos estos aspectos es crucial a la hora de calcular correctamente el pronóstico y establecer el tratamiento óptimo para hacer frente a esta enfermedad [85].

Desde un punto de vista morfológico, los carcinomas invasivos se clasifican en distintos subtipos de acuerdo una amplia gama de criterios, incluidos las características citológicas, el patrón y la arquitectura de crecimiento del tumor, la actividad secretora y las características del estroma, entre otros. Los principales tipos de tumores invasivos son los carcinomas infiltrantes ductales, lobulillares, mucinosos (coloides), tubulares, cribiformes, medulares, escamosos y papilares. El carcinoma ductal infiltrante, que se origina en las células que recubren el ducto galactóforo, es el subtipo histológico más común, el cual engloba el 70-80% de todas las lesiones invasivas [86]. Por otro lado, para determinar el grado de diferenciación de las células, se sigue el sistema de clasificación de Nottingham, el cual cataloga esta enfermedad en 3 niveles, del I al III, siendo este último el valor más indiferenciado y de peor pronóstico [87]. En cuanto al estadio del CM, este se define siguiendo las normas del sistema de estadificación TNM, un método estandarizado para clasificar la malignidad de un neoplasma según una lista de factores que involucran el tamaño del tumor primario (T), si afecta o no a los ganglios o nódulos linfáticos (N), y la aparición o no de metástasis distal (M) [88].

Esta clasificación que atiende solo a parámetros histológicos es insuficiente para reflejar la heterogeneidad biológica del CM en su totalidad. Otra forma de establecer el pronóstico y la progresión de este tipo de cáncer es jerarquizándolo a nivel molecular o, en otras palabras, clasificarlo en base a genes específicos y sus productos (ARN y proteínas). El abordaje molecular tradicional consiste en la identificación, mediante inmunohistoquímica o hibridación *in situ*, de los receptores presentes en el tumor,

específicamente los receptores hormonales de estrógenos (ER) y de progesterona (PR) y el receptor ERBB2 o HER2. La combinación de los patrones de expresión génica de estos receptores, junto con los de otros biomarcadores como Ki-67 (indicativo del grado de proliferación de las células tumorales), permite estratificar molecularmente esta patología en los siguientes subtipos: luminal A (ER+, PR \geq 20%, HER2-/+,Ki-67<14%), luminal B (ER+, PR<20%,HER2-/+,Ki-67 \geq 14%), HER2 positivo (ER-,PR-,HER2+) y triple negativo (ER-,PR-,HER2-) [89,90]. De entre todos ellos, el triple negativo es la forma más agresiva [91]. Además de las clases moleculares, se han identificado algunas firmas multigénicas pronósticas y predictivas basadas en la expresión diferencial de un conjunto seleccionado de genes, las cuales ofrecen información complementaria a los parámetros clínico-patológicos tradicionales, muy útil sobre todo en estadios tempranos de la enfermedad. Finalmente, otra forma de clasificar el tumor de seno es utilizando la susceptibilidad genética o multigénica al CM [85].

La estratificación precisa del cáncer de mama facilita el manejo clínico de las pacientes y la adopción de medidas terapéuticas personalizadas. De esta forma, por ejemplo, a los pacientes con el subtipo luminal se les aplica una terapia hormonal, mientras que a aquellos en los que está sobreexpresado el receptor HER2 se les administra una terapia biológica frente a este biomarcador [83].

2.1. CÁNCER DE MAMA HEREDITARIO

El cáncer de mama hereditario (CMH) representa del 5 al 10% de todos los casos de CM. Esta patología es atribuible a mutaciones deletéreas de la línea germinal en genes de riesgo, presenta mayormente un patrón de herencia autosómico dominante y exhibe una penetrancia incompleta. Interesantemente, el CMH muestra una marcada heterogeneidad fenotípica y genotípica, existiendo múltiples síndromes de cáncer hereditario y una amplia gama de genes que predisponen a esta enfermedad, aspecto muy importante a tener en cuenta sobre todo en el manejo clínico de los pacientes [92]. En general, lo más habitual en estas familias es la sobrerrepresentación de cánceres de ovario, cánceres de mama bilaterales y cánceres de mama masculinos, los cuales suelen manifestarse a edades muy tempranas.

El concepto de CM como enfermedad hereditaria fue descrito por primera vez por el médico francés Pierre Paul Broca en el año 1866. Su esposa padecía cáncer de mama desde una edad temprana y, después de estudiar la historia familiar, identificó cuatro

generaciones con varias mujeres afectadas [93,94] (**Figura 13**). Este informe «Broca» se considera el primero de muchos en sugerir la heredabilidad del cáncer de mama [95–98]. Más de un siglo después de esta observación, comenzaron a aparecer en la literatura descripciones de genealogías que se correspondían con distintos síndromes vinculados con el cáncer de mama hereditario, los cuales se caracterizan por diferentes combinaciones de tumores de aparición temprana; el más común de ellos es el síndrome de cáncer hereditario de mama y ovario, descrito por primera vez durante la década de 1970 [99,100].

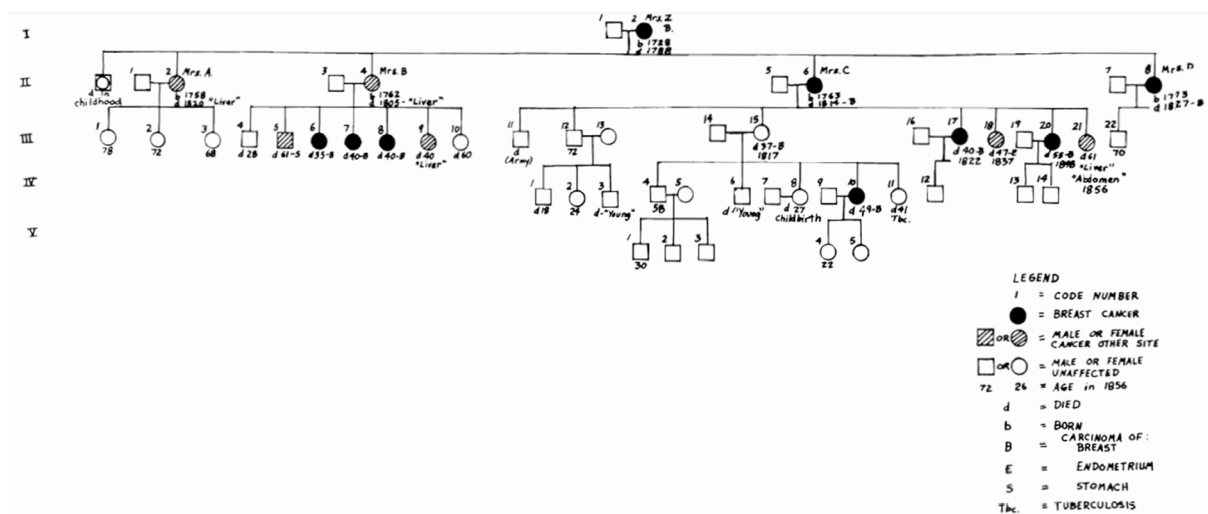


Figura 13. Árbol genealógico de Broca. Árbol genealógico de la familia de Mr. Z. recogido en el libro de Pierre Paul Broca de título *Traité des tumeurs*, publicado el año 1866 [93,94].

En 1988, tras el análisis de 1579 familias de alto riesgo con múltiples individuos afectados, se demostró que el agrupamiento de cáncer en estas familias se explicaba a través de un modelo de transmisión autosómico dominante con un alelo de predisposición altamente penetrante [101]. Dos años más tarde, los estudios de ligamiento revelaron que el locus de susceptibilidad a CM mapeaba en el cromosoma 17 [102]. Posteriormente, se descubrió que *BRCA1* era el gen causal y, poco después, se identificó *BRCA2*, localizado en el cromosoma 13, como el segundo gen de predisposición a CM [103,104]. Hasta ese momento, las pacientes de alto riesgo recibían asesoramiento genético basado únicamente en la historia familiar de cáncer. El descubrimiento de *BRCA1* y *BRCA2* y la rápida aparición de pruebas genéticas cambió por completo el panorama, facilitando enormemente el diagnóstico del CMH [92,105]. Tanto *BRCA1* como *BRCA2* son genes supresores de tumores que desempeñan un papel central en la reparación de roturas de doble cadena (DSB) en el ADN por recombinación homóloga (HR), proceso clave en el mantenimiento de la estabilidad genómica. Además, se ha demostrado que los

BRCA también están implicados en otras actividades celulares esenciales, como el control de la dinámica del centrosoma, la segregación cromosómica, la citocinesis y la estabilización del genoma temporal y espacialmente durante el ciclo celular [106]. No obstante, estos genes *BRCA* explicarían solo en torno al 25% de todos los casos de CMH [107].

Durante los últimos años, se han ido incorporando a la lista otros genes de susceptibilidad o predisposición a este tipo de cáncer (genes CM). La mayoría de ellos son supresores tumorales que también están participando en esta ruta de reparación del ADN por HR. En particular, dos estudios caso-control publicados recientemente, BRIDGES y CARRIERS, donde se evaluaron los datos de secuenciación de unas 180 000 mujeres, incluidas pacientes con CM no seleccionadas y controles (individuos no afectados), identificaron un riesgo elevado de cáncer de mama en mujeres con variantes patogénicas en otros 6 genes, además de los *BRCA*: *PALB2*, *ATM*, *CHEK2*, *RAD51C*, *RAD51D* y *BARD1* [108,109] (**Figura 14**). Estas investigaciones revelaron que las variantes truncantes en los genes *BRCA1*, *BRCA2* y *PALB2* se asocian a un riesgo elevado de CM (riesgos relativos entre 5,0 y 10,6), mientras que *ATM*, *CHEK2*, *RAD51C*, *RAD51D* y *BARD1* confieren un riesgo moderado (1,7-2,5) [110]. Además de incrementar el riesgo de CM, las variantes deletéreas en estos 8 genes se vinculan a otros tipos de neoplasias malignas, como, por ejemplo, cáncer de ovario (*BRCA1*, *BRCA2*, *PALB2*, *ATM*, *RAD51C*, *RAD51D*), páncreas (*BRCA1*, *BRCA2*, *PALB2*, *ATM*), próstata (*BRCA2*, *CHEK2*, *ATM*), estómago (*BRCA1*, *BRCA2*, *ATM*), colorrectal (*ATM*) y tiroides (*CHEK2*) [111–116]. Asimismo, existen otros genes relacionados con el CMH, tales como los principales responsables de síndromes hereditarios menos frecuentes que también predisponen a cáncer de mama, como el síndrome de Li-Fraumeni, atribuible a mutaciones en el gen *TP53*; el síndrome de Cowden, asociado a variantes del gen *P TEN*; el síndrome de Peutz-Jeghers, achacable a cambios en el gen *STK11*; el cáncer gástrico difuso hereditario, ligado a mutaciones inactivantes en el gen *CDH1*, y la neurofibromatosis tipo 1, vinculada a cambios en el ADN del gen *NF1* [107]. El modelo que mejor explica como una única copia mutada de estos genes supresores tumorales es suficiente como para desencadenar un CM es el de Knudson, la denominada teoría *two-hit*. Según este modelo, la predisposición hereditaria al cáncer implica una mutación en la línea germinal, mientras que la tumorigénesis requiere de una segunda variante somática [117].

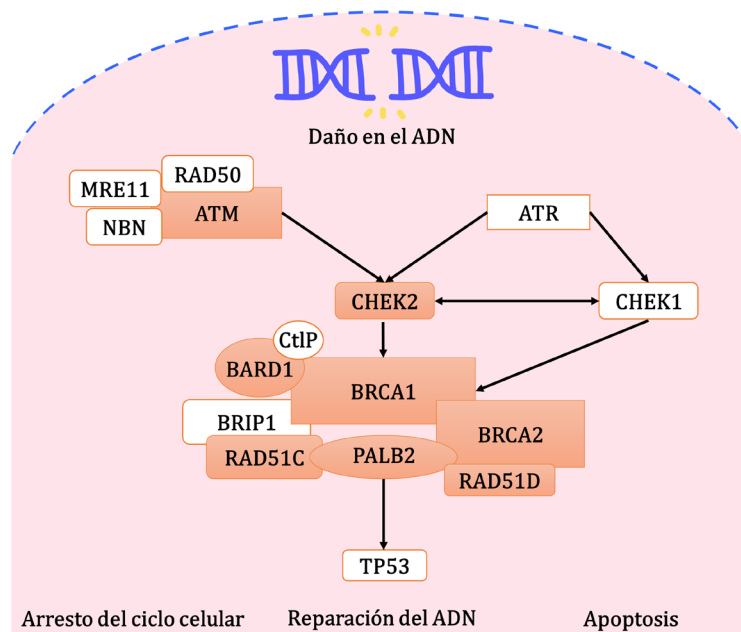


Figura 14. Representación esquemática de los genes de susceptibilidad a cáncer de mama que participan en la vía de reparación de DSB por HR. Dentro de las figuras geométricas de color naranja aparecen los principales genes de riesgo a cáncer de mama, de acuerdo con los resultados obtenidos en los proyectos BRIDGES y CARRIERS [91].

Con todos los datos recopilados hasta este momento, las variantes patogénicas en los genes de susceptibilidad de penetrancia moderada y alta explican aproximadamente el 25-30% del riesgo de CMH. Otro 18% es atribuible al riesgo poligénico, es decir, a la combinación de diferentes alelos de predisposición comunes de penetrancia baja (riesgo relativo <1,5) [107]. Esta información nos lleva a concluir que, hoy por hoy, alrededor de un 50% de los casos de CMH quedan sin explicar a nivel genético.

2.2. MANEJO CLÍNICO DE PACIENTES CON CÁNCER DE MAMA HEREDITARIO

En la actualidad, los avances en oncología de precisión han incrementado las indicaciones para la realización de pruebas genéticas de la línea germinal con el fin de guiar el tratamiento, tanto sistémico como quirúrgico, aumentando, lógicamente, la demanda de estas pruebas. Concretamente, el proceso de atención a pacientes con CMH incluye un primer paso de identificación de las familias de alto riesgo, seguido de la prestación de los servicios de asesoramiento y pruebas genéticas y, finalmente, el proceso se cierra con la comunicación de los resultados, el correcto seguimiento de las familias y/o la aplicación de las acciones preventivas recomendadas de acuerdo con los resultados obtenidos [118].

El análisis genético del CMH requiere realizar una selección inicial precisa de aquellas familias de alto riesgo en las que está indicado el estudio genético, tal y como acabamos de mencionar. En España, la selección de estas familias se basa en los criterios establecidos por la Sociedad Española de Oncología Médica (SEOM; <https://seom.org/>), los cuales están recogidos en la **Tabla 1**. Estos criterios están asociados con una probabilidad de detección de variantes patogénicas $\geq 10\%$ y se basan en factores de riesgo clínicos como la edad de aparición, el estado de los receptores hormonales, la ascendencia con mutaciones fundadoras y los antecedentes personales y familiares de cáncer [119].

Tabla 1. Criterios de selección para el testado germinal de mutaciones asociadas al cáncer de mama establecidos por la SEOM. En esta tabla se detallan los criterios de selección de familias para su inclusión en las pruebas genéticas de CMH, los cuales están recogidos en la guía clínica de la SEOM de cáncer de mama y ovario hereditario del año 2019 [119].
Independientemente del historial familiar:
• Mujeres con cáncer de mama y ovario sincrónico o metacrónico
• Cáncer de mama ≤ 40 años
• Cáncer de mama bilateral (el primero diagnosticado ≤ 50 años)
• Cáncer de mama triple negativo ≤ 60 años
• Cáncer de ovario epitelial no mucinoso de alto grado (o cáncer de trompa de Falopio o peritoneal primario)
• Ascendencia con mutaciones fundadoras
• Mutación somática <i>BRCA</i> detectada en cualquier tipo de tumor con una frecuencia alélica $>30\%$ (si se conoce)
• Pacientes con cáncer de mama metastásico HER2 negativo aptas para considerar el tratamiento con inhibidores de PARP
2 o más familiares de primer grado con cualquiera de las siguientes características de alto riesgo:
• Cáncer de mama bilateral + otro cáncer de mama <60 años
• Cáncer de mama <50 años y cáncer de próstata o páncreas <60 años
• Cáncer de mama masculino
• Cáncer de mama y ovario
• Dos casos de cáncer de mama diagnosticados antes de los 50 años
3 o más familiares directos con cáncer de mama y/o cáncer de ovario y/o, cáncer de páncreas o de próstata

Una vez identificadas las familias de alto riesgo, el siguiente paso consiste en derivarlas a una consulta de asesoramiento genético. El asesoramiento genético en cáncer hereditario es un proceso muy complejo, destinado a ayudar a los pacientes a comprender y prepararse para las consecuencias médicas, psicológicas y familiares de las contribuciones genéticas a la enfermedad. Posiblemente, la evaluación del riesgo de cáncer sea la parte más importante de este proceso a la hora de establecer un plan de

manejo clínico adecuado para estas familias, que puede incluir un incremento de la frecuencia de pruebas de detección del CM, opciones de prevención (quimioprevención o cirugía profiláctica) y recomendaciones sobre el estilo de vida [120]. De entre todos los elementos necesarios para el cálculo de riesgo, el examen genético es el más relevante. Hasta hace poco, el testado de los genes *BRCA* se realizaba mediante métodos arduos, laboriosos y muy costosos [121]. No obstante, la enorme reducción de los costes y las mejoras recientes aplicadas en las técnicas de secuenciación de nueva generación (NGS) han permitido la introducción de las pruebas de panel multigénico, simplificando en gran medida el estudio de enfermedades hereditarias complejas, como el cáncer. De hecho, la secuenciación de paneles de genes CM conocidos o candidatos se ha convertido en la técnica predilecta para la identificación de variantes clínicamente accionables en individuos de alto riesgo [122]. La alta sensibilidad de la secuenciación NGS permite la detección de miles de variantes nuevas, lo que a su vez conlleva una desventaja consistente en el incremento en el número de variantes de significado clínico incierto (VUS) identificadas [123].

En paralelo, se han desarrollado una serie de normas para la clasificación clínica de las variantes detectadas. En concreto, en el año 2015 las asociaciones ACMG (*American College of Medical Genetics*), AMP (*Association for Molecular Pathology*) y CAP (*College of American Pathologists*) publicaron una guía para estandarizar los criterios de clasificación de las variantes genéticas [124]. En ella, se establecen 28 tipos diferentes de evidencias asociadas a características inherentes a cada variante, como la frecuencia alélica, los análisis funcionales, las predicciones *in silico*, etc. A estas evidencias se les asigna un código que valora su contribución al nivel de patogenicidad, a saber: independiente (BA1), fuerte (BS1-4), o leve [BP1-6]) si la evidencia tiene un impacto benigno, y muy fuerte (PVS1), fuerte (PS1-4), moderado (PM1-6), o leve (PP1-5) si la evidencia tiene un papel patogénico. Estos códigos funcionan como una especie de puntuación, combinándose entre sí para clasificar las variantes en 5 niveles distintos: clase I, benigna (B; probabilidad de patogenicidad <0,001), clase II, probablemente benigna (LB; probabilidad de patogenicidad entre 0,001 y 0,049), clase III, variante de significado clínico incierto (probabilidad de patogenicidad entre 0,05 y 0,949), clase IV, probablemente patogénica (LP; probabilidad de patogenicidad entre 0,95 y 0,99) y clase V, patogénica (P; probabilidad de patogenicidad <0,99). De acuerdo con estas normas, las variantes *nonsense* y *frameshift*, la delección de uno o más exones y los cambios en el ADN que afectan a las posiciones intrónicas $\pm 1,2$ de los sitios de *splicing* y al codón de iniciación de la

traducción, son mutaciones consideradas prácticamente de forma automática como P/LP. De facto, son el único tipo de variantes a las que se les puede asignar un nivel muy fuerte de patogenicidad basándose únicamente en las predicciones computacionales, o bien se pueden catalogar como LP simplemente al aplicarles al menos un criterio moderado, como coincidir con el fenotipo de un paciente [125].

La detección de mutaciones patogénicas en genes CM facilita el manejo clínico de las familias con esta enfermedad, de hecho, existen pautas de manejo establecidas por tipo de tumor y gen implicado, como, por ejemplo, las de la *National Comprehensive Cancer Network* (NCCN *Clinical Practice Guidelines in Oncology*). Específicamente, el manejo de familias portadoras de variantes truncantes en genes CM puede incluir: (i) recomendaciones sobre el estilo de vida; (ii) toma de decisiones quirúrgicas como la mastectomía y la salpingo-ooforectomía profilácticas; (iii) quimioprevención primaria con tamoxifeno u otros moduladores selectivos de los receptores de estrógenos; (iv) no uso de radioterapia en pacientes con síndrome de Li-Fraumeni (*TP53* alterado), ya que induce significativamente la aparición de tumores malignos en estos pacientes; (v) incremento de la vigilancia en portadoras (examen clínico de mama, mamografías, resonancia magnética y ecografías), y (vi) tratamiento con iPARP para portadoras de mutaciones en *BRCA1/2* con cáncer de mama metastásico (NCCN *Guidelines: Breast Cancer Risk Reduction Version 1.2023, Breast Cancer Screening and Diagnosis Version 1.2023 & Breast Cancer Version 4.2023*) [126,127]. No obstante, la detección de VUS en familias CMH equivale, en la práctica, a un resultado negativo del test genético, puesto que son variantes cuyo impacto sobre la función génica es desconocido y, por eso, no queda clara cuál es su asociación con la enfermedad. Es por ello por lo que los cálculos de riesgo en las pacientes portadoras de VUS se basan exclusivamente en los antecedentes personales y familiares de cáncer [128]. Ante este desafío, es prioritario determinar el rol de estas variantes en la enfermedad.

2.3. GENES DE SUSCEPTIBILIDAD A CÁNCER DE MAMA/OVARIO INVESTIGADOS

En esta tesis, nos hemos centrado básicamente en el análisis funcional de *splicing* de las variantes candidatas de los genes CM *CHEK2* y *RAD51C*, y en el estudio de la regulación de exones sometidos a *splicing* alternativo (exones 8 y 10 de *CHEK2*) y de exones especiales de los genes *BARD1* (exón 4-macroexón) y *BRIP1* (exón 1-donador GC).

▪ *RAD51C*

Las variantes deletéreas en el gen *RAD51C*, también conocido como *FANCO*, otorgan un riesgo moderado de CM (riesgo relativo: $\approx 1,6$; riesgo absoluto: 17-30% a los 80 años) y están principalmente asociadas a los tumores ER negativos (riesgo relativo: 5,7). Las mutaciones en este gen también incrementan el riesgo de cáncer de ovario (riesgo relativo: 5,6) [108,109,129,130]. *RAD51C* se sitúa en el cromosoma 17 (17q22) y se expande a lo largo de ≈ 43 kilobases (kb; chr17: 58 692 573-58 735 611; GRCh38). El transcrito más expresado de *RAD51C* (NM_058216/ENST00000337432.9) contiene 9 exones y codifica una proteína de 376 aminoácidos y unos 42 kDa. Estructuralmente, la proteína *RAD51C* presenta las siguientes regiones críticas para su función: (i) dominios Walker-A (p.125-132) y Walker-B (p.238-242) esenciales para la hidrólisis de ATP, (ii) cadena $\beta 1$ (p.120-124), cadena $\beta 2$ (p.154-159), cadena $\beta 3$ (p.207-212), cadena $\beta 4$ (p.237-242), cadena $\beta 5$ (p.280-285), cadena $\beta 6$ (p.312-318), cadena $\beta 7$ (p.321-326), cadena $\beta 8$ (p.336-342) y cadena $\beta 9$ (p.345-347), y (iii) la señal de localización nuclear (p.366-370) [131].

RAD51C es un gen supresor de tumores perteneciente a la familia de parálogos de *RAD51*, que está compuesta por seis proteínas: *RAD51B*, *RAD51C*, *RAD51D*, *XRCC2*, *XRCC3* y *SWSAP1* [132,133]. Estos parálogos de *RAD51* comparten del 20-30% de identidad a nivel de aminoácidos con *RAD51* y entre sí, excepto *SWSAP1*. Específicamente, las secuencias conservadas se encuentran sobre todo en los dominios Walker A y B, que les confieren una actividad ATPasa débil. Estos parálogos de *RAD51* ejercen sus funciones organizándose en distintos complejos proteicos estables; los clásicos son el denominado BCDX2 (*XRCC2* y *RAD51 B, C y D*), y CX3 (*RAD51C* y *XRCC3*). Además, *RAD51C* es integrante de un complejo adicional con *PALB2* y *BRCA2* (complejo HR) que promueve la reparación a través del proceso de recombinación homóloga. Finalmente, *SWSAP1* interacciona con diversas proteínas formando el complejo Shu. Estos cuatro complejos desempeñan funciones clave en el proceso de reparación del ADN por HR [134,135].

En concreto, se ha propuesto que *RAD51C* ejerce las siguientes funciones [135-139]:

- (i) Contribuir a la translocación de *RAD51* desde el citoplasma hacia el núcleo en respuesta al daño en el ADN.
- (ii) Asegurar el reclutamiento de *RAD51* a las regiones dañadas y la estabilización de los filamentos de *RAD51* sobre el ADN monocatenario.

- (iii) Facilitar la fosforilación/activación de CHK2 dependiente de ATM.
- (iv) Intervenir en la resolución de las uniones de Holliday para finalizar el proceso de reparación del ADN por HR.
- (v) Responder a enlaces entrecruzados en el ADN (ICL), ruta de reparación que está alterada en la anemia de Fanconi, al asociarse a otros parálogos o formando parte del complejo HR junto con PALB2/FANCN y BRCA2/FANCD1 (**Figura 15**).
- (vi) Participar en la respuesta al estrés replicativo interaccionando con otros parálogos.

Sin embargo, a pesar de que *RAD51C* es un gen central tanto en el proceso de recombinación homóloga como en la evasión del colapso replicativo, sus funciones exactas siguen sin estar claras.

En un principio, *RAD51C* se reportó como un gen de susceptibilidad hereditaria a cáncer de mama y ovario [129]. Al mismo tiempo, se notificó una mutación bialélica en *RAD51C* en una familia con anemia de Fanconi, enfermedad que presenta un patrón de herencia autosómico recesivo y en la que el efecto clínico es una depleción progresiva de la función de la médula ósea, malformaciones congénitas y un alto riesgo de desarrollar tumores sólidos y hematológicos a una edad muy temprana [140]. Desde el año 2010, numerosos trabajos han corroborado la asociación de mutaciones patogénicas de la línea germinal en este gen con el cáncer de ovario [141]. No obstante, la conexión entre *RAD51C* y cáncer de mama tardó una década en confirmarse, en concreto, gracias a dos grandes estudios epidemiológicos, BRIDGES y CARRIERS [108,109].

Según las recomendaciones de la NCCN, las mujeres que portan una variante deletérea en *RAD51C* deben someterse a mamografías anuales a partir de los 40 años, pudiendo considerarse la realización de exámenes de resonancia magnética de mama con contraste. En este caso, no se recomienda la mastectomía profiláctica, pero sí la salpingo-ooforectomía bilateral preventiva entre los 45 y 50 años, o incluso antes, en función de la historia familiar (NCCN *Guideline: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic Version 1.2024*) [142].

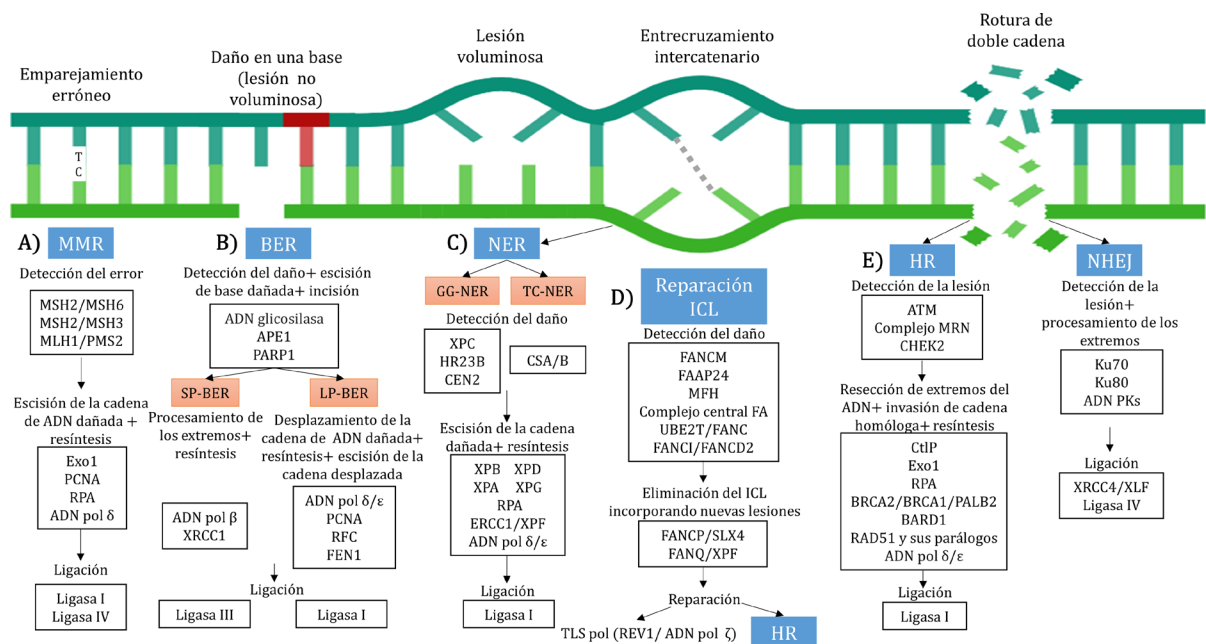


Figura 15. Resumen de las principales rutas moleculares de reparación de daños en el ADN. Esquema de las distintas lesiones que pueden generarse en el ADN y las vías moleculares más relevantes para repararlas. A) La vía de reparación por emparejamiento erróneo (MMR; *mismatch repair*) corrige básicamente las bases mal incorporadas en el ADN. B) La reparación por escisión de base (BER; *base excision repair*) es una ruta empleada normalmente para solventar roturas de simple cadena y bases dañadas que no generan una lesión voluminosa. Dicha ruta se divide en dos subvías: *short-patch* BER y *long-patch* BER. Las roturas de simple cadena pueden repararse también a través de la vía SSB (single-strand break repair) que, aunque no se muestra en esta imagen, es muy similar a la BER. C) Para las lesiones voluminosas, se sigue mayormente la reparación por escisión de nucleótidos (NER; *nucleotide excision repair*), que elimina la base dañada y varios nucleótidos adyacentes. Este mecanismo cuenta con dos vías para el reconocimiento de la lesión: *global genome* NER (GG-NER) y *transcription-coupled* NER (TC-NER). D) La reparación de ICL (*interstrand cross-link*), también conocida como vía FA, resuelve los enlaces covalentes intercatenarios. E) Los principales mecanismos para la reparación de DSB son la HR y la unión de extremos no homólogos (NHEJ; *non-homologous end joining*). Existen dos rutas adicionales, la SSA (*single-strand annealing*) y la Alt-EJ (*alternative end joining*) no reflejadas en esta figura.

En ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/?term=rad51c%5Bgene%5D>, último acceso el 18 de septiembre del 2023) hay un total de 1699 variantes puntuales (<50 pares de bases [pb]) registradas del gen *RAD51C*, específicamente 869 *missense*, 105 *frameshift*, 60 *nonsense* y 63 localizadas en los sitios clásicos de *splicing*. En cuanto a su clasificación clínica, 34 (2,0%) están clasificadas como B, 462 (27,2%) como LB, 88 (5,2%) aparecen como LP, 165 (9,7%) son P y 923 (54,3%) son VUS.

▪ *CHEK2*

CHEK2 es un gen supresor tumoral que confiere un riesgo moderado a CM (riesgo relativo: $\approx 2,5$; riesgo absoluto: $\approx 25\%$ a los 80 años) y acumula casi el 25% de todas las variantes truncantes de proteína (solo superado por *BRCA2*) [108,109,143,144]. Este gen fue localizado en el cromosoma 22 (22q12.1) distribuido a lo largo de ≈ 55 kb (chr22: 28 687 743-28 742 422; GRCh38) [145]. El transcrito más expresado (NM_007194/ENST00000404276.6) consta de 15 exones con el inicio de la traducción localizado en el exón 2. Este ARNm codifica para una proteína, la serina/treonina quinasa nuclear CHK2, de 543 aminoácidos y aproximadamente unos 61 kDa, que contiene 3 dominios funcionales clave: (i) un dominio N-terminal, que se caracteriza por siete pares de residuos de serina-glutamina o treonina-glutamina (SQ/TQ), los cuales serán fosforilados por ATM y otras quinasas (SQ/TQ *cluster domain* [SCD]; p.19-69); (ii) el dominio FHA (*forkhead-associated domain*; p.92-205), que está estructurado en 11 cadenas β organizadas en forma de sándwich y media las interacciones proteína-proteína dependientes de la fosforilación de CHK2, y (iii) el dominio quinasa (KD; p. 306-501), que comprende casi la mitad de la secuencia proteica y consta de dos lóbulos que conforman un sitio de unión a ATP. Finalmente, en la región C-terminal se encuentra la señal de localización nuclear (NLS; p.515-522) [113].

CHK2 es un mediador fundamental en la señalización de DSB. En ausencia de daño en el ADN, la quinasa CHK2 permanece en su forma monomérica inactiva. En respuesta a DSB y al colapso de las horquillas replicativas, CHK2 es activada por fosforilación mediada por la proteína ATM, lo que desencadena la dimerización transitoria y la autofosforilación, seguida de la disociación de sus monómeros activos. CHK2 activada va a fosforilar diversas proteínas nucleares involucradas en la respuesta a daño en el ADN, por lo que se considera un guardián clave de la integridad del genoma. En concreto, modifica polipéptidos que participan en la ruta de reparación de DSB por HR (*BRCA1* y *BRCA2*) y también fosforila proteínas implicadas en la vía de reparación BER (*XRCC1* y *FOX-M1*) (**Figura 15**). Asimismo, promueve la relajación de la heterocromatina al actuar sobre *KAP-1*, posibilitando de esta manera el acceso a la lesión de las proteínas de respuesta a daño en el ADN, e induce la detención del ciclo celular a través de distintas vías mientras la rotura está siendo reparada. CHK2 también impulsa la apoptosis cuando el daño en el ADN es irreparable y, de este modo, evita la propagación de un genoma alterado y potencialmente dañino. Por último, esta proteína estimula, aunque no se sabe aún bien cómo, la entrada en un estado de senescencia celular, una barrera para la división de

células expuestas crónicamente a agresiones dañinas o con oncogenes activados que estimulan la replicación y generan lesiones en el ADN [146].

Inicialmente, las variantes deletéreas de la línea germinal del gen *CHEK2* se asociaron con el síndrome de Li-Fraumeni; no obstante, esta relación se ha cuestionado debido a las diferencias de fenotipo entre los pacientes con este síndrome y los portadores de variantes P/LP en *CHEK2*. Inmediatamente después, este gen se vinculó con un mayor riesgo de CM. En particular, variantes germinales de este gen están ligadas sobre todo a los tumores de mama ER positivos (riesgo relativo: 2,7), por lo que estas mujeres pueden ser candidatas de terapias hormonales antiestrogénicas, como el tratamiento con tamoxifeno, raloxifeno o inhibidores de la aromatasa. Por otro lado, la variante c.1100delC de *CHEK2*, la más comúnmente detectada, se asocia con un aumento de dos a tres veces del riesgo de CM en mujeres y de diez veces en hombres. Asimismo, variantes truncantes en *CHEK2* se han relacionado también con otros tipos de cáncer, incluidos el de próstata y tiroides [113,147].

De acuerdo con las guías de la NCCN, los individuos que portan una mutación P/LP en este gen deben someterse a mamografías anuales a partir de los 40 años y barajar la posibilidad de realizarles resonancias magnéticas de mama con contraste a partir de los 30-35 años. También se recomienda una colonoscopia a partir de los 40 años o, si hay antecedentes familiares, 10 años antes de la edad del familiar de primer grado en el momento del diagnóstico del cáncer colorrectal, y considerar el cribado del cáncer de próstata en varones también a partir de los 40 años. Finalmente, es importante destacar que no hay pruebas suficientes sobre los beneficios de la aplicación de la mastectomía profiláctica en estas pacientes, no obstante, se puede valorar la realización de este procedimiento basándonos en la historia familiar (NCCN *Guideline: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic Version 1.2024*) [110,147,148].

Finalmente, en cuanto a las mutaciones descritas en el gen *CHEK2*, la base de datos ClinVar recoge un total de 3371 variantes moleculares (<50 pb) de este gen (<https://www.ncbi.nlm.nih.gov/clinvar/?term=chek2%5Bgene%5D>; último acceso 18 de septiembre del 2023), de las cuales 1607 (47,7%) son *missense*, 370 (11%) están clasificadas como *frameshift*, 153 (4,5%) se catalogan como *nonsense* y 145 (4,3%) estarían alterando los sitios de *splicing*. De acuerdo con su significado clínico, 71 aparecen clasificadas como B, 776 como LB, 223 como LP, 497 como P y 1751 como VUS (52%).

▪ *BARD1*

El gen *BARD1* es un supresor tumoral relacionado con un riesgo moderado de CM (riesgo relativo: $\approx 1,7$; riesgo absoluto: 17-30% a los 80 años), sobre todo asociado a tumores ER negativos [108,109,149]. Este gen humano se descubrió en el año 1996 y está ubicado en el cromosoma 2 (2q35) extendiéndose a lo largo de unas 84 kb (chr2: 214 725 646-214 809 683; GRCh38). Su transcrito más abundante (NM_000465; ENST00000260947.9) consta de 11 exones, los cuales codifican una proteína de 777 aminoácidos y 87 kDa. Interesantemente, su exón número 4 es un exón largo o macroexón con 950 pb. Teniendo en cuenta que la longitud óptima de un exón para un *splicing* eficiente está entre 50 y 250 nt, el papel de los elementos reguladores va a ser crucial en el correcto reconocimiento de dicho exón [150]. A nivel estructural, la proteína BARD1 contiene en su región N-terminal un dominio RING (p. 49-100), al igual que BRCA1, a través del cual se unen formando un heterodímero estable con actividad ubiquitina ligasa E3, y su región C-terminal se compone de un dominio de repeticiones de anquirina en tándem (p. 425-555) y dos dominios BRCA1 C-terminales (p. 568-777), los cuales interactúan con diversas proteínas implicadas en la respuesta a daño en el ADN. Además, muestra señales de exportación nuclear que, junto con las NLS, son necesarias para su ubicación intracelular [151].

Funcionalmente, BARD1 actúa como un supresor de tumores en vías dependientes e independientes de BRCA1. Por un lado, la formación del heterodímero BRCA1-BARD1 es clave en las rutas de respuesta al daño en el ADN, la regulación del ciclo celular, la modulación de la estructura de la cromatina y la señalización hormonal, al controlar los niveles de receptores de estrógenos y de progesterona. La multifuncionalidad de este heterodímero BRCA1-BARD1 se debe mayormente a su actividad ubiquitina ligasa, la cual modifica múltiples proteínas implicadas en estas vías, estabilizando o alterando la funcionalidad de las de interés y marcando para degradar las innecesarias o las dañinas. Cabe destacar las ubiquitinaciones incorporadas en la RNA pol II, para inducir su degradación a través del proteosoma, y en la topoisomerasa II β , regulando así la transcripción en respuesta a daño en el ADN. También ejerce funciones clave en la reparación del ADN al interactuar directamente con proteínas que intervienen tanto en HR (BRCA2 y RAD51) como en la reparación por MMR (MSH2 y MSH6) (**Figura 15**). Por otro lado, BARD1 induce la apoptosis; inhibe de forma transitoria la poliadenilación en respuesta a daño en el ADN, evitando el procesamiento potencialmente nocivo del ARN;

actúa como supresor tumoral, al regular el factor de transcripción oncogénico NF-κB a través de su inhibidor BCL3, y recluta a BRCA1 a los sitios de daño. [116,152,153].

Tras la confirmación de la patogenicidad de variantes en *BRCA1* ubicadas en su dominio RING por estar impidiendo la interacción BRCA1-BARD1, se planteó la hipótesis de que mutaciones en *BARD1* podrían actuar del mismo modo y ser causantes de enfermedad. De hecho, dos años después de su descubrimiento, en el año 1998, se vincularon variantes germinales de *BARD1* con el cáncer de mama y ovario [154]. Su asociación con el cáncer de mama, al contrario de lo que ocurre con el de ovario, ha sido corroborada en estudios posteriores [108,109,155]. Asimismo, se ha demostrado que las isoformas aberrantes de *BARD1* muestran un papel oncogénico, habiéndose detectado en distintos tipos de cáncer, como el de pulmón, mama, colon y ovario, y su expresión se asocia significativamente con una disminución en la tasa de supervivencia [116,151].

Las pautas de manejo de la NCCN para pacientes portadoras de variantes patogénicas en *BARD1* recomiendan mamografías anuales a partir de los 40 años, pudiendo considerarse la obtención de imágenes por resonancia magnética de mama con contraste. La mastectomía preventiva no está justificada en estas pacientes, no obstante, el uso de este procedimiento puede valorarse en función de los antecedentes familiares (NCCN *Guideline: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic Version 1.2024*) [110].

ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/?term=bard1%5Bgene%5D>; último acceso 18 de septiembre del 2023) acumula un conjunto de 3384 variantes puntuales (<50 pb) de este gen, siendo las más abundantes las *missense* (1815; 53,6%), seguidas de lejos por las *frameshift* (188; 5,6%), las *nonsense* (126; 3,7%) y, por último, las localizadas en los sitios clásicos de *splicing* (76; 2,2%). Respecto a su significado clínico 1014 son B/LB, 401 P/LP y 1858 (54,9%) permanecen clasificadas como VUS.

▪ **BRIP1**

Variantes truncantes en *BRIP1*, denominado alternativamente *FANCF* o *BACH1*, no incrementan significativamente el riesgo de cáncer de mama (riesgo relativo: 1,1), aunque sí el de ovario (riesgo relativo: 4,9; riesgo absoluto: 5,8% a los 80 años) [108,109,130,156]. Este gen se localiza en el cromosoma 17 (17q23.2) ocupando unas 184 kb (chr17: 61 679 139-61 863 559; GRCh38). Su isoforma predominante (NM_032043; ENST00000259008.7) consta de 20 exones, 19 de los cuales (exones 2 a 20) son

codificantes. Curiosamente, su exón 1 no codificante usa un donador atípico GC, cuya detección probablemente dependa de la actuación de factores reguladores del *splicing* [157]. *BRIP1* se traduce en una proteína de 1249 aminoácidos y unos 141 kDa con función ATPasa ADN dependiente y helicasa 5'-3', perteneciente a la familia RecQ DEAH. Dicha proteína presenta un dominio helicasa altamente conservado, que abarca la mayor parte de su estructura, y está compuesto principalmente por 8 motivos ATPasa/helicasa (Q: p.17-25; I: p.39-57; Ia: p.245-258; II: p.385-398; III: p.610-624; IV: p.689-710; V: p.748-775; VI: p.819-836) y un centro de hierro-azufre (p.276-362), caracterizado por cuatro residuos de cisteína fuertemente conservados. Dentro de este dominio, encontramos también un sitio de unión a la proteína MLH1, que participa en la ruta de reparación MMR (**Figura 15**). Además, en su región C-terminal acumula una serie de sitios de unión a otras proteínas, incluyendo BRCA1 [158,159].

BRIP1 actúa como un supresor tumoral y desempeña un papel esencial en el mantenimiento de la integridad del genoma. Mediante su actividad helicasa ATP dependiente, *BRIP1* desenrolla una amplia gama de sustratos de ADN dúplex, así como estructuras secundarias de este ácido nucleico, como los G-cuádruplex. Gracias a esta actividad helicasa y a su interacción con una gran variedad de proteínas (como MLH1, BRCA1, CtIP, RPA, BLM y TOPBP1), *BRIP1* favorece una replicación eficiente y la progresión de las células a través de la fase S e interviene en la regulación de la respuesta al estrés replicativo y en la reparación del ADN [160,161].

Esta helicasa se identificó por primera vez en el año 2001 como una proteína que interactúa físicamente con BRCA1 y contribuye a su función en el proceso de reparación del ADN, lo cual convirtió rápidamente a *BRIP1* en un nuevo gen candidato de predisposición a CM [162]. En primer lugar, se demostró que *BRIP1* está inactivado bialélicamente en pacientes con anemia de Fanconi [163–165]. Poco después, se estableció *BRIP1* como un gen de predisposición a CM de baja penetrancia [166], aunque otros estudios posteriores más amplios y completos indican que no existe tal asociación [108,109]. En cambio, sí se ha verificado un vínculo entre variantes deletéreas de *BRIP1* y un incremento significativo del riesgo de cáncer de ovario [130,167]. Asimismo, se encontró una mutación en este gen involucrada en el síndrome de Li-Fraumeni, un trastorno autosómico dominante que confiere un alto riesgo a una amplia variedad de tumores malignos, como sarcoma de tejidos blandos, osteosarcoma, cáncer de mama, cáncer de cerebro, cáncer de glándula suprarrenal y leucemia, entre otros [168]. Cabe

destacar también la implicación de *BRIP1* en otro tipo de cánceres, como el de endometrio o el colorrectal [169,170].

Las guías de manejo de la NCCN no aportan ninguna recomendación respecto al cribado y la prevención del cáncer de mama en familias portadoras de una variante P/LP en este gen. En cambio, respecto al cáncer de ovario, aconsejan valorar la aplicación de la salpingo-ooforectomía bilateral profiláctica entre los 45 y 50 años (NCCN *Guidelines: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic Version 1.2024*) [142].

En ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/?term=brip1%5Bgene%5D>; último acceso el 18 de septiembre del 2023) hay 4769 variantes registradas de *BRIP1*, de las cuales 2441 (51,2%) son *missense*, 340 (7,1%) son *frameshift*, 199 (4,2%) son *nonsense* y 112 (2,3%) quedan ubicadas en los sitios de *splicing*. De acuerdo con su clasificación clínica, 1383 son B/LB y 668 son P/LP, mientras que 2603 están catalogadas como VUS.

2.4. *SPLICING* ABERRANTE EN CÁNCER DE MAMA HEREDITARIO

El *splicing* aberrante es la causa de muchas enfermedades humanas, incluido el cáncer. Análisis transcriptómicos han detectado >15 000 variantes de *splicing* específicas de esta patología en 27 tipos de cánceres diferentes, incluido el CM [171]. También se ha demostrado que los tumores exhiben hasta un 30% más de eventos de *splicing* alternativos que los tejidos normales [172]. En particular, las evidencias sugieren que esta desregulación del *splicing* está involucrada en la biogénesis y progresión de los tumores, incluida la proliferación celular, la apoptosis, la invasión, la metástasis tumoral, la angiogénesis y la resistencia a la quimioterapia/radioterapia [48,173]. De hecho, el AS aberrante se considera actualmente una marca distintiva del cáncer [174,175].

Específicamente, en cáncer de mama se ha demostrado que esta desregulación generalizada del AS tiene implicaciones fundamentales en la biología y la clínica de esta enfermedad. Es más, las firmas de AS pueden actuar como biomarcadores pronósticos del CM [176]. Adicionalmente, estos eventos de AS alterados también pueden estar vinculados al desarrollo, la progresión y la respuesta al tratamiento del CM [177]. En concreto, se ha comprobado que el *splicing* aberrante de varios genes, incluidos *HER2*, *DMTF1*, *BRCA1*, *FGFR1*, *FGFR2*, *KLF6*, *BIRC5* o *TP53*, contribuye a la carcinogénesis mamaria [178].

Además de alteraciones globales en el AS que pueden actuar como biomarcadores pronósticos y predictivos del CM, la presencia de variantes germinales de *splicing* en *cis* en cualquiera de los 8 genes CM principales puede inducir transcritos anómalos implicados en la enfermedad [59,61,72,179–188]. Varias investigaciones revelan que en los genes CM gran parte de las variantes nocivas están afectando al *splicing* del pre-ARNm o, lo que es lo mismo, la desregulación del *splicing* es un mecanismo etiopatogénico frecuente en CMH [189,190]. Por ende, los ensayos funcionales de *splicing* de variantes potencialmente espliceogénicas reportadas en los genes de susceptibilidad a cáncer de mama/ovario investigados en esta tesis (*RAD51C*, *CHEK2*, *BARD1* y *BRIP1*) permitirán la reclasificación de una gran fracción de VUS y, de este modo, se contribuirá a la elucidación del espectro de predisposición genética a esta enfermedad. El objetivo final es incrementar el número de pacientes y familias CMH que se pueden beneficiar de las medidas de prevención correspondientes y de la aplicación de tratamientos personalizados [191].

HIPÓTESIS Y OBJETIVOS

HIPÓTESIS

Recientemente, se ha determinado que variantes patogénicas en al menos 8 genes muestran una asociación significativa con el cáncer de mama: *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM*, *BARD1*, *RAD51C* y *RAD51D*. Asimismo, cerca de un 50% de las variantes reportadas en estos genes de predisposición son de significado clínico incierto, las cuales plantean un serio problema en el diagnóstico, la prevención y la terapia de la enfermedad.

Los genes de predisposición a cáncer de mama acumulan una alta proporción de variantes patogénicas que afectan al *splicing*, el cual constituye, por tanto, un mecanismo prevalente de patogenidad en esta afección. Las variantes espliceogénicas pueden ubicarse, o bien en las secuencias consenso de *splicing*, o bien en los elementos reguladores de este proceso, que son esenciales en el reconocimiento de los exones sujetos a *splicing* alternativo, aquellos que emplean sitios de *splicing* atípicos (p. ej., 5'ss-GC) y los exones de gran/pequeño tamaño. Cualquier tipo de variante en estas regiones puede inducir anomalías de *splicing* y transcritos aberrantes que inactiven el gen CM en cuestión y estén asociados a un incremento del riesgo de esta enfermedad.

Por tanto, los ensayos funcionales de *splicing* mediante minigenes híbridos de variantes de genes CM candidatas de alterar este mecanismo permitirán la identificación de mutaciones espliceogénicas y la interpretación clínica de las mismas, posibilitando la reclasificación de una gran cantidad de VUS. El fin último es mejorar el manejo clínico de las pacientes y aumentar el número de familias que se pueden beneficiar de medidas preventivas y/o terapéuticas personalizadas.

OBJETIVOS

El objetivo principal de esta tesis es el estudio de la desregulación del *splicing* en genes de susceptibilidad a cáncer de mama/ovario y determinar su papel en la enfermedad.

Objetivos específicos:

1. Construcción y validación de minigenes reporteros de *splicing* de los genes de predisposición a cáncer de mama/ovario *CHEK2*, *RAD51C*, *BARD1* y *BRIP1*.
2. Identificación de las variantes de sitios de *splicing* potencialmente espliceogénicas de los genes *CHEK2* y *RAD51C* mediante filtrado bioinformático.
3. Caracterización de variantes con impacto en el *splicing*: ensayos funcionales en minigenes híbridos.
4. Mapeo funcional de elementos reguladores del *splicing* de los exones alternativos 8 y 10 de *CHEK2*, el macroexón 4 de *BARD1* y el exón 1 de *BRIP1* (donador GC). Identificación de mutaciones espliceogénicas en los intervalos críticos mapeados.
5. Interpretación clínica de las variantes testadas siguiendo las normas ACMG/AMP. Evaluación general del papel del *splicing* aberrante como mecanismo etiopatogénico en cáncer de mama hereditario.

RESULTADOS

RAD51C

1. ARTÍCULO 1

Lara Sanoguera-Miralles, Alberto Valenzuela-Palomo, Elena Bueno-Martínez, Patricia Llovet, Beatriz Díez-Gómez, María José Caloca, Pedro Pérez-Segura, Eugenia Fraile-Bethencourt, Marta Colmena, Sara Carvalho, Jamie Allen, Douglas F Easton, Peter Devilee, Maaikje P G Vreeswijk, Miguel de la Hoya and Eladio A Velasco

Comprehensive Functional Characterization and Clinical Interpretation of 20 Splice-Site Variants of the *RAD51C* Gene

Cancers. 2020;12(12):3771.

DOI: <https://doi.org/10.3390/cancers12123771>

Factor de impacto 2020: 6,639

Introducción: El cáncer de mama/ovario hereditario es una enfermedad muy heterogénea con más de 10 genes de predisposición conocidos, entre los cuales se encuentra el gen *RAD51C*. En el marco del proyecto europeo BRIDGES, se secuenció un panel de 34 genes, incluido *RAD51C*, en 60 466 pacientes con cáncer de mama y 53 461 controles. Nuestro objetivo era caracterizar, desde un punto de vista funcional y clínico, las variantes candidatas de alterar el proceso de *splicing* identificadas en este gen.

Material y métodos: Se analizaron bioinformáticamente 40 variantes BRIDGES ubicadas en los límites intrón-exón del gen *RAD51C*, 20 de las cuales fueron seleccionadas para ser ensayadas funcionalmente. Para ello, se diseñó y construyó un minigén reportero de *splicing* que incluía los exones 2 a 8 de *RAD51C* (mgR51C_ex2-8). Este minigén produjo exclusivamente un transcrito *full-length* del tamaño (1062 nt), secuencia y estructura (V1-exones 2 a 8 de *RAD51C*- V2) esperados. Las 20 variantes candidatas se incorporaron en

el minigén WT por mutagénesis dirigida y se estudiaron funcionalmente en células MCF-7.

Resultados: Diecinueve variantes del gen *RAD51C* (95%) alteraron el mecanismo de *splicing* y 18 de ellas produjeron serias anomalías en este proceso, sin dejar trazas o generando cantidades residuales (2-5,4%) del transcrito *full-length*. Por análisis de fragmentos se identificaron 35 transcritos distintos en las variantes testadas: 16 truncantes de proteína, 6 que mantenían el ORF y 13 que se correspondían con isoformas menores no caracterizadas. Según criterios basados en las normas ACMG/AMP, 15 variantes fueron catalogadas como P/LP.

Discusión y conclusiones: La gran cantidad de variantes espliceogénicas detectadas (19/20) refleja una alta precisión del método bioinformático de selección empleado. Estos resultados evidenciaron que el *splicing* aberrante del gen *RAD51C* representa un mecanismo etiopatogénico relevante en la susceptibilidad al cáncer de mama. Por otro lado, los minigenes derivados del vector pSAD han demostrado ser una estrategia robusta y de alta capacidad para la caracterización inicial de las variantes potencialmente espliceogénicas, puesto que replican los resultados de *splicing* previamente detectados en ARN de paciente. Las principales ventajas de este enfoque minigénico son: no se necesitan muestras de individuos portadores de las variantes de interés para llevar a cabo el estudio funcional de *splicing*; vemos únicamente el efecto del alelo mutado sin la interferencia del alelo WT, a diferencia de lo que ocurre al analizar ARN de pacientes, y los ensayos funcionales se pueden realizar en líneas celulares relevantes para la enfermedad. Además, el análisis de fragmentos fluorescentes presenta una alta sensibilidad, distinguiendo hasta 35 transcritos diferentes y permitiendo la resolución de isoformas de ARN que difieren en tan solo 3-nt (∇ (E8p3) y *full-length*). Finalmente, siguiendo las normas ACMG/AMP, 15 variantes BRIDGES del gen de susceptibilidad a cáncer de mama *RAD51C* se clasificaron como P/LP: c.404G>A, c.405-6T>A, c.571+4A>G, c.571+5G>A, c.572-1G>T, c.705G>T, c.706-2A>C, c.706-2A>G, c.837+2T>C, c.905-3C>G, c.905-2A>C, c.905-2_905-1del, c.965+5G>A, c.1026+5_1026+7del y c.1026+5G>T. Estas variantes son clínicamente accionables y, por ende, claves para el correcto manejo clínico de las pacientes portadoras de las mismas.

2. ARTÍCULO 2

Lara Sanoguera-Miralles, Elena Bueno-Martínez, Alberto Valenzuela-Palomo, Ada Esteban-Sánchez, Inés Llinares-Burguet, Pedro Pérez-Segura, Alicia García-Álvarez, Miguel de la Hoya and Eladio A Velasco-Sampedro.

Minigene Splicing Assays Identify 20 Spliceogenic Variants of the Breast/Ovarian Cancer Susceptibility Gene *RAD51C*

Cancers. 2022, 14(12):2960.

DOI: <https://doi.org/10.3390/cancers14122960>

Factor de impacto 2022: 5,2

Introducción: Las variantes de pérdida de función del gen *RAD51C* se asocian a un mayor riesgo de cáncer de mama y ovario. Asimismo, las alteraciones en el *splicing* en este gen son un mecanismo frecuente de inactivación génica, como habíamos demostrado previamente. En este estudio, empleando de nuevo el minigén que abarca los exones 2-8 de *RAD51C* (mgR51C_ex2-8), procedimos a comprobar el impacto sobre el *splicing* de las variantes potencialmente espliceogénicas reportadas en la base de datos ClinVar.

Material y métodos: Se analizaron *in silico*, con la herramienta MaxEntScan (MES), un total de 141 variantes ClinVar ubicadas en los límites intrón/exón de *RAD51C*. De todas ellas, tras un filtrado bioinformático estricto, se seleccionaron 20, las cuales se introdujeron en el minigén WT mediante mutagénesis dirigida. Los minigenes WT y mutantes se transfectaron en distintas líneas celulares para efectuar el estudio funcional de las variantes candidatas.

Resultados: Las 20 variantes seleccionadas perturbaron el mecanismo de *splicing* (100% de sensibilidad de los criterios bioinformáticos de selección) y 18 mostraron anomalías importantes en este proceso, sin dejar ningún rastro o cantidades mínimas (<2,4%) del transcrito *full-length* del minigén (mgFL). Por electroforesis capilar de fragmentos fluorescentes, se identificaron 27 transcritos (incluidos el transcrito mgFL tanto WT como el que incluye la variante c.904G>A); de ellos, 14 truncaban la proteína RAD51C, 3 mantenían el ORF y 8 isoformas menores (1,1-4,7% de la expresión global) no pudieron caracterizarse. Estos datos funcionales de *splicing* generados por los minigenes

mutantes fueron cruciales para la interpretación clínica de las veinte variantes ensayadas basada en las recomendaciones ACMG/AMP, de modo que 16 de ellas fueron clasificadas como LP.

Discusión y conclusiones: Combinando los resultados de los dos estudios presentados de *RAD51C*, se puede concluir que buena parte de las variantes seleccionadas de este gen afecta al *splicing* (39/40), confirmando la precisión del método bioinformático de filtrado. Por otra parte, los ensayos funcionales empleando minigenes híbridos han demostrado ser herramientas valiosas para la caracterización inicial de posibles variantes espliceogénicas, mostrando numerosas ventajas, algunas de ellas indicadas en el resumen del artículo previo. Concretamente, el minigén mgR51C_ex2-8 proporcionó datos muy útiles de los efectos sobre el *splicing* de las 40 variantes de *RAD51C* analizadas, demostrando que cualquier cambio dentro de las secuencias consenso de *splicing* (PPT, 3'ss y 5'ss) es capaz de alterar este proceso clave de la expresión génica. En este estudio ClinVar, de las 20 variantes testadas, 16 se clasificaron como LP: c.146-3C>G, c.404+2T>C, c.405-1G>C, c.571+1del, c.572-3C>G, c.572-1G>C, c.705+1G>A, c.706-1G>T, c.837+1G>T, c.837+4_837+7del, c.838-2A>G, c.904G>A, c.904+1G>T, c.905-3_906del, c.905-2del y c.965+1G>A. En comparación con la clasificación recogida en la base de datos ClinVar, cuatro variantes previamente clasificadas como VUS (c.146-3C>G, c.572-3C>G, c.837+4_837+7del, y c.904G>A) fueron reclasificadas como LP, mientras que 2 mutaciones (c.146-4_146-2del y c.966-1G>C) se degradaron de LP a VUS. La clasificación clínica de las variantes de genes CM, como *RAD51C*, tiene implicaciones muy importantes en el asesoramiento de los pacientes con cáncer de mama hereditario.

CHEK2

1. ARTÍCULO 3

Lara Sanoguera-Miralles, Alberto Valenzuela-Palomo, Elena Bueno-Martínez, Ada Esteban-Sánchez, Víctor Lorca, Inés Llinares-Burguet, Alicia García-Álvarez, Pedro Pérez-Segura, Mar Infante, Douglas F. Easton, Peter Devilee, Maaïke P. G. Vreeswijk, Miguel de la Hoya and Eladio A. Velasco-Sampedro.

Systematic minigene-based splicing analysis and tentative clinical classification of 52 CHEK2 splice-site variants

Clinical Chemistry. 2023

DOI: <https://doi.org/10.1093/clinchem/hvad125>

Factor de impacto 2022: 9,3

Introducción: Las alteraciones en el *splicing* del pre-ARNm son un mecanismo deletéreo frecuente en el cáncer hereditario. Nuestro objetivo en este artículo era analizar funcionalmente, mediante el uso de plásmidos reporteros de *splicing*, las variantes potencialmente espliceogénicas reportadas en el proyecto europeo BRIDGES en el gen de predisposición al cáncer de mama *CHEK2*.

Material y métodos: En el proyecto BRIDGES se identificaron 128 variantes en los sitios de *splicing* (± 10 nt intrónicos y ± 2 nt exónicos) del gen *CHEK2*, las cuales fueron analizadas con la herramienta bioinformática MES. Para el ensayo funcional de las 52 variantes candidatas, se construyeron y se validaron tres minigenes de *CHEK2* (mgChk2_ex1-7, mgChk2_ex6-10 y mgChk2_ex11-15), que abarcan sus 15 exones. Las variantes seleccionadas fueron incorporadas en los minigenes WT mediante mutagénesis dirigida y se ensayaron en varias líneas celulares.

Resultados: Por un lado, se observó que 46 variantes (88,5%) alteraban el *splicing*. Algunas de ellas mostraban perfiles de *splicing* muy complejos, generando hasta 11 transcritos diferentes. Destacablemente, treinta y cuatro variantes produjeron graves anomalías en el proceso de *splicing*, sin dejar ninguna traza o cantidades insignificantes del mgFL. Por otro lado, se identificaron un total de 89 transcritos diferentes, 82 de los cuales pudieron caracterizarse. En concreto, además de los tres transcritos mgFL, cincuenta y nueve isoformas introducían PTC, siete mantenían el ORF, cinco eliminaban el codón de inicio de la traducción, seis afectaban a la región no traducida 5' (5'UTR) y dos incluían variantes *missense*. Estos datos funcionales fueron esenciales en el proceso de interpretación clínica de las variantes de *CHEK2* ensayadas siguiendo los criterios ACMG/AMP, que permitieron la clasificación de 27 variantes como P/LP. Curiosamente, el testado funcional de la variante c.684-2A>G reveló el uso de un sitio aceptor TG atípico (usado por $\approx 0,02\%$ de los aceptores humanos), cuya secuencia consenso está altamente conservada (17/23 posiciones). Seguidamente, se procedió a la búsqueda de regiones funcionales implicadas en el reconocimiento de este sitio no canónico TG, la cual desveló secuencias críticas en el exón 6 para su selección (c.685_698, c.709_722 y c.707_743).

Discusión y conclusiones: Se identificó una alta proporción de variantes *CHEK2* que afectan al *splicing*, lo cual nos lleva a concluir, de nuevo, que el método computacional de selección es muy sensible y que el *splicing* anómalo de genes de susceptibilidad es un mecanismo deletéreo frecuente en cáncer de mama hereditario. Por otro lado, el sistema de minigenes híbridos, cuyas ventajas principales están recogidas en el resumen del primer artículo, proporciona un método sencillo, versátil y adecuado para comprobar el efecto que ejercen sobre el *splicing* las variantes candidatas. Adicionalmente, la técnica de análisis de fragmentos fluorescentes presenta una alta sensibilidad y resolución, separando fragmentos que se diferencian en un solo nucleótido ($\Delta(E7p1)$ y *full-length*). En cuanto a la interpretación clínica de las variantes analizadas en este artículo, la incorporación de los datos de los ensayos funcionales de los minigenes de *CHEK2* a un esquema de clasificación basado en las directrices ACMG/AMP nos permitió catalogar 27 variantes de este gen de predisposición a cáncer de mama como P/LP: c.319+2T>A, c.319+4delA, c.320-3C>G, c.320-2A>G, c.444+1G>T, c.444+1G>A, c.445-2A>G, c.592+2T>G, c.684-2A>G, c.784_792+5del, c.792+1G>A, c.793-2A>G, c.846+1G>T, c.846+5G>A, c.846+6T>C, c.846+4_846+7del, c.847-14_847-2delinsGG, c.908+1G>T, c.908+3A>T, c.1096-6T>G, c.1260-8A>G, c.1375G>A, c.1375+1_1375+2del, c.1376-1G>C, c.1459C>T, c.1461+5G>T y c.1543-9_1546del. Sin embargo, 20 variantes (38%) seguían

siendo de significado clínico incierto debido, en parte, a los complejos patrones de *splicing* detectados en las variantes de *CHEK2* testadas. Este estudio muestra, de nuevo, la alta espliceogenicidad de cualquier cambio en la secuencia consenso de los sitios de *splicing*.

2. ESTUDIO DE REGULACIÓN DE LOS EXONES 8 Y 10 DE *CHEK2*

Lara Sanoguera-Miralles, Inés Llinares-Burguet, Elena Bueno-Martínez, Lobna Ramadane, Cristiana Stuani, Alberto Valenzuela-Palomo, Alicia García-Álvarez, Pedro Pérez-Segura, Emanuele Buratti, Miguel de la Hoya, and Eladio A. Velasco-Sampedro

Comprehensive splicing analysis of the alternatively spliced *CHEK2* exons 8 and 10 reveals three enhancer/silencer-rich regions and 38 spliceogenic variants

The Journal of Pathology (artículo en revisión)

Factor de impacto 2022: 7,3

El *splicing* está controlado por un conjunto amplio de elementos reguladores de este mecanismo (SRE), entre los que se incluyen *enhancers* y silenciadores. Ambos participan en la modulación del *splicing* alternativo, de modo que se espera una mayor concentración de estos SRE en exones sujetos a este fenómeno. Asimismo, las variantes exónicas que afecten a dichos motivos tendrán el potencial de alterar los patrones de *splicing* de un determinado exón. En este apartado, nos centramos en los exones 8 y 10 de *CHEK2*, que están sujetos a eventos prevalentes de *splicing* alternativo fisiológico, generando los transcritos $\Delta(E8)$ (exón 8 *skipping*) y $\Delta(E10)$ (exón 10 *skipping*). Nuestro objetivo era mapear las zonas ricas en SRE e identificar mutaciones espliceogénicas en dichos exones que puedan estar asociadas a un incremento del riesgo de cáncer de mama.

2.1. MAPEO FUNCIONAL DE INTERVALOS RICOS EN ESE

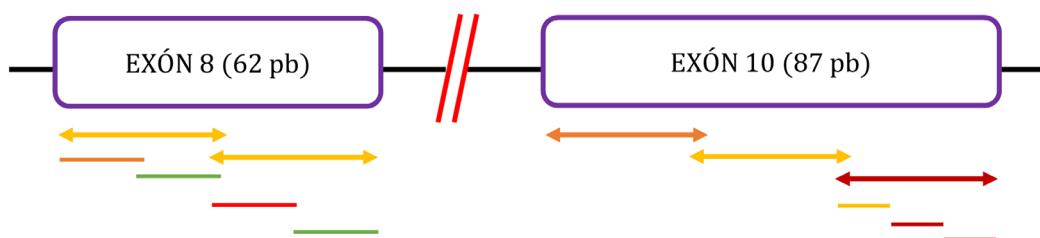
Para llevar a cabo el estudio de regulación de los exones alternativos 8 y 10 de *CHEK2*, se empleó el minigén híbrido que incluía los exones 6 a 10 de este gen (mgChk2_ex6-10), previamente usado para el testado de variantes BRIDGES (**Figuras Suplementarias 1-2**) [192], y se siguió la estrategia de microdeleciones (μ del) solapantes, método validado para la búsqueda de regiones críticas involucradas en el control de los exones diana. Inicialmente, incorporamos por mutagénesis dirigida 5 μ del exónicas de 30 nt en los exones 8 (c.849_878del y c.876_905del) y 10 (c.1011_1040del, c.1037_1066del y c.1063_1092del) de *CHEK2* empleando los primers detallados en la

Tabla Suplementaria 1, las cuales fueron más tarde ensayadas funcionalmente en células MCF-7.

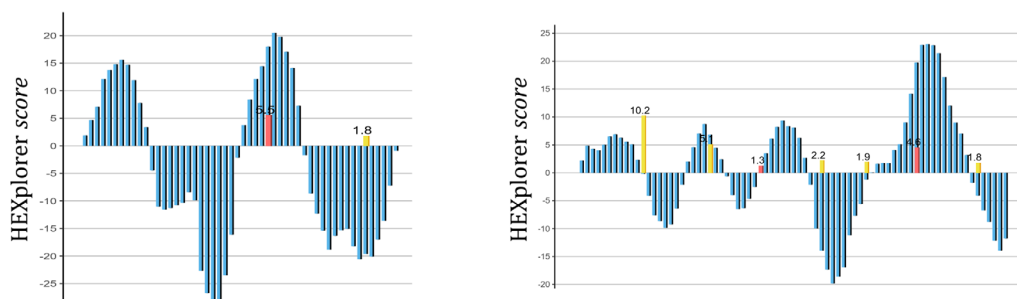
Las dos μ del de 30 nt ubicadas en el exón 8 alteraron débilmente el proceso de *splicing*. Según el perfil HEXplorer (<https://rna.hhu.de/HEXplorer/>), cada segmento de 30 nt contenía una región potenciadora (pico) y otra silenciadora (valle) que se compensarían entre sí, lo que explicaría el impacto débil de dichas μ del en el reconocimiento del exón 8. Para confirmar el perfil HEXplorer, decidimos analizar cuatro μ del internas de 15 nt (c.849_863del, c.863_877del, c.877_891del y c.891_905del). Mientras que dos μ del, c.863_877del y c.891_905del, no afectaron al *splicing* (presuntos intervalos ricos en ESS), las otras dos, c.849_863del y c.877_891del, alteraron significativamente el reconocimiento del exón 8 (regiones ricas en ESE), especialmente la μ del c.877_891del (mgFL=- 13,6%) (**Figura 16; Tabla 2**).

Por otro lado, las μ del c.1011_1040del y c.1063_1092del del exón 10 tuvieron un impacto significativo sobre el proceso de *splicing*, especialmente la c.1063_1092del, la más cercana al 5'ss, que no dejó ningún remanente del transcrito mgFL. Para refinar el mapa SRE del intervalo c.1063_1092, se testaron funcionalmente otras tres μ del internas de 10-11 nt (c.1063_1073del, c.1073_1083del y c.1083_1092del). Una de ellas, c.1073_1083del, tuvo un impacto total sobre el proceso de *splicing* (mgFL=0%), mientras que otra lo alteró fuertemente (c.1083_1092del; mgFL= 38,4%) (**Figura 16; Tabla 2**).

A)



B)



Pie de figura en la página siguiente.

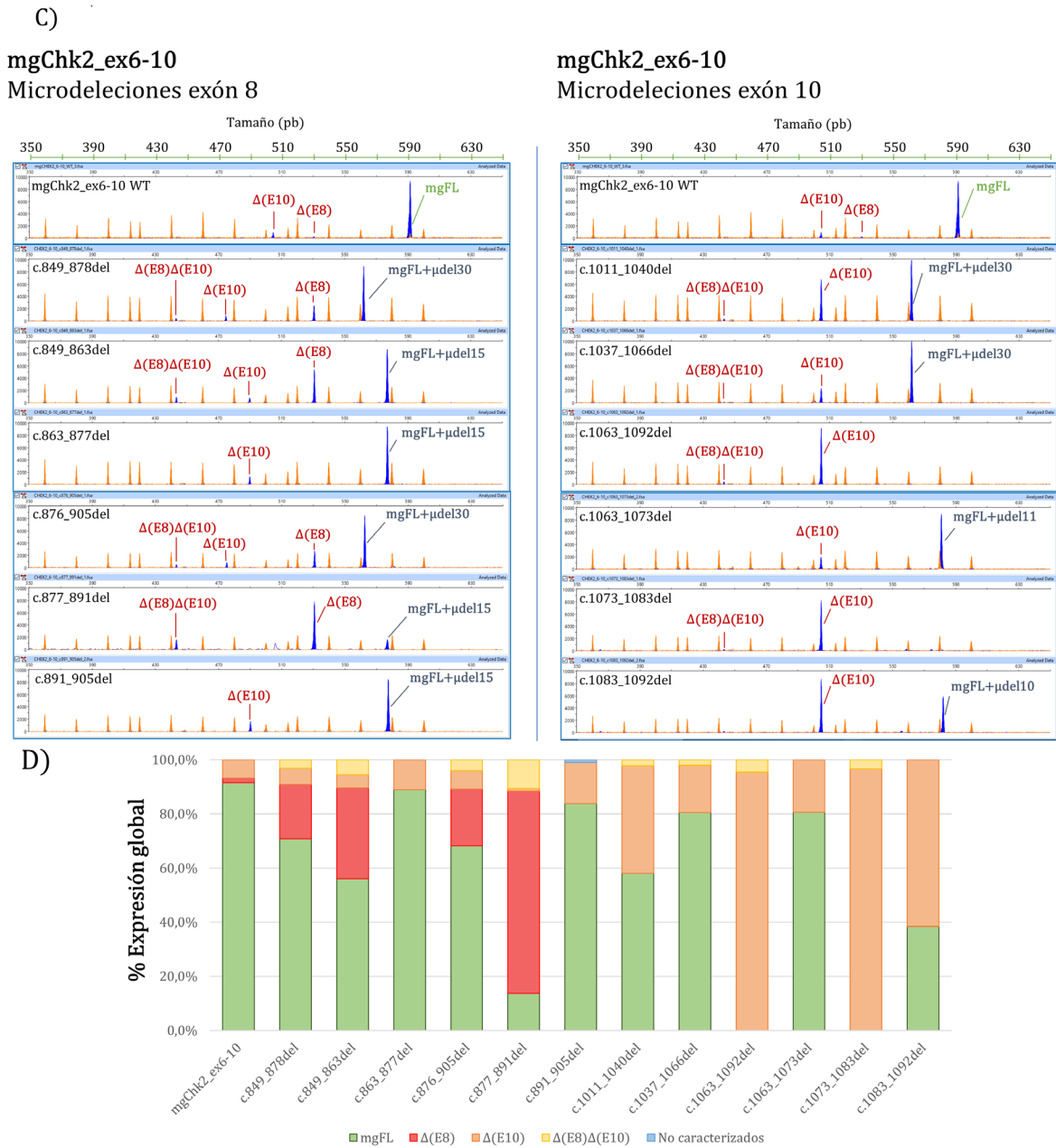


Figura 16. Mapeo de los intervalos ricos en SRE de los exones 8 y 10 de *CHEK2*. A) Mapa de μ del introducidas en el minigén WT mgChk2_ex6-10 (línea roja, impacto fuerte en el *splicing*; línea naranja, impacto débil o moderado; línea verde, impacto débil o nulo). B) Perfil extraído de HEXplorer de los exones 8 y 10 de *CHEK2*, donde los picos indican presuntas regiones ricas en ESE y los valles hipotéticas regiones ricas en ESS. C) Resultados de *splicing* de las doce μ del investigadas. Electroforesis capilar de fragmentos fluorescentes: los transcritos se indican mediante picos azules (productos marcados con FAM) y los picos naranjas representan el estándar de tamaño LIZ600. El eje x indica el tamaño en pb (electroferograma superior) y el eje y representa las unidades relativas de fluorescencia (RFU). D) El gráfico de barras ilustra las proporciones relativas de los diferentes tipos de transcritos generados por las μ del ensayadas.

Tabla 2. Resultados de *splicing* de las μ del incorporadas en los exones 8 y 10 de *CHEK2*.

Variante (HGVS) ¹	Transcritos ²			
	Canónico	PTC	<i>In-frame</i>	No caracterizados
WT (6 muestras)	91,5%± 1,5%	Δ (E8): 1,7%± 0,7%	Δ (E10): 6,8%± 0,9%	
c.849_878del	70,8%± 0,1%	Δ (E8): 20,1%± 0,1% Δ (E8) Δ (E10): 3,1%± 0,0%	Δ (E10): 6,0%± 0,1%	
c.849_863del	56,1%± 0,2%	Δ (E8): 33,5%± 0,0% Δ (E8) Δ (E10): 5,5%± 0,1%	Δ (E10): 4,9%± 0,1%	
c.863_877del	89,0%± 0,9%		Δ (E10): 11,0%± 0,9%	
c.876_905del	68,3%± 0,1%	Δ (E8): 20,9%± 0,1% Δ (E8) Δ (E10): 3,9%± 0,1%	Δ (E10): 6,9%± 0,1%	
c.877_891del	13,6%± 2,3%	Δ (E8): 74,8%± 2,6% Δ (E8) Δ (E10): 10,6%± 0,4%	Δ (E10): 1,0%± 0,1%	
c.891_905del	83,9%± 1,5%		Δ (E10): 15,1%± 1,4%	556-nt: 1,0%± 0,1%
c.1011_1040del	58,1% ± 1,7%	Δ (E8) Δ (E10): 2,2%± 0,2%	Δ (E10): 39,7% ± 1,5%	
c.1037_1066del	80,5%± 0,1%	Δ (E8): 2,0%± 0,1%	Δ (E10): 17,5%± 0,2%	
c.1063_1092del	-	Δ (E8) Δ (E10): 4,5%± 0,2%	Δ (E10): 95,5% ± 0,2%	
c.1063_1073del	80,6% ± 0,1%		Δ (E10): 19,4% ± 0,1%	
c.1073_1083del	-	Δ (E8) Δ (E10): 3,3%± 0,1%	Δ (E10): 96,7% ± 0,1%	
c.1083_1092del	38,4% ± 1,7%		Δ (E10): 61,6% ± 1,7%	

¹Las μ del en las que realizamos estudios sistemáticos de variantes se indican en negrita y la μ del en la que testamos las variantes reportadas en la base de datos ClinVar está subrayada. ²En cada μ del, el transcrito canónico y los otros transcritos que no incluyen la delección del exón 8 (para microdelecciones en este exón) y del exón 10 (para microdelecciones en este exón) incluyen una pérdida de 10, 11, 15 o 30 nt dependiendo de su tamaño.

2.2. ESTUDIO FUNCIONAL DE VARIANTES SRE

A continuación, en los tres intervalos espliceogénicos detectados, testamos funcionalmente 87 variantes exónicas, con el propósito de encontrar mutaciones SRE con impacto en el *splicing* y clínicamente accionables (69 *missense*, 12 sinónimas, 5 *nonsense* y 1 *frameshift*). Nueve de ellas se localizan entre los nucleótidos 1083 y 1092 y proceden de la base de datos ClinVar. Las setenta y ocho variantes restantes se ubican en las regiones altamente espliceogénicas de los exones 8 y 10 (c.877_891 y c.1073_1083), donde se ensayaron todas las variantes posibles para cada posición. Los 87 cambios en el ADN se introdujeron en el minigén mgChk2_ex6-10 WT mediante mutagénesis dirigida, empleando los primers detallados en la **Tabla Suplementaria 2**, y se analizaron en células MCF-7. Los resultados revelaron que 38 variantes (43,7%) alteraban el *splicing* (al menos un 10% de reducción del transcrito mgFL, punto de corte $\leq 81,5\%$), cuatro de las cuales (c.883G>A, c.883G>T, c.884A>T y c.1080G>T) daban lugar a graves anomalías de *splicing*, que no dejaban rastro o en las que solo se detectaban cantidades insignificantes (<5%) del transcrito mgFL (**Figuras 17-18; Tabla 3; Figuras Suplementarias 3-4; Tabla Suplementaria 3**). Además, otras 6 variantes mostraron impactos fuertes (mgFL: 20-50%; c.886G>A, c.886G>T, c.1075G>A, c.1075G>T, c.1076A>T y c.1078G>T), 27 exhibieron impactos moderados (mgFL: 50-80%; c.883G>C, c.884A>C, c.884A>G, c.885A>T, c.886G>C, c.887A>G, c.887A>T, c.890A>T, c.1073A>T, c.1075G>C, c.1076A>C, c.1076A>G, c.1077A>C, c.1077A>T, c.1078G>A, c.1078G>C, c.1079A>G, c.1079A>T, c.1081G>A, c.1081G>C, c.1081G>T, c.1082A>G, c.1082A>T, c.1083C>A, c.1083C>G, c.1083C>T y c.1086T>G) y una indujo un impacto débil (mgFL>80%; c.885A>G). Para comprobar la reproducibilidad de nuestro minigén, se ensayaron cuatro variantes también en células HeLa: c.883G>T, c.884A>G, c.1075G>T y c.1080G>T. Todas estas variantes producían perfiles de *splicing* similares en ambos tipos celulares (**Figura Suplementaria 5**).

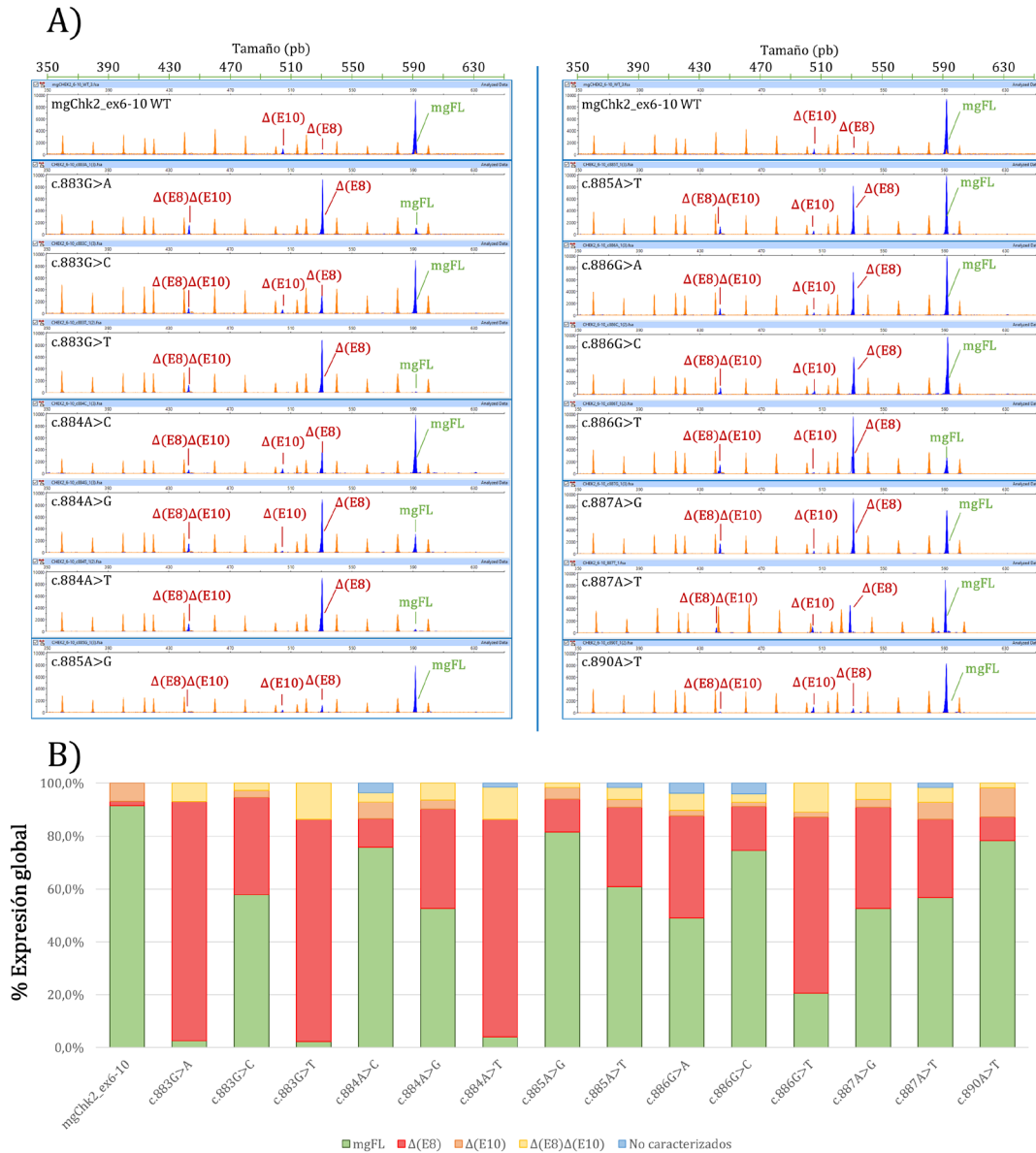


Figura 17. Ensayos de *splicing* de las variantes espliceogénicas del exón 8 de *CHEK2*.
 A) Análisis de fragmentos fluorescentes: los transcritos se indican mediante picos azules (productos marcados con FAM) y los picos naranjas representan el estándar de tamaño LIZ600. El eje x indica el tamaño en pb (electroferograma superior) y el eje y representa las RFU. B) El gráfico de barras ilustra las proporciones relativas de los diferentes tipos de transcritos inducidos por las variantes exónicas espliceogénicas del exón 8 de *CHEK2*.



Figura 18. Ensayos de *splicing* de las variantes espliceogénicas del exón 10 de *CHEK2*. A) Análisis de fragmentos fluorescentes: los transcritos se indican mediante picos azules (productos marcados con FAM) y los picos naranjas representan el estándar de tamaño LIZ600. El eje x indica el tamaño en pb (electroferograma superior) y el eje y representa las RFU. B) El gráfico de barras ilustra las proporciones relativas de los diferentes tipos de transcritos inducidos por las variantes exónicas espliceogénicas del exón 10 de *CHEK2*.

Tabla 3. Resultados de <i>splicing</i> de las variantes exónicas espliceogénicas de los exones 8 y 10 de <i>CHEK2</i> .					
Variante (HGVS) ¹	Proteína (HGVS)	Transcritos ²			
		Canónico	PTC	<i>In-frame</i>	No caracterizados
mgChk2_ex11-15 WT (6 ensayos)		91,5%± 1,5%	Δ(E8): 1,7%± 0,7%	Δ(E10): 6,8%± 0,9%	
<u>Ex8</u>					
c.883G>A	p.(Glu295Lys)	2,6%± 0,2%	Δ(E8): 90,5%± 0,1% Δ(E8)Δ(E10):6,9%± 0,1%		
c.883G>C	p.(Glu295Gln)	57,8%± 0,7%	Δ(E8): 36,8%± 0,8% Δ(E8)Δ(E10):2,7%± 0,0%	Δ(E10): 2,7%± 0,0%	
c.883G>T	p.(Glu295*)	2,2%± 0,1%	Δ(E8): 84,1%± 0,4% Δ(E8)Δ(E10):13,7%± 0,3%		
c.884A>C	p.(Glu295Ala)	75,7%± 0,3%	Δ(E8): 10,9%± 0,3% Δ(E8)Δ(E10):3,5%± 0,1%	Δ(E10): 6,3%± 0,4%	360-nt: 1,9%± 0,3% 321-nt: 1,7%± 0,0%
c.884A>G	p.(Glu295Gly)	52,7%± 1,1%	Δ(E8): 37,5%± 1,0% Δ(E8)Δ(E10):6,4%± 0,2%	Δ(E10): 3,4%± 0,1%	
c.884A>T	p.(Glu295Val)	4,0%± 0,6%	Δ(E8): 82,3%± 0,3% Δ(E8)Δ(E10):12,2%± 0,8%		596-nt: 1,5%± 0,4%
c.885A>G	p.(Glu295=)	81,5%± 0,3%	Δ(E8): 12,5%± 0,1% Δ(E8)Δ(E10):1,6%± 0,1%	Δ(E10): 4,4%± 0,2%	
c.885A>T	p.(Glu295Asp)	60,9%± 0,2%	Δ(E8): 30,0%± 0,3% Δ(E8)Δ(E10):4,6%± 0,1%	Δ(E10): 2,9%± 0,1%	321-nt: 1,6%± 0,1%
c.886G>A	p.(Asp296Asn)	49,0%± 0,2%	Δ(E8): 38,6%± 0,4% Δ(E8)Δ(E10):6,4%± 0,1%	Δ(E10): 2,2%± 0,0%	306-nt: 2,4%± 0,2% 321-nt: 1,4%± 0,1%
c.886G>C	p.(Asp296His)	74,6%± 3,2%	Δ(E8): 16,6%± 2,2% Δ(E8)Δ(E10):3,2%± 1,3%	Δ(E10): 1,6%± 1,6%	321-nt: 4,0%± 0,3%
c.886G>T	p.(Asp296Tyr)	20,6%± 0,5%	Δ(E8): 66,5%± 0,7% Δ(E8)Δ(E10):10,9%± 0,4%	Δ(E10): 2,0%± 0,6%	
c.887A>G	p.(Asp296Gly)	52,7%± 0,3%	Δ(E8): 38,2%± 0,2% Δ(E8)Δ(E10):6,1%± 0,2%	Δ(E10): 3,0%± 0,0%	

Variante (HGVS) ¹	Proteína (HGVS)	Transcritos ²			
		Canónico	PTC	<i>In-frame</i>	No caracterizados
c.887A>T	p.(Asp296Val)	56,7%± 2,4%	Δ(E8): 29,7%± 0,6% Δ(E8)Δ(E10):5,7%± 0,8%	Δ(E10): 6,3%± 0,5%	580-nt: 1,6%± 0,8%
c.890A>T	p.(Tyr297Phe)	78,3%± 1,5%	Δ(E8): 8,9%± 0,8% Δ(E8)Δ(E10):1,8%± 0,1%	Δ(E10): 11,0%± 0,7%	
<u>Ex10</u>					
c.1073A>T	p.(Gln358Leu)	76,6%± 1,1%	Δ(E8): 2,8%± 0,2% Δ(E8)Δ(E10): 1,0%± 0,1%	Δ(E10): 19,6%± 0,8%	
c.1075G>A	p.(Glu359Lys)	38,4%± 0,3%	Δ(E8)Δ(E10): 2,1%± 0,1%	Δ(E10): 59,5%± 0,3%	
c.1075G>C	p.(Glu359Gln)	75,7%± 0,2%	Δ(E8): 2,4%± 0,1% Δ(E8)Δ(E10): 1,0%± 0,0%	Δ(E10): 20,9%± 0,2%	
c.1075G>T	p.(Glu359*)	27,2%± 1,0%	Δ(E8)Δ(E10): 3,2%± 0,1%	Δ(E10): 69,6%± 1,2%	
c.1076A>C	p.(Glu359Ala)	73,6%± 1,0%	Δ(E8): 1,5%± 0,1%	Δ(E10): 24,9%± 1,0%	
c.1076A>G	p.(Glu359Gly)	71,1%± 3,0%	Δ(E8): 2,2%± 0,3%	Δ(E10): 26,7%± 1,0%	
c.1076A>T	p.(Glu359Val)	38,7%± 4,0%	Δ(E8)Δ(E10): 2,2%± 0,2%	Δ(E10): 59,1%± 3,8%	
c.1077A>C	p.(Glu359Asp)	73,2%± 0,3%	Δ(E8): 2,1%± 0,0% Δ(E8)Δ(E10): 1,0%± 0,0%	Δ(E10): 23,7%± 0,3%	
c.1077A>T	p.(Glu359Asp)	74,6%± 0,6%	Δ(E8): 2,1%± 0,0%	Δ(E10): 23,3%± 0,5%	
c.1078G>A	p.(Glu360Lys)	64,6%± 0,5%	Δ(E8): 1,1%± 0,0% Δ(E8)Δ(E10):c1,0%± 0,1%	Δ(E10): 33,3%± 0,4%	
c.1078G>C	p.(Glu360Gln)	58,0%± 0,6%	Δ(E8): 1,2%± 0,1% Δ(E8)Δ(E10):1,2%± 0,0%	Δ(E10): 39,6%± 0,5%	
c.1078G>T	p.(Glu360*)	26,1%± 1,9%	Δ(E8)Δ(E10): 2,1%± 0,0%	Δ(E10): 71,8%± 1,8%	
c.1079A>G	p.(Glu360Gly)	65,1%± 1,0%	Δ(E8): 1,8%± 0,1% Δ(E8)Δ(E10): 1,2%± 0,1%	Δ(E10): 29,6%± 0,7%	596-nt: 2,3%± 0,7%
c.1079A>T	p.(Glu360Val)	77,4%± 0,4%	Δ(E8): 1,5%± 0,1%	Δ(E10): 19,8%± 0,3%	572-nt: 1,3%± 0,2%
c.1080G>T	p.(Glu360Asp)	4,8%± 1,6%	Δ(E8)Δ(E10): 1,6%± 0,4%	Δ(E10): 93,6%± 1,9%	
c.1081G>A	p.(Asp361Asn)	75,2%± 4,7%	Δ(E8): 1,7%± 0,3%	Δ(E10): 23,1%± 4,4%	
c.1081G>C	p.(Asp361His)	76,0%± 0,4%	Δ(E8): 1,3%± 0,3%	Δ(E10): 21,0%± 0,3%	572-nt: 1,7%± 0,4%

Variante (HGVS) ¹	Proteína (HGVS)	Transcritos ²			
		Canónico	PTC	<i>In-frame</i>	No caracterizados
c.1081G>T	p.(Asp361Tyr)	50,5%± 3,5%	$\Delta(E8): 1,1\% \pm 0,3\%$ $\Delta(E8)\Delta(E10): 1,8\% \pm 0,1\%$	$\Delta(E10): 46,6\% \pm 3,7\%$	
c.1082A>G	p.(Asp361Gly)	68,2%± 4,8%	$\Delta(E8): 1,6\% \pm 0,0\%$ $\Delta(E8)\Delta(E10): 1,3\% \pm 0,3\%$	$\Delta(E10): 28,9\% \pm 4,5\%$	
c.1082A>T	p.(Asp361Val)	76,7%± 1,4%	$\Delta(E8): 1,7\% \pm 0,1\%$	$\Delta(E10): 21,6\% \pm 1,3\%$	
c.1083C>A	p.(Asp361Glu)	72,6%± 2,3%	$\Delta(E8): 1,6\% \pm 0,6\%$	$\Delta(E10): 25,8\% \pm 1,8\%$	
c.1083C>G	p.(Asp361Glu)	76,3%± 0,2%	$\Delta(E8): 1,3\% \pm 0,0\%$	$\Delta(E10): 20,8\% \pm 0,2\%$	572-nt: 1,6%± 0,1%
c.1083C>T	p.(Asp361=)	77,0%± 0,7%	$\Delta(E8): 3,1\% \pm 0,1\%$ $\Delta(E8)\Delta(E10): 1,2\% \pm 0,1\%$	$\Delta(E10): 18,7\% \pm 0,5\%$	
c.1086T>G	p.(Cys362Trp)	79,7%± 0,3%	$\Delta(E8): 3,4\% \pm 0,1\%$ $\Delta(E8)\Delta(E10): 1,2\% \pm 0,0\%$	$\Delta(E10): 15,7\% \pm 0,4\%$	

¹Se muestran en negrita las 10 variantes con impactos severos (<5% del transcrito mgFL) y fuertes (mgFL: 20-50%). ²Lógicamente, en cada variante, los transcritos en los que no se suprime el exón mutado incorporarán la mutación puntual investigada.

El análisis de fragmentos fluorescentes detectó hasta 16 transcritos diferentes (incluido el transcrito mgFL), 12 de los cuales eran isoformas menores no caracterizadas ($\leq 4\%$ de la expresión global). Los transcritos $\Delta(E8)$ (r.847_908del; p.Pro283Aspfs*8) y $\Delta(E8)\Delta(E10)$ (r.[847_908del; 1009_1095del]; p.Pro283Aspfs*8), introducen PTC, por lo que probablemente estén asociados a un mayor riesgo de cáncer de mama. En cambio, la isoforma $\Delta(E10)$ (*skipping* del exón 10), mantiene el ORF (r.1009_1095del; p.Tyr337_Lys365del). Este transcrito *in-frame* introduce en la proteína una delección de 29 aminoácidos, 21 de los cuales están altamente conservados en vertebrados (**Figura Suplementaria 6**), lo que sugiere un papel relevante de esta región en la función de la proteína. Es más, esta isoforma elimina parte del dominio quinasa de la proteína CHK2 [113,146,193]. Por tanto, la pérdida del exón 10 podría inactivar CHK2.

Además, los resultados de estas variantes se utilizaron para evaluar la precisión de tres herramientas de predicción ESE/ESS (HEXplorer, HOT-SKIP [<https://hot-skip.img.cas.cz/>], HExoSplice [<http://bioinfo.univ-rouen.fr/HExoSplice/inputs.php>]) y SpliceAI (**Tabla 4; Figura 19; Tabla Suplementaria 4-5**) [194]. Sorprendentemente, las variantes espliceogénicas c.1080G>T, que altera drásticamente el reconocimiento del exón 10 ($\Delta(E10)$: 93,6%), y c.1081G>A ($\Delta(E10)$: 23,6%) no habrían sido seleccionadas por ninguno de los programas bioinformáticos testados. De los cuatro algoritmos analizados, HEXplorer presentó la sensibilidad más alta (92,1%), mientras que HOT-SKIP mostró la más baja (23,7%), con 29 falsos negativos. Por otro lado, SpliceAI mostró la mayor especificidad (85,7%), seguido de HOT-SKIP (70,8%), HExoSplice (60,4%) y HEXplorer (44,9%). Por último, HExoSplice ofreció la precisión más elevada (72,1%) definida como la capacidad de detección de variantes verdaderamente positivas y verdaderamente negativas. Curiosamente, el análisis empleando la herramienta SpliceAI sugirió un mejor rendimiento para las variantes en el exón 8 (80% de precisión) que para las variantes ubicadas en el exón 10 (45,2% de precisión) (**Tabla Suplementaria 6**).

Tabla 4. Sensibilidad, especificidad y precisión de las herramientas de predicción de ESE/ESS.

Herramientas bioinformáticas	Espliceogénicas (Sensibilidad) ¹		No espliceogénicas (Especificidad) ²		Precisión ³
	Verdadero positivo	Falso negativo	Verdadero negativo	Falso positivo	
HEXplorer	35/38 (92,1%)	3	22/49 (44,9%)	27	57/87 (65,5%)
HOT-SKIP	9/38 (23,7%)	29	34/48 (70,8%)	14	43/86 (50%)
HExoSplice	33/38 (86,8%)	5	29/48 (60,4%)	19	62/86 (72,1%)
SpliceAI	13/38 (34,2%)	25	42/49 (85,7%)	7	55/87 (63,2%)

¹Las variantes con efectos muy débiles (<10% de reducción del transcrito canónico) no se consideraron espliceogénicas. Sensibilidad (entre paréntesis) = Verdadero positivo/Número total de variantes espliceogénicas. ²Especificidad (entre paréntesis) = Verdadero negativo/Número total de variantes no espliceogénicas. ³La precisión (entre paréntesis) se calculó como [Verdadero positivo + Verdadero negativo]/Variantes totales.

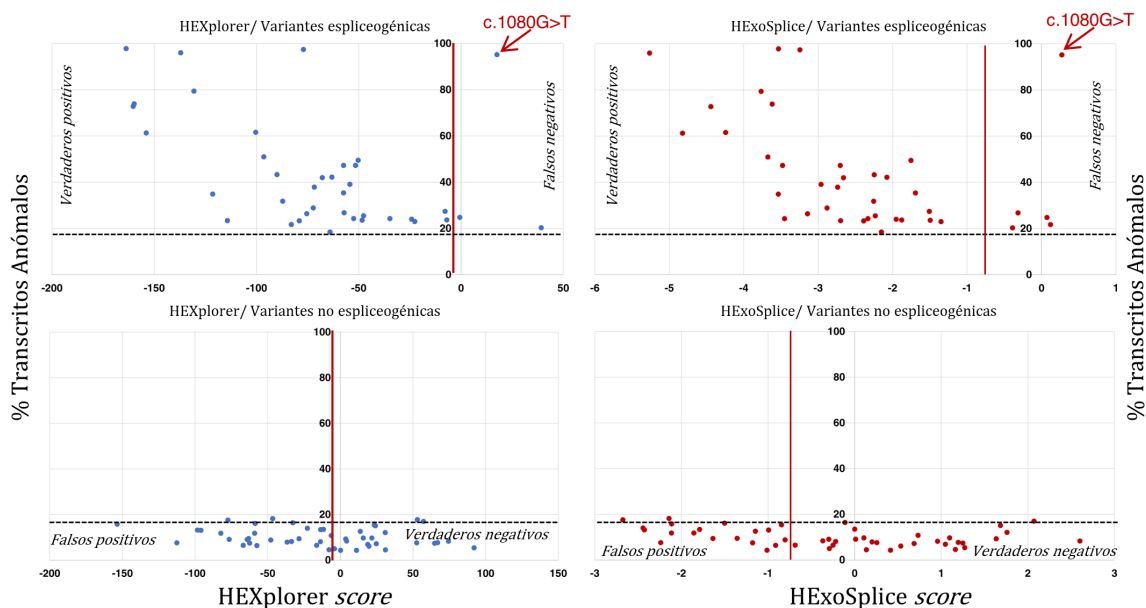


Figura 19. Correlación entre las predicciones SRE de HEXplorer y HExoSplice y los resultados funcionales. Los valores de HEXplorer y HExoSplice se han representado gráficamente frente a la proporción de transcritos anómalos inducidos por cada variante. Los puntos de corte (líneas rojas verticales) para la selección de variantes fueron ≤ -5 (HEXplorer) y $\leq -0,75$ (HExoSplice). Según esto, las variantes situadas a la izquierda de las líneas rojas serían seleccionables para ser ensayadas funcionalmente. El punto de corte para los transcritos anómalos fue el 18% de la expresión global (líneas discontinuas), que representa una reducción de aproximadamente el 10% del mgFL del minigén WT (91,5% de la expresión global). Por lo tanto, las variantes (puntos) por encima de la línea discontinua son espliceogénicas. La variante c.1080G>T (>95% de transcritos aberrantes), falso negativo, se indica con una flecha roja.

2.3. INTERPRETACIÓN CLÍNICA DE VARIANTES SIGUIENDO LAS NORMAS ACMG/AMP

Tras el estudio de *splicing* empleando el minigén mgChk2_ex6-10, clasificamos las 87 variantes exónicas del gen *CHEK2* testadas siguiendo un sistema de puntuación basado en las normas ACMG/AMP, que incluye las especificaciones del gen *CHEK2* desarrolladas internamente [192,195]. El sistema de clasificación incorpora las lecturas de los minigenes como códigos PVS1(ARN)/BP7(ARN) de fortaleza variable dependiendo del resultado experimental real, tal y como recomendó recientemente el Subgrupo de *Splicing* ClinGen SVI (*Clinical Genome Resource Sequence Variant Interpretation Working Group*) [196]. Para las variantes que producen lecturas complejas (dos o más transcritos), primero asignamos códigos PVS1(ARN)/BP7(ARN) a cada transcrito individual y, posteriormente, estos se combinan para asignar un código PVS1(ARN) (o BP7(ARN)) global, basado en las fortalezas de los códigos individuales de cada transcrito y su contribución relativa a la expresión global (véase la **Figura Suplementaria 7** para más detalles).

En concreto, de las 38 variantes espliceogénicas identificadas, una variante fue clasificada como P, dos como LP, y 35 de ellas como VUS (**Tabla 5; Tabla Suplementaria Excel 1**). La elevada tasa de VUS refleja el hecho de que la mayoría de estas variantes espliceogénicas muestran efectos parciales, expresando cantidades variables de transcritos aberrantes y transcrito *full-length* (estos últimos portadores de la correspondiente variante investigada). Notablemente, nuestro enfoque acaba clasificando dos variantes *missense* del exón 8 (c.883G>A p.(Glu295Lys) y c.884A>T p.(Glu295Val)) como LP, ya que ambas inducen fundamentalmente transcritos truncantes (mayoritariamente *skipping* del exón 8). Es más, nuestro estudio muestra que los datos funcionales recientes de *CHEK2* basados en ADNc que apoyan la benignidad para algunas variantes *missense* (en particular, para c.883G>C p.(Glu295Gln), c.887A>G p.(Asp296Gly), y c.1076A>G p.(Glu359Gly)) deben interpretarse con cautela, ya que estas variantes tienen un impacto significativo en el *splicing* [197]. Por último, nuestro sistema de clasificación cataloga dos variantes *nonsense* localizadas en el exón 10 (c.1075G>T p.(Glu359*) y c.1078G>T p.(Glu360*)) como variantes de significado clínico incierto, puesto que en ambas el *skipping* del exón 10, que introduce una delección *in-frame*, es el transcrito mayoritario (69,6% y 71,8%, respectivamente).

Tabla 5. Clasificación ACMG/AMP de las 38 variantes espliceogénicas testadas en los exones 8 y 10 de <i>CHEK2</i> .							
c.HGVS¹	p.HGVS¹	ACMG/AMP		PVS1(ARN)/BP7(ARN)²	PM2³	PM5⁴	ClinVar⁵
c.883G>A	p.(Glu295Lys)	LP	+9	+8	+1	0	VUS (2)
c.883G>C	p.(Glu295Gln)	VUS	+1	0	+1	0	VUS (1)
c.883G>T	p.(Glu295*)	P	+10	+8	+1	+1	P (1)
c.884A>C	p.(Glu295Ala)	VUS	+1	0	+1	0	(-)
c.884A>G	p.(Glu295Gly)	VUS	+1	0	+1	0	VUS (2)
c.884A>T	p.(Glu295Val)	LP	+9	+8	+1	0	(-)
c.885A>G	p.(Glu295=)	VUS	+1	0	+1	0	VUS (1); LB (2)
c.885A>T	p.(Glu295Asp)	VUS	+1	0	+1	0	VUS (2)
c.886G>A	p.(Asp296Asn)	VUS	+1	0	+1	0	VUS (1)
c.886G>C	p.(Asp296His)	VUS	+1	0	+1	0	VUS (2)
c.886G>T	p.(Asp296Tyr)	VUS	+1	0	+1	0	VUS (5)
c.887A>G	p.(Asp296Gly)	VUS	+1	0	+1	0	VUS (3)
c.887A>T	p.(Asp296Val)	VUS	+1	0	+1	0	VUS (3)
c.890A>T	p.(Tyr297Phe)	VUS	+1	0	+1	0	(-)
c.1073A>T	p.(Gln358Leu)	VUS	+1	0	+1	0	(-)
c.1075G>A	p.(Glu359Lys)	VUS	+1	0	+1	0	LB (1)
c.1075G>C	p.(Glu359Gln)	VUS	+1	0	+1	0	(-)
c.1075G>T	p.(Glu359*)	VUS	+5	+4	+1	0	(-)
c.1076A>C	p.(Glu359Ala)	VUS	+1	0	+1	0	(-)
c.1076A>G	p.(Glu359Gly)	VUS	0	0	0	0	VUS (6)
c.1076A>T	p.(Glu359Val)	VUS	+1	0	+1	0	(-)
c.1077A>C	p.(Glu359Asp)	VUS	+1	0	+1	0	(-)
c.1077A>T	p.(Glu359Asp)	VUS	+1	0	+1	0	(-)
c.1078G>A	p.(Glu360Lys)	VUS	+1	0	+1	0	VUS (8)
c.1078G>C	p.(Glu360Gln)	VUS	+1	0	+1	0	(-)
c.1078G>T	p.(Glu360*)	VUS	+5	+4	+1	0	(-)
c.1079A>G	p.(Glu360Gly)	VUS	+1	0	+1	0	VUS (1)
c.1079A>T	p.(Glu360Val)	VUS	+1	0	+1	0	VUS (1)

c.HGVS ¹	p.HGVS ¹	ACMG/AMP		PVS1(ARN)/BP7(ARN) ²	PM2 ³	PM5 ⁴	ClinVar ⁵
c.1080G>T	p.(Glu360Asp)	VUS	+5	+4	+1	0	VUS (1)
c.1081G>A	p.(Asp361Asn)	VUS	+1	0	+1	0	(-)
c.1081G>C	p.(Asp361His)	VUS	+1	0	+1	0	VUS (5)
c.1081G>T	p.(Asp361Tyr)	VUS	+1	0	+1	0	VUS (1)
c.1082A>G	p.(Asp361Gly)	VUS	+1	0	+1	0	(-)
c.1082A>T	p.(Asp361Val)	VUS	+1	0	+1	0	(-)
c.1083C>A	p.(Asp361Glu)	VUS	+1	0	+1	0	(-)
c.1083C>G	p.(Asp361Glu)	VUS	+1	0	+1	0	VUS (1)
c.1083C>T	p.(Asp361=)	VUS	+1	0	+1	0	LB (4)
c.1086T>G	p.(Cys362Trp)	VUS	+1	0	+1	0	VUS (1)

La tabla muestra solo las evidencias ACMG/AMP que han contribuido a la clasificación final. Un análisis más detallado se muestra en la **Tabla Suplementaria Excel 1**.

¹Nomenclatura HGVS usando como referencia la secuencia NM_007194. Las variantes clasificadas como P/LP se muestran en negrita. ²Fortaleza del código PVS1(ARN)/BP7(ARN) derivada de las lecturas de los minigenes mgChk2_ex6-10 mutantes. Ver **Tabla Suplementaria Excel 1** para más detalles. ³Hemos aplicado la evidencia patológica de rareza como una fuerza de apoyo, tal y como recomendó recientemente el ClinGen SVI. ⁴Hemos aplicado el PM5 como una fuerza de apoyo a los PTC. No se ha aplicado a las variantes c.1075G>T p.(Glu359*) y c.1078G>T p.(Glu360*), ya que el análisis empleando el minigén mgChk2_ex6-10 ha revelado que estas dos mutaciones no son PTC *bonafide*. ⁵A efectos comparativos, resumimos el estado de ClinVar de las 38 variantes investigadas (último acceso 14/08/2023). La información detallada de ClinVar se proporciona en la **Tabla Suplementaria Excel 1**.

La clasificación de las 49 variantes no espliceogénicas se muestra en la **Tabla Suplementaria Excel 2**. La mayoría de las variantes (N=31) terminaron catalogándose como variantes de significado clínico incierto. Sin embargo, pudimos clasificar todas las variantes sinónimas (N=10) como B/LB y todas las variantes que introducen un PTC (N=3) como P, destacando la relevancia de un resultado de *splicing* negativo a la hora de clasificar variantes exónicas de cualquier tipo.

En conclusión, el mapeo funcional mediante microdeleciones exónicas es una estrategia valiosa para la localización de regiones ricas en SRE, el cual ha permitido la detección de 3 regiones de este tipo en los exones 8 y 10 de *CHEK2*. El estudio sistemático de 87 variantes ubicadas en dichos intervalos identificó 38 variantes exónicas espliceogénicas, de las cuales 3 fueron clasificadas como P/LP.

BARD1

1. CONSTRUCCIÓN Y VALIDACIÓN DEL MINIGÉN *WILD TYPE*

Para el estudio de los elementos funcionales implicados en la regulación del exón largo de *BARD1* (exón 4; 950 pb), se construyó por clonaje tradicional un vector reportero de *splicing* que incluía los exones 3 a 5 de este gen, basándonos en la secuencia de *BARD1* NM_000465.4. Así pues, el exón 4 de interés fue clonado en el vector de *splicing* pSAD (Patente P201231427, Consejo Superior de Investigaciones Científicas) junto con sus exones vecinos naturales 5' y 3', para que las reacciones de *splicing* tengan lugar en su contexto genómico natural.

Inicialmente, se incorporó en el vector de *splicing* el exón 4. En este sentido, el inserto de interés, que incluía el exón 4 y 253 pb del final del intrón 3 y 300 pb del principio del intrón 4, se amplificó a partir de ADN genómico con la polimerasa *Phusion* de alta fidelidad (*Thermo Fisher Scientific*, Waltham, MA, Estados Unidos) y los primers indicados en la **Tabla Suplementaria 7**. A continuación, tanto el vector como el inserto se cortaron empleando las enzimas de restricción *Bam*HI y *Kpn*I y, tras la ligación, obtuvimos el minigén mgBARD1_ex4 (**Figura Suplementaria 8A**). La validación funcional de este minigén dio como resultado un transcrito *full-length* del tamaño (1131 nt), secuencia y estructura (V1- *BARD1* exón 4-V2) esperados (**Figura Suplementaria 9A**).

Seguidamente, se agregó el exón 5 a este constructo. Para ello, el inserto con el exón 5, 147 pb del intrón 4 y 200 pb del intrón 5 se amplificó con la polimerasa *Phusion* y los primers detallados en la **Tabla Suplementaria 7**. Posteriormente, este inserto fue incorporado en mgBARD1_ex4 por *overlap extension PCR* (**Figura Suplementaria 8A**). El minigén mgBARD1_ex4-5 produjo un transcrito *full-length* del tamaño (1212 nt), secuencia y estructura (V1- exones 4 a 5 de *BARD1*- V2) esperados, junto con varios eventos alternativos que implican el uso de donadores crípticos localizados 137, 261 y

408 nt *upstream* del sitio 5'ss canónico del exón 4, así como el *skipping* de este macroexón (**Figura Suplementaria 9B**).

Finalmente, se añadió el exón 3 al minigén mgBARD1_ex4-5. A tal efecto, el inserto que abarca 218 pb del intrón 2, el exón 3 y 270 pb del intrón 3 se amplificó igual que los anteriores y se incluyó en el minigén mgBARD1_ex4-5 mediante corte con las enzimas de restricción *SacII* y *BamHI* (**Figura Suplementaria 7; Figura Suplementaria 8A**). En este caso, la validación del minigén mgBARD1_ex3-5 reveló que una pequeña proporción de los transcritos producidos se correspondían con el mgFL, sin embargo, la mayoría derivaban del uso de un aceptor alternativo por parte del exón 4 ubicado 151 nt *downstream* del exón 3 (**Figura Suplementaria 9C**). Con el fin de estabilizar el minigén, se introdujo una delección de 250 pb que afectaba a ambas partes del intrón 3, dejando 122 nt de la región 5' de este intrón y 157 nt de su región 3' (**Figura Suplementaria 8B**). De esta forma, logramos obtener un minigén estable, el cual produjo un transcrito *full-length* del tamaño (1361 nt), secuencia y estructura (V1- exones 3 a 5 de *BARD1*- V2) previstos, acompañado por las mismas isoformas alternativas descritas en la validación del mgBARD1_ex4-5 (**Figura 20**).

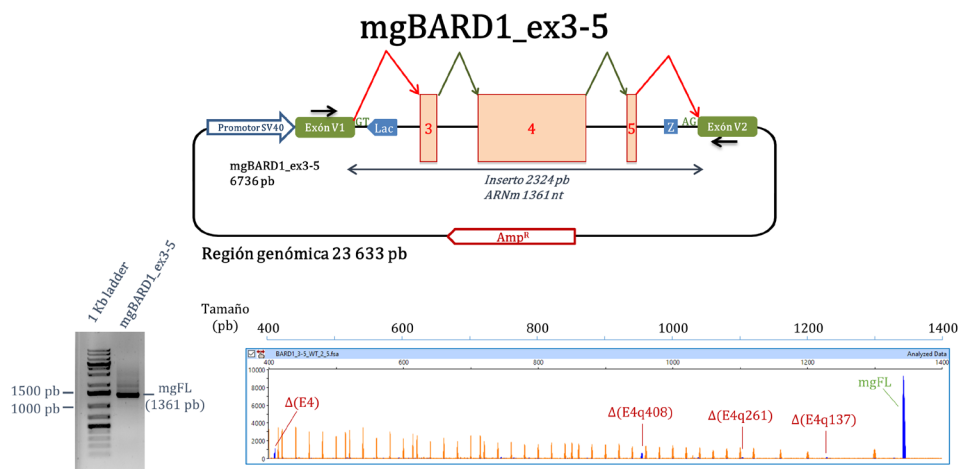
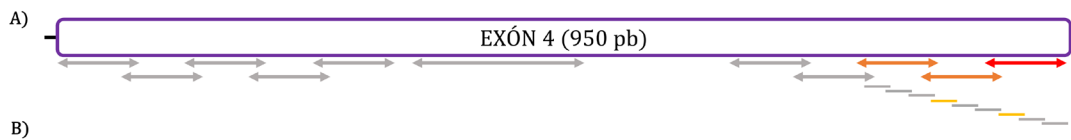


Figura 20. Estructura y validación funcional del minigén final de *BARD1*. En la parte superior de la figura aparece la representación gráfica del minigén mgBARD1_ex3-5 final. Los exones están dentro de recuadros de color rojizo; las flechas negras sobre los exones V1 y V2 indican los primers de RT-PCR específicos del vector. En la parte inferior de la figura se muestra la validación funcional del minigén: los productos RT-PCR se analizaron mediante electroforesis de agarosa (izquierda) y electroforesis capilar de fragmentos fluorescentes (derecha). En el electroferograma del análisis de fragmentos fluorescentes, los picos azules son los transcritos producidos por el minigén, mientras que los naranjas representan el patrón de tamaño LIZ1200. Por otro lado, el eje x refleja el tamaño en pb y el eje y representa las RFU.

2. MAPEO FUNCIONAL DE LOS MOTIVOS REGULADORES

El mapeo de los motivos reguladores de *BARD1* implicados en el reconocimiento y la inclusión de su exón largo se llevó a cabo mediante microdeleciones exónicas solapantes, inicialmente, de 30 nt. En concreto, se incorporaron tanto en la región 5' como en la región 3' del exón 4 cinco μ del de 30 nt con 5 nt solapantes, dejando intactos los dos primeros y los tres últimos nucleótidos del exón (primers en **Tabla Suplementaria 8**). Además, se generó una microdelección más grande de 104 nt (c.581_684del), diseñada para eliminar una región rica en *enhancers* Tra2 β , de acuerdo con las predicciones de la herramienta *Human Splicing Finder* (HSF, <http://www.umd.be/HSF3/>, herramienta actualmente no disponible; **Tabla Suplementaria 9**) [198]. De las 11 microdeleciones ensayadas, tres (c.1232_1261del, c.1257_1286del y c.1282_1311del) alteraban considerablemente el proceso de *splicing* basándonos en los resultados de electroforesis en gel de agarosa (**Figura Suplementaria 10A**). En estas 3 áreas que tuvieron un mayor impacto sobre el proceso de *splicing*, se introdujeron tres μ del de 12 nt con 2 nt solapantes para afinar aún más la región de interés. Sin embargo, estas microdeleciones más pequeñas afectaban débilmente al reconocimiento del exón 4 (**Figura Suplementaria 10B**).

Los transcritos generados por las μ del espliceogénicas se analizaron también por electroforesis capilar de fragmentos fluorescentes, cuyos resultados se muestran en la **Figura 21** y en la **Tabla 6**. Así pues, por análisis de fragmentos identificamos 16 isoformas de *splicing*, de las cuales una engloba los distintos transcritos *full-length* (WT y con las diferentes μ del), tres mantienen el ORF, otras tres incorporan PTC y nueve se corresponden con transcritos menores no caracterizados. Concretamente, las 6 isoformas caracterizadas, aparte de los distintos transcritos *full-length*, son las siguientes: Δ (E4q137) (p.(Thr394Argfs*6)), Δ (E4q261) (p.(Val352_Lys438del)), Δ (E4q408) (p.(Val303_Lys438del)), Δ (E4q843) (p.(Val158_Lys438del)), Δ (E4q863) (p.(Ser151Argfs*7)) y Δ (E4) (p.(Asp122Glyfs*7)).



B)

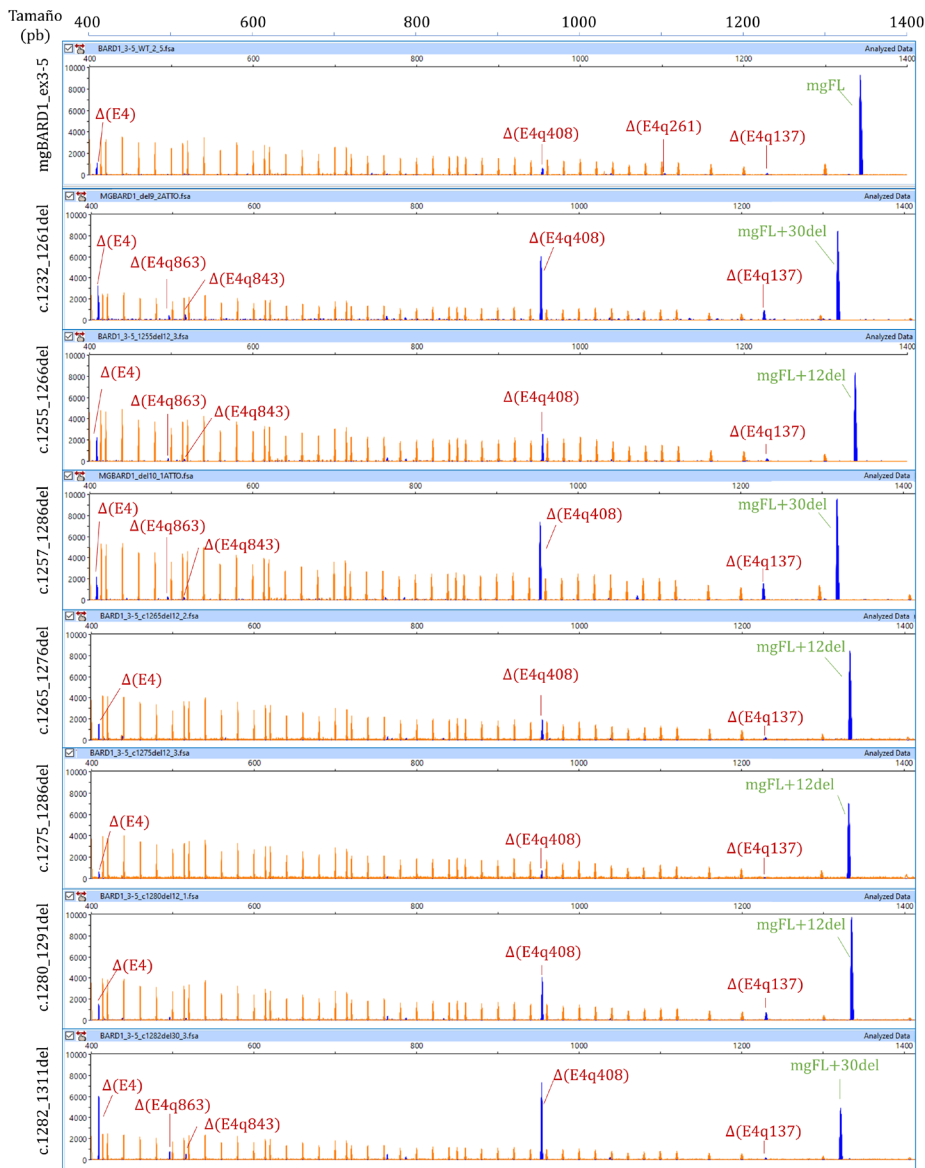


Figura 21. Ensayos funcionales de *splicing* de las μ del de *BARD1* espliceogénicas. A) Mapa de las μ del ensayadas. En gris aparecen las que no alteran el proceso de *splicing*, de acuerdo con los resultados de electroforesis en gel de agarosa, y en color amarillo, naranja y rojo las que tienen un impacto débil, moderado y alto, respectivamente. B) Análisis de fragmentos fluorescentes de los transcritos generados por los minigenes WT y con las distintas μ del. En los electroferogramas de análisis de fragmentos fluorescentes, los picos azules se corresponden con los transcritos producidos por el minigén, mientras que los naranjas representan el patrón de tamaño LIZ1200. Por otro lado, el eje x refleja el tamaño en pb y el eje y representa las RFU.

Tabla 6. Resultados de *splicing* de las microdeleciones espliceogénicas ensayadas en mgBARD1_ex3-5.

Microdelección ¹	Transcritos ²			
	<i>Full-length</i>	PTC	<i>In-frame</i>	No caracterizados
mgBARD1_ex3-5 (6 muestras)	81,4%±1,5%	Δ(E4): 7,2%±0,9% Δ(E4q137): 1,1%±0,2%	Δ(E4q408): 5,0%±0,9% Δ(E4q261): 1,0%±0,2%	1574-nt: 4,3%±1,0%
c.1232_1261del	42,6%±1,2%	Δ(E4): 14,5%±1,0% Δ(E4q137): 4,5%±0,2% Δ(E4q863): 2,1%±0,1%	Δ(E4q408): 31,0%±0,2% Δ(E4q843): 2,3%±0,1%	764-nt: 1,7%±0,1% 787-nt: 1,3%±0,3%
c.1255_1266del	60,1%±3,1%	Δ(E4): 11,4%±0,4% Δ(E4q137): 1,9%±0,1% Δ(E4q863): 1,2%±0,1%	Δ(E4q408): 16,9%±2,1% Δ(E4q843): 1,1%±0,1%	1584-nt: 3,6%±0,4% 787-nt: 1,4%±0,2% 764-nt: 2,4%±0,2%
c.1257_1286del	48,4%±4,0%	Δ(E4): 8,9%±0,7% Δ(E4q137): 6,3%±0,8% Δ(E4q863): 1,3%±0,1%	Δ(E4q408): 31,4%±3,1% Δ(E4q843): 1,1%±0,2%	1073-nt: 1,5%±0,2% 764-nt: 1,1%±0,1%
c.1265_1276del	72,4%±0,8%	Δ(E4): 7,9%±0,6% Δ(E4q137): 1,5%±1,3%	Δ(E4q408): 14,1%±0,2%	764-nt: 2,9%±0,9% 438-nt: 1,2%±1,1%
c.1275_1286del	79,8%±0,3%	Δ(E4): 4,2%±0,1% Δ(E4q137): 1,2%±0,1%	Δ(E4q408): 7,3%±0,2%	1548-nt: 4,6%±0,2% 1525-nt: 1,5%±0,2% 764-nt: 1,4%±0,1%
c.1280_1291del	66,5%±2,5%	Δ(E4): 5,5%±0,7% Δ(E4q137): 4,1%±0,5%	Δ(E4q408): 20,9%±1,4%	764-nt: 1,9%±0,1% 438-nt: 1,1%±1,2%
c.1282_1311del	25,6%±3,1%	Δ(E4): 23,4%±3,9% Δ(E4q863): 2,8%±0,4% Δ(E4q137): 1,1%±0,2%	Δ(E4q408): 38,6%±2,3% Δ(E4q843): 2,1%±0,4%	764-nt: 2,3%±0,1% 787-nt: 1,4%±0,0% 1561-nt: 1,4%±0,2% 1038-nt: 1,3%±0,2%

¹En negrita aparecen las μ del que exhiben impactos fuertes en el *splicing*. ²Los transcritos están anotados de la siguiente manera: Δ (*skipping* de secuencias exónicas), E (exón) y, cuando es necesario, q (desplazamiento del donador) + nt insertados o eliminados.

3. MOTIVOS SILENCIADORES PUTATIVOS PRÓXIMOS A LOS DONADORES CRÍPTICOS

El exón 4 de *BARD1* tiene varios donadores crípticos que, generalmente, no son usados por la maquinaria de *splicing*. Asimismo, HSF detectó una alta concentración de silenciadores hnRNPA1 putativos alrededor de los sitios crípticos que podrían ser responsables de su represión, por lo que decidimos investigar esta hipótesis. Para ello, destruimos, mediante la incorporación de una serie de mutaciones puntuales, dichos motivos silenciadores próximos a los sitios 5'ss crípticos localizados 137 (c.1175_1183), 408 (c.904_912) y 843 (c.469_477) nt *upstream* del canónico, los cuales se han denominado sitio críptico 6, 4 y 2, respectivamente (**Figura Suplementaria 11**). Seleccionamos estos donadores crípticos en base tanto a su fortaleza (puntuación MES) como a su frecuencia de uso en nuestro minigén y en condiciones fisiológicas. Luego, incorporamos por mutagénesis dirigida las variantes puntuales necesarias para lograr nuestro propósito, empleando los primers detallados en la **Tabla Suplementaria 10** [199]. En un principio, solo la interrupción de los silenciadores próximos al sitio críptico 6 producía una alteración en el proceso de *splicing* (**Figura 22**). En este caso, por secuenciación Sanger, comprobamos que lo que ocurría era que se estaba empleando como donador el sitio críptico 4, ubicado 408 nt *upstream* del canónico, y se estaba reconociendo otro exón dentro del propio exón 4 de 137 pb, el cual usa un aceptor adyacente al sitio críptico 6 (c.1158_1180) y el donador clásico del exón 4 de *BARD1* (**Figura Suplementaria 11**). La puntuación MES revela que una de las mutaciones puntuales que incluimos para destruir uno de los silenciadores próximos al sitio críptico 6 incrementa la fortaleza de este aceptor críptico (secuencia WT: ATTCATTAGTCTTTCACCAG|GTA→MES=3,70; secuencia mutada: ATTCATTGGTCTTTCACCAG|GTA→MES=7,73).

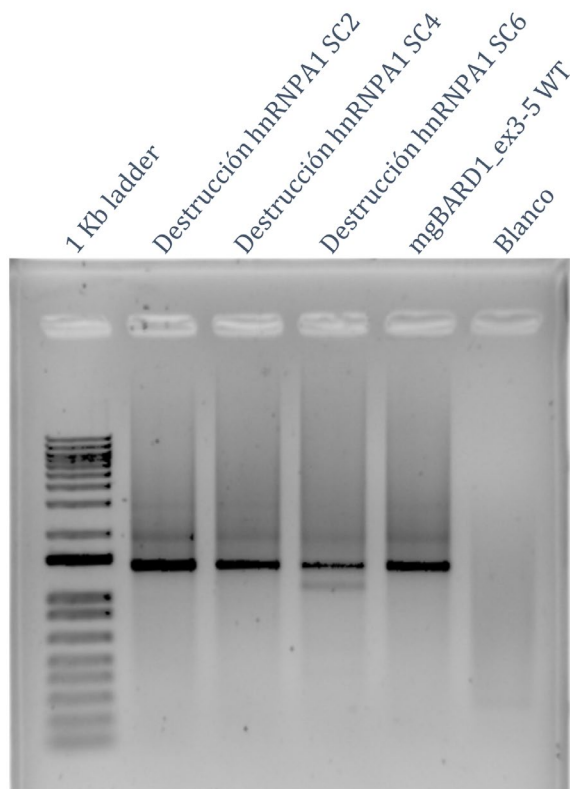


Figura 22. Electroforesis en gel de agarosa de los transcritos generados tras la destrucción de los motivos hnRNPA1 predichos próximos a los donadores críticos del exón 4 de *BARD1*. En esta figura se muestra el gel de agarosa (1%) de los productos de RT-PCR generados tanto por el minigén mgBARD1_ex3-5 WT como por los distintos minigenes mutantes en los que se han interrumpido los ESS putativos para el factor de *splicing* hnRNPA1 próximos a los sitios críticos 2 (SC2), 4 (SC4) y 6 (SC6).

En conclusión, en el exón 4 de *BARD1* se identificaron tres regiones ricas en SRE (c.1232_1261, c.1257_1286 y c.1282_1311) donde podrían ubicarse mutaciones potencialmente espliceogénicas. Por otra parte, el fortalecimiento de sitios críticos o la generación de sitios *de novo* por ciertas variantes podría provocar su uso y el consiguiente acortamiento de los macroexones, de modo que su estudio funcional podría ser una estrategia óptima para la identificación de mutaciones potencialmente deletéreas del gen *BARD1*. Por último, se dispone de un constructo valioso, validado funcionalmente y apto para el estudio de cualquier variante reportada de los exones 3 a 5 de *BARD1*.

BRIP1

1. CONSTRUCCIÓN Y VALIDACIÓN DEL MINIGÉN *WILD TYPE*

El exón 1 del gen *BRIP1* posee un donador atípico GC, cuya regulación procedimos a investigar. A tal efecto, diseñamos un minigén reportero de *splicing* que incluía el primer y el segundo exón de dicho gen, basándonos en la secuencia de *BRIP1* NM_032043.3.

Empezamos incorporando el exón 2 en el vector de *splicing* pSAD. En este sentido, amplificamos el inserto con el exón 2 y 156 y 252 nt *upstream* y *downstream* de este exón a partir de ADN genómico, empleando la polimerasa *Phusion* y los primers listados en la **Tabla Suplementaria 11**. A continuación, el inserto fue clonado en el vector mediante corte con enzimas de restricción (*XbaI* y *BamHI*) y ligación, obteniéndose el minigén mgBRIP1_ex2 (**Figura Suplementaria 12A**). La validación funcional de este minigén dio como resultado un transcrito *full-length* que mostraba un tamaño (304 nt), secuencia y estructura (V1- exón 2 de *BRIP1*-V2) esperados, el cual aparecía acompañado por un transcrito que incluía el *skipping* de este exón y una inserción intrónica no caracterizada (**Figura Suplementaria 13A**).

A este vector que portaba el exón 2 de *BRIP1*, se le incorporó el exón 1, generando un exón quimérico V1- ex1 *BRIP1*. El inserto de 490 nt, el cual engloba 240 pb del exón 1 y 250 pb del intrón 1, se amplificó de la misma forma que el anterior, pero, en este caso, se clonó en el minigén por *overlap extension PCR*, sustituyendo 62 pb del exón V1 y 363 nt del intrón del vector (**Figura Suplementaria 12B**). Así, se obtuvo el minigén mgBRIP1_ex1-2, un minigén cuyo transcrito principal empleaba inesperadamente un donador críptico (MES score = 10,19) en el exón 1, localizado 190 nt *upstream* del canónico (**Figura Suplementaria 13B**). Este se estabilizó introduciendo una delección de 52 nt en la región 5' del exón 1 de *BRIP1*, que eliminaba dicho donador críptico. El minigén resultante generó un transcrito *full-length* del tamaño (408 pb), secuencia y estructura

(V1- exón 1 de *BRIP1*- exón 2 de *BRIP1*-V2) esperados, acompañado de la isoforma fisiológica $\Delta(E2p4)$ (**Figura 23; Figura Suplementaria 12C**).

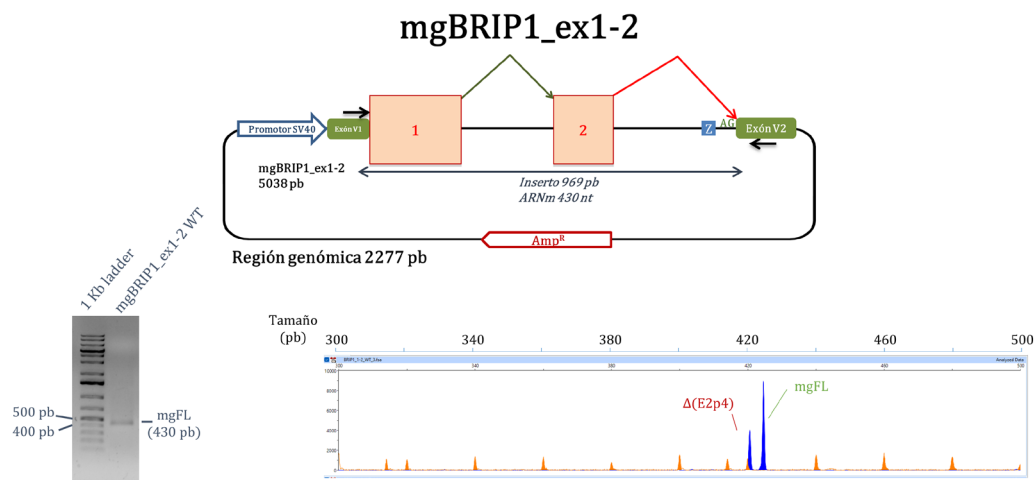


Figura 23. Estructura y validación funcional del minigén final de *BRIP1*. En la parte superior de la figura aparece la representación gráfica del minigén mgBRIP1_ex1-2 final. Los exones están dentro de recuadros de color rojizo; las flechas negras sobre los exones V1 y V2 indican los primers de RT-PCR específicos del vector. En la parte inferior de la figura se muestra la validación funcional del minigén: los productos RT-PCR se analizaron mediante electroforesis en gel de agarosa (izquierda) y electroforesis capilar de fragmentos fluorescentes (derecha). En el electroferograma del análisis de fragmentos fluorescentes, los picos azules se corresponden con los transcritos producidos por el minigén, mientras que los naranjas representan el patrón de tamaño LIZ600. El eje x refleja el tamaño en pb y el eje y las RFU.

2. MAPEO FUNCIONAL DE LOS MOTIVOS REGULADORES

La búsqueda de los motivos reguladores de *BRIP1* involucrados en el reconocimiento del donador atípico GC de su exón 1 se realizó mediante μ del exónicas solapantes. Todas las μ del fueron incorporadas en mgBRIP1_ex1-2 mediante mutagénesis dirigida, utilizando los primers detallados en la **Tabla Suplementaria 12**. Específicamente, se introdujeron al final del exón 1 y al principio del 2 dos μ del de 30 nt con 5 nt solapantes, sin deleccionar los dos primeros y los tres últimos nucleótidos de dichos exones. Ninguna de estas microdelecciones tuvo un impacto relevante sobre el *splicing* del gen, conforme a las imágenes de electroforesis en gel de agarosa (**Figura 24A**). A pesar de ello, decidimos agregar 2 μ del adicionales de 16 nt en la región del exón 1 más cercana al donador GC, las cuales tampoco alteraron el *splicing* de *BRIP1* (**Figura 24B**). Por tanto, podemos concluir que el reconocimiento del donador GC del exón 1 de *BRIP1* no está regulado por secuencias exónicas de la región 3' de este exón y

de la región 5' del exón 2, lo cual puede ser debido a la alta conservación de la secuencia consenso del sitio 5' de *splicing* (8/9 posiciones; CAG|GCAAGT vs. CAG|GTAAGT).

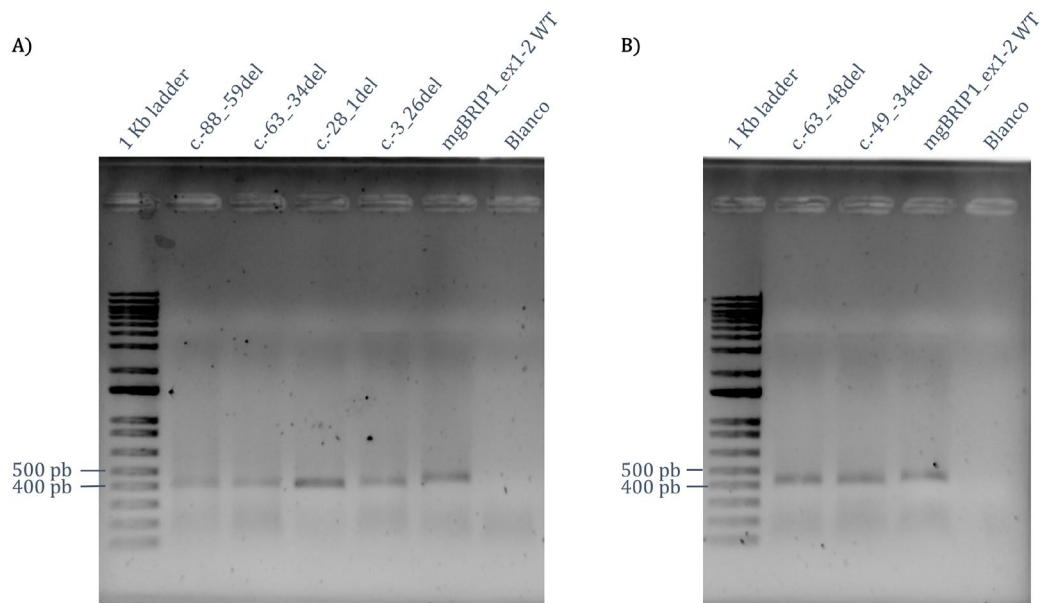


Figura 24. Electroforesis en gel de agarosa de los resultados del ensayo funcional de las μ del incorporadas en el minigén mgBRIP1_ex1-2. A) Electroforesis en gel de agarosa (1%) de las cuatro μ del de 30 nt incorporadas en la región 3' del exón 1 (c.-88_-59del y c.-63_-34del) y en la región 5' del exón 2 (c.-28_1del y c.-3_26del). B) Electroforesis en gel de agarosa (1%) de las dos μ del de 16 nt próximas al donador GC (c.-63_-48del y c.-49_-34del).

DISCUSIÓN

DISCUSIÓN

El cáncer de mama es la neoplasia maligna más común en todo el mundo. La aparición de esta patología se asocia con muchos factores de riesgo, incluida la predisposición genética y hereditaria. Específicamente, las formas hereditarias explican entre el 5% y el 10% de todos los CM. *BRCA1* y *BRCA2* fueron los primeros genes de predisposición descubiertos y tradicionalmente se han considerado como los principales en esta enfermedad. Sin embargo, los avances en el campo de la secuenciación masiva, unidos a una disminución progresiva de su coste, han permitido la identificación de nuevos genes y de nuevas mutaciones involucradas en el desarrollo de tumores mamarios malignos. Estas técnicas son la base de grandes estudios caso-control, como BRIDGES, en el que se analizaron 34 genes en más de 113 000 mujeres de 25 países distintos (60 466 mujeres con CM y 53 461 controles), y CARRIERS, que investigó 28 genes en unas 64 000 mujeres estadounidenses (32 247 pacientes CM y 32 544 controles). En ambos estudios, se ha demostrado que variantes truncantes en otros 6 genes están significativamente asociadas con un incremento del riesgo de CM: *PALB2*, *ATM*, *CHEK2*, *BARD1*, *RAD51C* y *RAD51D* [108,109]. En cambio, los genes *PTEN*, *TP53*, *CDH1*, *STK11* y *NF1*, cada uno de los cuales se relaciona con un síndrome clínico distinto, todos ellos vinculados al cáncer de mama, a pesar de haber sido, por regla general, considerados genes de alta penetrancia, no mostraron una asociación significativa en estos estudios, presumiblemente debido a la baja prevalencia de mutaciones [200].

Los paneles multigénicos, además de ser herramientas muy útiles a nivel experimental, están siendo usados de forma rutinaria en todo el mundo con fines diagnósticos y su demanda se ha incrementado considerablemente en los últimos años con la promesa de una atención personalizada. Es verdad que estos exámenes ofrecen muchas ventajas, destacando la reducción de los costes de las pruebas genéticas y la posibilidad de analizar simultáneamente varios genes de interés con una gran profundidad de cobertura, lo cual facilita notablemente el estudio de condiciones hereditarias heterogéneas, como el cáncer, y permite la búsqueda de marcadores teragnósticos. No obstante, la otra cara de la moneda es que no se detectan variantes estructurales, existe una falta de estandarización de estas pruebas y la detección de variantes está superando la capacidad de interpretación de las mismas, lo que irremediablemente conduce a una acumulación de variantes de significado clínico incierto [122,123]. De hecho, en los paneles utilizados para pacientes de alto riesgo de

CM, la probabilidad de encontrar una VUS es significativamente mayor que la probabilidad de detectar una variante P/LP [201]. Este fenómeno fue corroborado por el proyecto CARRIERS, en el cual se observó que la prevalencia de variantes patogénicas en 12 genes de predisposición al cáncer de mama (*ATM*, *BARD1*, *BRCA1*, *BRCA2*, *CDH1*, *CHEK2*, *NF1*, *PALB2*, *PTEN*, *RAD51C*, *RAD51D* y *TP53*) era del 5,01% en las pacientes y del 1,63% en controles, mientras que más del 18% de los individuos en ambos grupos presentaban una VUS [109]. Más aún, según la base de datos ClinVar ([https://www.ncbi.nlm.nih.gov/clinvar?term=%22clinvar_all%22\[Filter\]](https://www.ncbi.nlm.nih.gov/clinvar?term=%22clinvar_all%22[Filter]); último acceso 18 de septiembre del 2023), solo el 11,5% de las variantes reportadas son P/LP, el 38,4% están catalogadas como B/LB y las más abundantes serían las VUS, que suponen el 47,5% de todas las variantes registradas en esta base de datos. En particular, en los 8 genes CM principales, el porcentaje de VUS en ClinVar es del 32,5% (**Tabla 7**). Estas variantes de significado clínico desconocido limitan la naturaleza informativa de las pruebas genéticas y, por ende, suponen un gran desafío para el proceso de asesoramiento genético y la gestión clínica de los pacientes. De ahí que la reclasificación de estas VUS como patogénicas o benignas sea esencial para guiar las decisiones y recomendaciones médicas e impacte sustancialmente en las opciones con las que cuentan los pacientes en cuanto a prevención del cáncer, detección, terapia dirigida contra esta enfermedad y decisiones quirúrgicas y reproductivas [202].

Tabla 7. Variantes P/LP y VUS registradas en la base de datos ClinVar en los genes que conforman el núcleo central de la susceptibilidad a CM. Último acceso 18 de septiembre de 2023.

Gen	Variantes ClinVar	VUS	% VUS	P/LP	% P/LP
<i>BRCA1</i>	13 896	1727	12,43%	4212	30,31%
<i>BRCA2</i>	17 422	3256	18,69%	5125	29,42%
<i>CHEK2</i>	3567	1795	50,32%	869	24,36%
<i>ATM</i>	14 575	7073	48,53%	3082	21,15%
<i>PALB2</i>	5195	2568	49,43%	1237	23,81%
<i>RAD51C</i>	1811	945	52,18%	341	18,83%
<i>RAD51D</i>	1575	767	48,70%	240	15,24%
<i>BARD1</i>	3512	1888	53,76%	497	14,15%
TOTAL	61 553	20 019	32,52%	15 603	25,35%

Durante muchos años, el filtrado de variantes teragnósticas candidatas se ha centrado en las consecuencias sobre la codificación proteica, pasando por alto muchas variantes genéticas implicadas en otros mecanismos clave de la expresión génica, como la transcripción o el *splicing*. Es más, las variantes que generan PTC (*nonsense* o *frameshift*) son mutaciones consideradas prácticamente de forma automática como P/LP. Sorprendentemente, dentro de este conjunto de variantes truncantes de proteína, algunos autores incluyen también a los cambios intrónicos que afectan a las posiciones altamente conservadas de los sitios de *splicing* ($\pm 1,2$). Nuestra experiencia en el ámbito de la desregulación del *splicing* evidencia que esta asunción es un error, puesto que algunas de estas variantes introducen deleciones o inserciones *in-frame* en el ARN maduro y otras generan sitios atípicos que pueden ser reconocidos por el espliceosoma, como ciertas mutaciones +2T>C (GT→donador atípico GC), pudiéndose mantener en ambos casos la funcionalidad del transcrito principal [124,125]. Aparte de las mutaciones $\pm 1,2$, existen otras muchas posibles variantes de *splicing*, generalmente clasificadas como VUS a la espera de un estudio funcional que determine su efecto exacto y proporcione evidencias relevantes para su clasificación clínica [203]. Asimismo, se ha demostrado que la disrupción del *splicing* es un mecanismo etiopatogénico muy frecuente, sobre todo, en cáncer, constituyendo un sello distintivo de esta patología [190,204]. Por lo tanto, los ensayos de *splicing* pueden aportar una información muy importante para la interpretación clínica de VUS y la reclasificación de variantes previamente catalogadas como benignas o patogénicas. Conocer el impacto exacto de una variante sobre el *splicing* puede ser también muy valioso para el diseño de estrategias terapéuticas alternativas, como, por ejemplo, la modulación del *splicing* anómalo mediante oligonucleótidos antisentido [57].

Tradicionalmente, los ensayos funcionales de *splicing* se han realizado a partir de ARN del paciente. Sin embargo, cuando la muestra de ARN del paciente es inaccesible, los vectores reporteros de *splicing* o minigenes híbridos representan una herramienta genética idónea para la caracterización preliminar de posibles variantes espliceogénicas. De hecho, están recogidos como una de las opciones para realizar los estudios funcionales de *splicing* en las guías de clasificación ACMG/AMP [124]. Este sistema de minigenes ofrece las siguientes ventajas: i) análisis aislado del efecto del alelo mutante sin la interferencia del alelo WT, a diferencia de lo que ocurre al emplear muestras de pacientes portadoras; ii) identificación y cuantificación precisa de todos los transcritos producidos por la variante testada gracias al uso de inhibidores del NMD (cicloheximida), que

impiden la degradación de los transcritos con PTC; iii) estudio de múltiples variantes empleando un único constructo; iv) posibilidad de realizar el ensayo funcional en diferentes tipos celulares relevantes para la enfermedad en cuestión, y v) alta reproducibilidad de los patrones de *splicing* fisiológicos y patológicos. Además, la electroforesis capilar fluorescente de los transcritos generados ofrece una alta resolución y sensibilidad, siendo capaz de separar fragmentos que se diferencian en un solo nt [184,205]. En conclusión, todos estos datos sugieren que los minigenes proporcionan una información muy útil para la interpretación clínica inicial de variantes candidatas.

Cabe destacar que el grupo de Splicing y Susceptibilidad Genética a Cáncer del IBGM, en el cual se ha realizado esta tesis doctoral, tiene una vasta experiencia en el diseño y la construcción de minigenes híbridos y el testado funcional de variantes. Particularmente, ha llevado a cabo más de 300 ensayos funcionales de variantes del gen *BRCA2* empleando este sistema [61,72,181–185]. Más recientemente, debido a su participación en el proyecto europeo BRIDGES como parte del *Work Package 4 (Functional analysis of gene variants)*, este grupo ha efectuado estudios sistemáticos de *splicing* mediante minigenes de los principales genes CM, en concreto, *PALB2*, *ATM*, *CHEK2*, *RAD51C* y *RAD51D*, algunos de los cuales son la base de esta tesis doctoral [59,180,186,192,206]. Adicionalmente, a través de su servicio de minigenes (<http://www.ibgm.med.uva.es/servicios/servicio-de-splicing-minigenes/>), ha investigado la disrupción del proceso de *splicing* en genes implicados en otras enfermedades genéticas: *MLH1* (síndrome de Linch), *COL1A1* (osteogénesis imperfecta), *SERPINA1* (deficiencia de alfa-1 antitripsina), *CHD7* (síndrome CHARGE), *GRN* (demencia frontotemporal), *UGT1A1* (síndrome de Crigler-Najjar), *TP53* (síndrome de Li-Fraumeni), *TRPM4* (cáncer colorrectal) y *SGMS2* (síndrome de lesiones en anillo del cuero cabelludo-fragilidad ósea) [207–210].

Concretamente, en esta tesis nos hemos centrado en el estudio de los genes *CHEK2*, *RAD51C*, *BARD1* y *BRIP1*, en los que más de la mitad de las variantes reportadas en la base de datos ClinVar son VUS (**Tabla 7**). A excepción de *BRIP1*, en el que solo está confirmada su vinculación con el cáncer de ovario, el resto de los genes estudiados forman parte del núcleo central de la susceptibilidad hereditaria a cáncer de mama [108,109,211]. Todos ellos están incluidos en las guías de manejo de la NCCN, recomendándose, salvo en *BRIP1*, una vigilancia intensificada sin cirugía profiláctica mediante mamografías y otras técnicas de diagnóstico por imagen. Por otro lado, en *RAD51C* y *BRIP1*, debido al riesgo

significativo de cáncer de ovario, se aconseja la salpingo-ooforectomía bilateral preventiva entre los 45 y 50 años, y en *CHEK2* se recomienda también una vigilancia incrementada a través de colonoscopias para la detección precoz del cáncer colorrectal y hacer un cribado de cáncer de próstata a partir de los 40 años (NCCN *Guideline: Genetic/Familial High-Risk Assessment: Breast; Ovarian, and Pancreatic Version 1.2024*). En consecuencia, el estudio de la desregulación del *splicing* en estos genes va a contribuir a la dilucidación del espectro de predisposición genética a cáncer de mama/ovario y va a ser clave en el manejo clínico óptimo de las pacientes.

1. VARIANTES DE SITIOS DE *SPLICING*

Uno de los objetivos de esta tesis es el estudio de las variantes de sitios de *splicing* potencialmente espliceogénicas de los genes *RAD51C* y *CHEK2*. A tal efecto, en estos genes se recopilieron, respectivamente, un total de 40 y 128 variantes BRIDGES de los límites intrón/exón, de las cuales se seleccionaron 20 y 52 variantes candidatas tras el filtrado bioinformático (**Tabla Suplementaria 1, Artículo 1; Tabla Suplementaria 1, Artículo 3**) [212,213]. Adicionalmente, en *RAD51C* se escogieron otras 20 variantes de sitios de *splicing* potencialmente espliceogénicas procedentes de la base de datos pública ClinVar (**Tabla Suplementaria 1, Artículo 2**). Cabe destacar que la mayor parte de las variantes seleccionadas están ubicadas en la región donadora (54; 59%) (**Tabla 8**). Algo lógico teniendo en cuenta estudios previos, que demuestran que son más frecuentes las variantes que afectan al sitio donador que las que alteran el sitio de *splicing* 3'[56]. Por otro lado, el análisis de las 92 variantes candidatas requirió la construcción de 4 minigenes, uno que incluye siete de los nueve exones de *RAD51C* (exones 2-8; mgR51C_ex2-8) y 3 minigenes que engloban todos los exones de *CHEK2* (mgChk2_ex1-7, mgChk2_ex6-10 y mgChk2_ex11-15) (**Figura 1, Artículo 1; Figura Suplementaria 1, Artículo 1; Figura 1, Artículo 3; Figura Suplementaria 2, Artículo 3**).

Tabla 8. Resumen del análisis bioinformático, resultados de *splicing* y clasificación clínica siguiendo las normas ACMG/AMP de las 92 variantes de sitios de *splicing* testadas en esta tesis.

Variante (HGVS) ¹	Gen ²	Resumen bioinformático		% Transcritos aberrantes ⁵	ACMG/AMP ⁶
		MES ³	SpliceAI ⁴		
Variantes en la posición intrónica -8					
c.1260-8A>G	<i>CHEK2</i> (ivs11)	[-]3'ss (2,0⇒-3,8) [+]3'ss (9,7) 7-nt upstream [=] Cr, 3'ss (3,8) 33-nt downstream	▼(E12p7)	100% (▼(E12p7))	LP (+9)
Variantes en la posición intrónica -6					
c.405-6T>A	<i>RAD51C</i> (ivs2)	[-]3'ss (7,7⇒2,2) [+] 3'ss (8,6) 4-nt upstream	▼(E3p4)	100% (▼(E3p4))	LP
c.1096-6T>G	<i>CHEK2</i> (ivs10)	[-]3'ss (9,2⇒2,0) [+]3'ss (3,2) 5-nt upstream	▼(E11p5)+Δ(E11)	100% (▼(E11p5))	LP (+9)
Variantes en la posición intrónica -5					
c.1009-5T>A	<i>CHEK2</i> (ivs9)	[-]3'ss (6,9⇒3,8)	▼(E10p142)+ Δ(E10)#	98,2% (Δ(E10))	VUS (+5)
Variantes en la posición intrónica -3					
c.146-3C>G	<i>RAD51C</i> (ivs1)	[-]3'ss (9,5⇒1,9)	Δ(E2)	100% (Δ(E2))	LP (+9)
c.146-3C>T	<i>RAD51C</i> (ivs1)	[↓]3'ss (9,5⇒8,7)	FL	0%	VUS
c.320-3C>G	<i>CHEK2</i> (ivs2)	[-]3'ss (7,7⇒1,7) [+]3'ss (3,9) 2-nt upstream	Δ(E3)#+▼(E3p2)	100% (Δ(E3))	LP (+9)
c.572-3C>G	<i>RAD51C</i> (ivs3)	[-]3'ss (7,4⇒-1,4)	Δ(E4)	100% (Δ(E4))	LP (+9)
c.905-3C>G	<i>RAD51C</i> (ivs6)	[-]3'ss (8,2⇒-4,9)	Δ(E7)	100% (Δ(E7))	LP
c.966-3C>A	<i>RAD51C</i> (ivs7)	[-]3'ss (7,3⇒4,4) [↓] Cr, 3'ss (5,4) 3-nt upstream	Δ(E8)	98% (Δ(E8))	VUS
Variantes en la posición intrónica -2					
c.320-2A>G	<i>CHEK2</i> (ivs2)	[-]3'ss (7,7⇒-0,3)	Δ(E3)#+▼(E3p28)	100% (Δ(E3))	LP (+9)
c.445-2A>G	<i>CHEK2</i> (ivs3)	[-]3'ss (11,3⇒3,3)	Δ(E4)	100%(Δ(E4)▼(E4A38))	LP (+9)
c.684-2A>G	<i>CHEK2</i> (ivs5)	[-]3'ss (10,2⇒2,3) [+]3'ss-TG (3,4) 1-nt upstream	Δ(E6p14)	100%(▼(E6p1))	LP (+9)
c.706-2A>C	<i>RAD51C</i> (ivs4)	[-]3'ss (11,1⇒3,1) [+]3'ss (3,3) 10-nt downstream	Δ(E5p10)+Δ(E5)	100% (Δ(E5p10))	LP
c.706-2A>G	<i>RAD51C</i> (ivs4)	[-]3'ss (11,1⇒3,1) [+]3'ss (3,2) 10-nt downstream	Δ(E5p10)+Δ(E5)	100% (Δ(E5))	P
c.793-2A>G	<i>CHEK2</i> (ivs6)	[-]3'ss (6,2⇒-1,8) [+]3'ss (5,6) 1-nt upstream	▼(E7p1)	100%(▼(E7p1))	LP (+9)
c.838-2A>G	<i>RAD51C</i> (ivs5)	[-]3'ss (10,2⇒2,2)	Δ(E6)	100%(Δ(E6))	LP (+9)

Variante (HGVS) ¹	Gen ²	Resumen bioinformático		% Transcritos aberrantes ⁵	ACMG/AMP ⁶
		MES ³	SpliceAI ⁴		
c.905-2A>C	<i>RAD51C</i> (ivs6)	[-]3'ss (8,2⇒0,1)	Δ(E7)	100%(Δ(E7))	P
c.966-2A>G	<i>RAD51C</i> (ivs7)	[-]3'ss (7,3⇒-0,7) [↑] Cr, 3'ss (7) 3-nt upstream	Δ(E8)	100%(Δ(E8))	VUS
c.966-2A>T	<i>RAD51C</i> (ivs7)	[-]3'ss (7,3⇒-1,1) [↑] Cr, 3'ss (7,6) 3-nt upstream	Δ(E8)	100%(Δ(E8))	VUS
Variantes en la posición intrónica -1					
c.405-1G>C	<i>RAD51C</i> (ivs2)	[-]3'ss (7,7⇒-0,4) [+]3'ss (4,2) 7-nt downstream	Δ(E3p7)+Δ(E3)	100%(Δ(E3p7))	LP (+9)
c.572-1G>C	<i>RAD51C</i> (ivs3)	[-]3'ss (7,4⇒-0,6)	Δ(E4)	100%(Δ(E4))	LP (+9)
c.572-1G>T	<i>RAD51C</i> (ivs3)	[-]3'ss (7,4⇒-1,2)	Δ(E4)	100% (Δ(E4))	LP
c.706-1G>T	<i>RAD51C</i> (ivs4)	[-]3'ss (11,1⇒2,5) [+]3'ss (4,3) 10-nt downstream	Δ(E5p10)	100%(Δ(E5p10))	LP (+9)
c.793-1G>T	<i>CHEK2</i> (ivs6)	[-]3'ss (6,2⇒-2,4)	Δ(E7)+Δ(E7p7)	100%(Δ(E7))	VUS (+8)
c.793-1G>A	<i>CHEK2</i> (ivs6)	[-]3'ss (6,2⇒-2,6) [+]3'ss (4) 1-nt downstream	Δ(E7p1)+Δ(E7)#	100%(Δ(E7p1))	VUS (+8)
c.966-1G>C	<i>RAD51C</i> (ivs7)	[-]3'ss (7,3⇒-0,8) [↑] Cr, 3'ss (6,8) 3-nt upstream	Δ(E8)	100%(▼(E8p3))	VUS (+2)
c.1376-1G>C	<i>CHEK2</i> (ivs12)	[-]3'ss (6,0⇒-2,1)	Δ(E13)+Δ(E13p17)	100%(Δ(E13))	LP (+9)
Variantes en el primer nt exónico					
c.-6G>A	<i>CHEK2</i> (Ex2)	[-]3'ss (1,7⇒0,3)	FL	25,1%(Δ(E4)▼(E4A38)))	LB (-4)
Variantes en el segundo nt exónico					
c.-5T>G	<i>CHEK2</i> (Ex2)	[-]3'ss (1,7⇒-3,2)	FL	50,7%(▼(I1 ^{mg}))	VUS (0)
Pequeñas deleciones que afectan al sitio 3'ss					
c.-6-6_-6-4del	<i>CHEK2</i> (ivs1)	[-]3'ss (1,7⇒-3,7)	▼(E2p48)	7,5% (▼(I1 ^{mg}))	LB (-3)
c.146-4 146-2del	<i>RAD51C</i> (ivs1)	[-]3'ss (9,5⇒-0,8) [+]3'ss (5,9) 3-nt downstream	FL	100% (Δ(E2))	VUS (+2)
c.593-11_593-7del	<i>CHEK2</i> (ivs4)	[↓]3'ss (9,2⇒7,4)	Δ(E5)#+Δ(E5p27)	87,7% (Δ(E5))	VUS (0)
c.847-14 847-2delinsGG⁷	<i>CHEK2</i> (ivs7)	[-]3'ss (9,4⇒-8,3)		100% (Δ(E8))	LP (+9)
c.905-3 906del	<i>RAD51C</i> (ivs6)	[-]3'ss (8,2⇒-8,6) [+]3'ss (4,5) 7-nt downstream	Δ(E7)	100% (Δ(E7))	LP (+9)
c.905-2del	<i>RAD51C</i> (ivs6)	[-]3'ss (8,2⇒2,1)	Δ(E7)	100% (Δ(E7))	LP (+9)
c.905-2 905-1del	<i>RAD51C</i> (ivs6)	[-]3'ss (8,2⇒0,6)	Δ(E7)	100% (Δ(E7))	P
c.1543-9 1546del	<i>CHEK2</i> (ivs14/Ex15)	[-]3'ss (8,5⇒-8,8) [+]3'ss (6,4) 11-nt downstream	Δ(E15p11)+▼(I14 ^{mg})	100%(Δ(E15p11))	LP (+9)

Variante (HGVS) ¹	Gen ²	Resumen bioinformático		% Transcritos aberrantes ⁵	ACMG/AMP ⁶
		MES ³	SpliceAI ⁴		
Variantes en el antepenúltimo nt exónico					
c.1459C>T p.(Gln487Ter)	CHEK2 (Ex13)	[↓]5'ss (8,6⇒5,6)	FL	26,6%(Δ(E13))	LP (+9)
Variantes en el último nt exónico					
c.-7G>A	CHEK2 (Ex1)	[↓]5'ss (10,5⇒7,4) [=] Cr, 5'ss (8,1) 13-nt <i>upstream</i>	Δ(E1q13)#	93,5% (Δ(E1q13))	VUS (+1)
c.-7G>T	CHEK2 (Ex1)	[↓]5'ss (10,5⇒7,4) [=] Cr, 5'ss (8,1) 13-nt <i>upstream</i>	Δ(E1q13)#	75,8% (Δ(E1q13))	VUS (+1)
c.319G>A p.(Glu107Lys)	CHEK2 (Ex2)	[-]3'ss (8,9⇒3,8)	Δ(E2)	74% (Δ(E2))	VUS (0)
c.404G>A p.(Cys135Tyr)	RAD51C (Ex2)	[-]5'ss (4,8⇒-3,5) [=]Cr, 5'ss (5,4) 27-nt <i>downstream</i>	Δ(E2)+▼(E2q27)	100%(▼(E2q27))	LP
c.705G>T p.(Lys235Asn)	RAD51C (Ex4)	[-]5'ss (9,1⇒2,6)	Δ(E4)	100%(Δ(E4))	LP
c.904G>A p.(Gly302Arg)	RAD51C (Ex6)	[-]5'ss (5,6⇒1) [=] Cr, 5'ss (6,2) 4-nt <i>downstream</i>	Δ(E6)+▼(E6q4)	97,6%(▼(E6q4))	LP (+9)
c.1095G>C p.(Lys365Asn)	CHEK2 (Ex10)	[-]5'ss (10,6⇒7,7)	Δ(E10)	94,2%(Δ(E10))	VUS (+5)
c.1375G>A p.(Ala459Thr)	CHEK2 (Ex12)	[-]5'ss (9,1⇒3,9)	Δ(E12)	100%(Δ(E12))	LP (+9)
c.1542G>T p.(Gln514His)	CHEK2 (Ex14)	[-]5'ss (6,5⇒-4,3)	Δ(E14q12)	100%(Δ(E14q12))-	VUS (+2)
Variantes en la posición intrónica +1					
c.-7+1G>T	CHEK2 (ivs1)	[-]5'ss (10,5⇒2,0) [=] Cr, 5'ss (8,1) 13-nt <i>upstream</i>	Δ(E1q13)	100%(Δ(E1q13))	VUS (+1)
c.444+1G>T	CHEK2 (ivs3)	[-]5'ss (8,1⇒-0,4) [+]5'ss (6,3)1-nt <i>upstream</i>	Δ(E3q1)	100%(Δ(E3q1))	LP (+9)
c.444+1G>A	CHEK2 (ivs3)	[-]5'ss (8,1⇒-0,1) [=]5'ss (7,6) 4-nt <i>downstream</i>	▼(E3q4)	100%(▼(E3q4))	P (+12)
c.705+1G>A	RAD51C (ivs4)	[-]5'ss (9,1⇒0,9)	Δ(E4)	100%(Δ(E4))	LP (+9)
c.792+1G>A	CHEK2 (ivs6)	[-]5'ss (9,1⇒0,9)	▼(E6q119)+Δ(E6)	100%(▼(E6q119))	LP (+9)
c.837+1G>T	RAD51C (ivs5)	[-]5'ss (8,6⇒0,1)	Δ(E5)	100%(Δ(E5))	LP (+9)
c.846+1G>T	CHEK2 (ivs7)	[-]5'ss (8,3⇒-0,2)	Δ(E7)	100%(Δ(E7))	LP (+9)

Variante (HGVS) ¹	Gen ²	Resumen bioinformático		% Transcritos aberrantes ⁵	ACMG/AMP ⁶
		MES ³	SpliceAI ⁴		
<u>c.904+1G>T</u>	<i>RAD51C</i> (ivs6)	[-]5'ss (5,6⇒-3,0) [=] Cr, 5'ss (6,2) 4-nt downstream	Δ(E6)	100%(▼(E6q4))	LP (+9)
<u>c.908+1G>T</u>	<i>CHEK2</i> (ivs8)	[-]5'ss (8,5⇒0,0)	Δ(E8)	100%(Δ(E8))	LP (+9)
<u>c.965+1G>A</u>	<i>RAD51C</i> (ivs7)	[-]5'ss (8,7⇒0,5)	Δ(E7)	100%(Δ(E7))	LP (+9)
<u>c.1095+1G>A</u>	<i>CHEK2</i> (ivs10)	[-]5'ss (10,6⇒2,4)	Δ(E10)	100%(Δ(E10))	VUS (+5)
Variantes en la posición intrónica +2					
<u>c.319+2T>A</u>	<i>CHEK2</i> (ivs2)	[-]5'ss (8,9⇒0,7)	Δ(E2)	100%(Δ(E2))	P (+12)
<u>c.404+2T>C</u>	<i>RAD51C</i> (ivs2)	[-]5'ss (4,8⇒-3,0) [=] Cr, 5'ss (5,4) 27-nt downstream	Δ(E2)+▼(E2q27)	100%(▼(E2q27))	LP (+9)
<u>c.592+2T>G</u>	<i>CHEK2</i> (ivs4)	[-]5'ss (8,5⇒0,8)	Δ(E4)	100%(Δ(E4)▼(E4A38))-	LP (+9)
<u>c.837+2T>C</u>	<i>RAD51C</i> (ivs5)	[-]5'ss (8,6⇒0,8)	Δ(E5)	100%(Δ(E5))	LP
Variantes en la posición intrónica +3					
<u>c.404+3A>G</u>	<i>RAD51C</i> (ivs2)	[-]5'ss (4,8⇒0,6) [=] Cr, 5'ss (5,4) 27-nt downstream	▼(E2q27)+ Δ(E2)	73,7%(▼(E2q27))	VUS (+1)
<u>c.444+3A>G</u>	<i>CHEK2</i> (ivs3)	[↓]5'ss (8,1⇒5,4) [↓] Cr, 5'ss (5,7) 4-nt downstream	FL	26,1%(Δ(E4)▼(E4A38))	LB (-4)
<u>c.592+3A>T</u>	<i>CHEK2</i> (ivs4)	[-]5'ss (8,5⇒1,8)	Δ(E4)	100%(Δ(E4)▼(E4A38))	VUS (+8)
<u>c.683+3A>G</u>	<i>CHEK2</i> (ivs5)	[↓]5'ss (8,9⇒6,4)	Δ(E5)	86,4%(Δ(E5))	VUS (+1)
<u>c.705+3A>G</u>	<i>RAD51C</i> (ivs4)	[↓]5'ss (9,1⇒4,6) [+] 5'ss (6,1) 2-nt downstream	FL	78,7%(Δ(E4))	VUS (+1)
<u>c.908+3A>T</u>	<i>CHEK2</i> (ivs8)	[-]5'ss (8,5⇒-3,4)	Δ(E8)	100%(Δ(E8))	LP (+9)
<u>c.1542+3A>G</u>	<i>CHEK2</i> (ivs14)	[-]5'ss (6,5⇒1,2)	Δ(E14q12)#	100%(▼(E14q2))-	VUS (+5)
Variantes en la posición intrónica +4					
<u>c.571+4A>G</u>	<i>RAD51C</i> (ivs3)	[↓]5'ss (10,5⇒8,1) [+] 5'ss (5,5) 4-nt downstream	Δ(E3)+▼(E3q4)	94,6%(Δ(E3))	LP
<u>c.592+4A>G</u>	<i>CHEK2</i> (ivs4)	[↓]5'ss (8,5⇒5,7)	Δ(E4)#	64,7%(Δ(E4)▼(E4A38))	VUS (+1)
<u>c.1008+4A>G</u>	<i>CHEK2</i> (ivs9)	[↓]5'ss (8,6⇒6,7)	FL	7,2%(Δ(E10))	LB (-3)

Variante (HGVS) ¹	Gen ²	Resumen bioinformático		% Transcritos aberrantes ⁵	ACMG/AMP ⁶
		MES ³	SpliceAI ⁴		
Variantes en la posición intrónica +5					
c.319+5G>T	CHEK2 (ivs2)	[↓]5'ss (8,9⇒4,2)	Δ(E2)#	77,2%(Δ(E2))	VUS (+1)
c.571+5G>A	RAD51C (ivs3)	[↓] 5'ss (10,5⇒5,8)	Δ(E3)	100%(Δ(E3))	P
c.705+5G>C	RAD51C (ivs4)	[↓]5'ss (9,1⇒7,2)	FL	48,8%(Δ(E4))	VUS
c.846+5G>A	CHEK2 (ivs7)	[-]5'ss (8,3⇒2,1)	Δ(E7)	100%(Δ(E7))	LP (+9)
c.965+5G>A	RAD51C (ivs7)	[↓]5'ss (8,7⇒3,8)	Δ(E7)	100%(Δ(E7))	LP
c.1026+5G>T	RAD51C (ivs8)	[-]5'ss (2,00⇒-6,3)	Δ(E8)	100%(Δ(E8))	LP
c.1259+5G>C	CHEK2 (ivs11)	[↓]5'ss (9,5⇒4,6)	Δ(E11)	31,7%(Δ(E11))	VUS (+1)
c.1461+5G>T	CHEK2 (ivs13)	[-]5'ss (8,6⇒2,2)	Δ(E13)	100%(Δ(E13))	LP (+9)
Variantes en la posición intrónica +6					
c.444+6T>C	CHEK2 (ivs3)	[↓]5'ss (8,1⇒7,3)	FL	26,7%(Δ(E4)▼(E4A38))	LB (-3)
c.683+6T>C	CHEK2 (ivs5)	[↓]5'ss (8,9⇒8,1)	FL	33%(Δ(E1q13))	VUS (+1)
c.846+6T>C	CHEK2 (ivs7)	[↓]5'ss (8,3⇒5,6)	Δ(E7)	91,8%(Δ(E7))	LP (+9)
Pequeñas deleciones que afectan al sitio 5'ss					
c.-7+4 -7+13del	CHEK2 (ivs1)	[-]5'ss (10,5⇒3,1) [=] Cr, 5'ss (8,1) 13-nt upstream	Δ(E1q13)	100%(Δ(E1q13))	VUS (+1)
c.319+4del	CHEK2 (ivs2)	[-]5'ss (8,9⇒-0,6)	Δ(E2)	100%(Δ(E2))	LP (+9)
c.571+1del	RAD51C (ivs3)	[-]5'ss (10,5⇒-14,1) [+]5'ss (11,1) 1-nt upstream	Δ(E3q1)	100%(Δ(E3q1))	LP (+9)
c.784 792+5del	CHEK2 (Ex6/ivs6)	[-]5'ss (9,1⇒-40,6)	▼(E6q119) ⁸ +Δ(E6)	100% (Δ(E6q9)▼(E6q114))	LP (9)
c.837+4 837+7del	RAD51C (ivs5)	[-]5'ss (8,6⇒-8,9)	Δ(E5)	100%(Δ(E5))	LP (+9)
c.846+4 846+7del	CHEK2 (ivs7)	[-]5'ss (8,3⇒-4,4)	Δ(E7)	100%(Δ(E7))	LP (+9)
c.1026+5 1026+7del	RAD51C (ivs8)	[-]5'ss (2,00⇒-6,3)	Δ(E8)	100%(Δ(E8))	P
c.1375+1 1375+2del	CHEK2 (ivs12)	[-]5'ss (9,1⇒-1,8)	Δ(E12)+▼(E12q30) ⁹	100%(Δ(E12))	LP (+9)
<p>¹Se subrayan las variantes que no dejan rastro (o ≤5%) del transcrito <i>full-length</i>. ²Gen y exón (Ex) y/o intrón (ivs) en el que se localiza la mutación investigada. ³[-] Interrupción del sitio de <i>splicing</i>; [+] sitio nuevo; [↓] se reduce la fortaleza del sitio de <i>splicing</i>; [↑] se incrementa la fortaleza del sitio de <i>splicing</i>; [=] no se altera la fortaleza del sitio de <i>splicing</i>, y Cr, sitio críptico. ⁴Los transcritos predichos por SpliceAI están anotados como sigue: Δ (<i>skipping</i> de secuencias exónicas), ▼ (inclusión de secuencias intrónicas), E (exón) y, cuando es necesario, p (desplazamiento del aceptor) y q (desplazamiento del donador) + nt insertados o eliminados. Simplemente con fines informativos, cuando el Δ <i>score</i> de SpliceAI es bajo (<0,20), pero los resultados predichos están apoyados por los datos experimentales se indican con #.⁵Se indica el porcentaje de transcritos aberrantes y entre paréntesis el transcrito aberrante más abundante. ⁶Clasificación clínica de las 92 variantes testadas en esta tesis siguiendo las normas ACMG/AMP. ⁷c.847-14_847-2delinsGG se ha excluido del análisis, ya que SpliceAI actualmente no admite InDels complejas. ⁸▼(E6q119) para simplificar. Dado que la alteración del <i>splicing</i> está causada por una deleción de 14 nt, el tamaño real de la inserción es 14 nt más corto. ⁹▼(E12q30) para simplificar. Dado que la alteración del <i>splicing</i> está causada por una deleción de 2 nt, el tamaño real de la inserción es de 28 nt.</p>					

Los resultados obtenidos confirmaron que los ensayos empleando el minigén mgR51C_ex2-8 son capaces de reproducir los patrones de *splicing* detectados en ARN de pacientes. Específicamente, cuatro variantes inducían resultados muy similares o incluso idénticos a los descritos en ARN de individuos portadores: c.571+4A>G ($\Delta(E3)$) [214], c.706-2A>G ($\Delta(E5)$) [215], c.905-2_905-1del ($\Delta(E7)$) [216] y c.1026+5_1026 +7del ($\Delta(E8)$) [217]. Asimismo, las mutaciones c.837+2T>C y c.905-3C>G/c.905-3_906del/c.905-2A>C/c.905-2del replicaron los perfiles de *splicing* encontrados en muestras de pacientes portadoras de las variantes c.837+1G>A y c.905-2A>G, que afectan a los mismos sitios de *splicing*, respectivamente [218,219]. Por último, los cambios puntuales c.404G>C/G>T, localizados en una posición idéntica a la variante c.404G>A testada en nuestro minigén, promovieron el uso del mismo sitio de *splicing* crítico 27-nt *downstream* ($\nabla(E2q27)$) del donador canónico (**Tabla 1, Artículo 1; Figura 2, Artículo 1; Tabla 1, Artículo 2; Figura 1, Artículo 2**) [220]. En conjunto, estos datos apoyan la reproducibilidad del enfoque minigénico. No obstante, es interesante remarcar que en las muestras de pacientes portadoras de las variantes c.571+4A>G, c.706-2A>G, y c.1026+5_1026+7del se detecta únicamente el transcrito anómalo principal producido por nuestros minigenes mutantes ($\Delta(E3)$ -76,5%, $\Delta(E5)$ -65,4%, y $\Delta(E8)$ -78,0%, respectivamente), pero no el resto de isoformas de ARN que aparecen en proporciones menores. Estas ligeras variaciones pueden deberse a varias razones, entre ellas: (i) el *splicing* alternativo específico de tejido, ya que las RT-PCR de pacientes se realizan normalmente a partir de ARN de sangre; (ii) la alta sensibilidad del análisis de fragmentos fluorescentes, que permite la identificación de isoformas raras; (iii) el uso de inhibidores del NMD en los ensayos funcionales con minigenes híbridos (algo que no suele hacerse en las muestras procedentes de pacientes), permitiendo la detección de transcritos PTC que habitualmente son degradados a través de esta vía; (iv) la presencia del alelo de tipo WT en muestras de pacientes que interfiere en la interpretación de los resultados, y (v) la alta tasa de transcripción en los minigenes híbridos desencadenada por un promotor SV40 fuerte. Del mismo modo, el constructo WT que inducía un único transcrito del tamaño (1062 nt), secuencia y estructura (V1- *RAD51C* exones 2 a 8- V2) esperados, no replicó exactamente el patrón de *splicing* de células MCF-7 o de ARN de mama control, los cuales mostraron transcritos alternativos menores (**Figura 1, Artículo 1**). Por lo tanto, adicionalmente, deben tenerse en cuenta otros factores propios de los minigenes, como la ausencia del contexto genómico natural debida al acortamiento de los intrones clonados. Por ende, podríamos afirmar que la ausencia de posibles SRE intrónicos y la arquitectura

natural exón/intrón podrían influir de alguna manera en los resultados de *splicing* que ofrecen los minigenes WT y mutantes [27]. Otra forma de validar la reproducibilidad de este minigén fue comprobar que 3 variantes (c.405-1G>C, c.706-1G>T y c.904G>A), así como el minigén WT, exhibían el mismo efecto en el *splicing* al ser testadas en células MCF-7 y MDA-MB-231 (**Figura 2, Artículo 2**).

En relación con los minigenes de *CHEK2*, los tres reproducen con bastante precisión los patrones de *splicing* previamente detectados en condiciones fisiológicas, de modo que son adecuados para el testado funcional de las variantes candidatas [191,221]. Ahora bien, es importante destacar que la falta de contexto genómico en la región 3' del exón 10 y la región 5' del 11 podría representar una limitación importante para el análisis de las variantes ubicadas en los sitios 5'ss del exón 10 y 3'ss del exón 11 (c.1095G>C, c.1095+1G>A y c.1096-6T>G). Pese a ello, las predicciones de SpliceAI para estas variantes coinciden con los resultados funcionales de los minigenes mutantes, respaldando la reproducibilidad de los mismos (**Tabla 1, Artículo 3; Figura 2, Artículo 3; Figura 3, Artículo 3; Tabla Suplementaria 7, Artículo 3**). Más aún, los resultados de seis variantes de *CHEK2* mimetizaron los perfiles de *splicing* previamente identificados en ensayos de ARN de portadores: c.444+1G>T (Δ (E3q1)) [222], c.444+1G>A (∇ (E3q4)) [223], c.793-1G>A (Δ (E7p1)) [224], c.846+1G>T (Δ (E7)) [225], c.846+4_846+7del (Δ (E7)) y c.1260-8A>G (∇ (E12p7)) [226]. En cambio, el análisis de ARN de portadoras de la variante c.592+3A>T mostró la producción de dos transcritos aberrantes diferentes, uno carente del exón 4 y otro desprovisto de los exones 4 y 5. Sin embargo, no se detectó el uso de un exón 4 no constitutivo (∇ (E4A38)), la isoforma más abundante de acuerdo con los resultados encontrados tras el estudio funcional de esta variante empleando el minigén mgChk2_ex1-7, lo cual podría explicarse por las razones enumeradas en el párrafo anterior [227,228]. Adicionalmente, 12 variantes (4 por minigén) fueron ensayadas en células HeLa, mostrando patrones de *splicing* similares, salvo la variante c.320-3C>G, en la que se observaron proporciones ligeramente diferentes entre ambas líneas celulares de los transcritos más expresados (**Figura Suplementaria 4, Artículo 3**).

Sorprendentemente, una elevada proporción de las variantes analizadas (85/92; 92,4%) disrumpían el *splicing*, de las cuales 69 producían anomalías graves (<5% de transcrito *full-length*), corroborando la precisión de nuestro método bioinformático de selección (**Tabla 1, Artículo 1; Figura 2, Artículo 1; Tabla 1, Artículo 2; Figura 1,**

Artículo 2; Tabla 1, Artículo 3; Figura 2, Artículo 3; Figura 3, Artículo 3). En total, 42 variantes espliceogénicas estaban afectando a los nucleótidos intrónicos clásicos ($\pm 1,2$ nt), mientras que las 43 restantes alteraban otras posiciones de los sitios de *splicing*, poniendo de relieve el potencial deletéreo de cualquier cambio en los nucleótidos conservados de los sitios 5'ss y 3'ss. De todas las variantes con efecto en el *splicing*, diecisiete produjeron cantidades variables del mgFL. Desafortunadamente, se desconoce la cantidad mínima de expresión de los genes *RAD51C* y *CHEK2* necesaria para conferir haplosuficiencia, de ahí que estos ensayos de *splicing* no se consideraran informativos (PVS1_O_N/A). Respecto a las 7 variantes no espliceogénicas, 5 (c.146-3C>T de *RAD51C*; c.444+3A>G, c.444+6T>C, c.683+6T>C y c.1008+4A>G de *CHEK2*) modificaron débilmente los valores MES de los sitios de *splicing* canónicos y en tres de ellas (c.146-3C>T de *RAD51C*; c.444+3A>G y c.1008+4A>G de *CHEK2*) la sustitución en cuestión añade el segundo nucleótido más prevalente en dichas posiciones, lo que justifica la ausencia de impacto de las mismas.

En cuanto a las mutaciones que están perturbando la región aceptora, las ubicadas en la posición -3 que introducen un cambio de C/T a A/G es muy probable que produzcan alteraciones severas del proceso de *splicing* (p. ej., las variantes c.905-3C>G y c.966-3C>A de *RAD51C*). Por otra parte, si una mutación en el tracto de polipirimidina combina el debilitamiento del sitio aceptor y la generación de un sitio *de novo*, es previsible que exhiba un impacto fuerte sobre el *splicing* (p. ej., la variante c.405-6T>A de *RAD51C*; **Tabla 8**). En lo que respecta a las mutaciones ubicadas en el sitio 5'ss, despiertan nuestro interés principalmente las sustituciones +2T>C, ya que estarían creando un donador GC atípico, los cuales representan menos del 1% de todos los sitios donadores humanos [25]. Se observó que alrededor del 15-18% de los cambios +2T>C conservan la actividad del sitio 5'ss, induciendo la producción de entre un 1% y un 84% de transcrito *full-length* [229]. Aun así, ninguna de las dos variantes +2T>C testadas (c.404+2T>C y c.837+2T>C) indujeron el uso del dinucleótido GC atípico creado *de novo*, al igual que la variante de *PALB2* c.48+2T>C [59,206]. Por el contrario, la mutación c.108+2T>C de *PALB2* creaba un donador GC funcional que producía un 85% de transcrito *full-length* [59]. Este último resultado puede estar relacionado con el hecho de que el resto de las posiciones que configuran el 5'ss muestran el nucleótido más común según la secuencia consenso de este sitio de *splicing* (CAG|GCAAGT). Por otro lado, además de los cambios en las posiciones +1,2, las variantes que modifican otras posiciones conservadas de la región donadora, como por ejemplo los últimos nt del exón o las posiciones +3, +4, +5 y +6, también son

capaces de inducir patrones de *splicing* anómalos. De hecho, 29 de las 36 mutaciones ensayadas en estas posiciones exhiben un impacto fuerte o muy fuerte, confirmando, de nuevo, la elevada espliceogenicidad de cualquier cambio en la secuencia consenso de los sitios de *splicing* (**Tabla 8**).

Además, cabe destacar los distintos perfiles generados por tres cambios genéticos puntuales que modifican el mismo sitio de *splicing* en *RAD51C*: c.706-2A>C/-2A>G/-1G>T (**Tabla 1, Artículo 1; Figura 2, Artículo 1; Tabla 1, Artículo 2; Figura 1, Artículo 2**). Las variantes c.706-2A>C y c.706-1G>T promovieron principalmente el uso de un 3' ss críptico 10 nt *downstream* (Δ (E5p10); 91,4% y 100%, respectivamente), mientras que la c.706-2A>G indujo mayoritariamente el *skipping* del exón 5 (Δ (E5); 65,4%), el cual se detectaba también junto con el transcrito Δ (E5p10) (33,5%). Una explicación plausible de estos resultados de *splicing* tan diversos es que las mutaciones c.706-2A>C y c.706-1G>T, al introducir un cambio de purina a pirimidina, estarían reforzando el PPT del sitio aceptor críptico 10 nt *downstream*, algo que no ocurre con la variante c.706-2A>G que incorpora en el ADN otra purina. Otros resultados de *splicing* interesantes son los originados por las 4 mutaciones que afectan a la región aceptora del exón 8 (c.966-3C>A, c.966-2A>G, c.966-2A>T y c.966-1G>C), en las que, a pesar de que los algoritmos bioinformáticos empleados ofrecen unas predicciones similares para todas ellas, se observan diferencias importantes en los patrones de *splicing*. Mientras que en las variantes c.966-3C>A, c.966-2A>G y c.966-2A>T el *skipping* del exón 3 es el evento más frecuente (Δ (E3); 86,8%, 86,7% y 89,1%, respectivamente), la mutación c.966-1G>C induce predominantemente el uso de un aceptor críptico 3 nt *upstream* (∇ (E8p3); 79,4%). Así pues, teniendo en cuenta la imprevisibilidad de los resultados de *splicing*, los ensayos de ARN son muy recomendables para investigar el impacto exacto de las variantes.

A su vez, la electroforesis capilar fluorescente de los productos de RT-PCR generados por los minigenes de *RAD51C* y *CHEK2* WT y mutantes logró identificar 142 isoformas de ARN diferentes, 115 de las cuales pudieron caracterizarse (**Tabla Suplementaria 2, Artículo 1; Tabla Suplementaria 3, Artículo 2; Tabla Suplementaria 4, Artículo 3**). En concreto, 82 de los transcritos descritos introducen PTC, 15 isoformas de ARN mantienen el ORF, 5 eliminan el codón de inicio, 2 incorporan InDels en la región 5'UTR y 11 de estos productos RT-PCR se corresponden con el mgFL, WT o con mutaciones exónicas puntuales. Esta técnica analítica de

separación ofreció una alta resolución y sensibilidad, siendo capaz de detectar isoformas de ARN raras y distinguir transcritos que difieren en unos pocos nucleótidos, como, por ejemplo, el transcrito mgFL de los productos ▼(E8p3)-a,b,c,d, de *RAD51C* en los que se insertan 3 nt. Las anotaciones ▼(E8p3)-a,b,c,d hacen referencia a cuatro versiones diferentes de esta inserción intrónica de 3 nt en el 3' ss del exón 8: ▼(E8p3)-a (r.[966-3c>a,965_966ins966-3_966-1]; 9,7%), ▼(E8p3)-b (r.[966-2a>g,965_966ins966-3_966-1]; 11,0%), ▼(E8p3)-c (r.[966-2a>u,965_966ins966-3_966-1]; 5,9%) y ▼(E8p3)-d (r.[966-1g>c,965_966ins966-3_966-1]; 79,4%), producidas por las variantes c.966-3C>A, c.966-2A>G, c.966-2A>T y c.966-1G>C, respectivamente. Esto provocaría tres efectos diferentes a nivel proteico, a saber: p.Arg322dup, p.Arg322delinsSerGly, p.Arg322delinsSerTrp y p.Arg322delinsSerThr. También en *RAD51C* observamos dos anotaciones distintas para el evento de *splicing* que se genera por el uso de un donador críptico ubicado 27 nt *downstream* del canónico en el exón 2 (▼(E2q27)-a,b; un transcrito incluye la mutación r.404+2c y el otro la variante r.404+3g) y para la inserción de 4 nt que afecta a la región donadora del exón 6 (▼(E6q4)-a,b; una isoforma incorpora la variante r.904a y la otra la mutación r.904+1u). Más aún, un total de 13 transcritos (Δ(E2p3), ▼(E2q27), Δ(E2q175), Δ(E2q22), Δ(E2), Δ(E3), Δ(E4), Δ(E4_E5), Δ(E5), Δ(E7), Δ(E7_E8), Δ(E8) y ▼(E8p3)) de los 33 caracterizados en *RAD51C* se habían descrito previamente como isoformas naturales de dicho gen, lo que sugiere que los eventos alternativos fisiológicos pueden predecir de algún modo los perfiles de *splicing* de las variantes (Tabla 1, Artículo 1; Figura 2, Artículo 1; Tabla 1, Artículo 2; Figura 1, Artículo 2) [187,230,231]. Esta hipótesis está también apoyada por los resultados obtenidos en *CHEK2*, en el cual 8 transcritos (Δ(E1q13), Δ(E4), ▼(E4A38), Δ(E5), Δ(E4_E5), Δ(E8), Δ(E10) y Δ(E12p33)) de los 82 caracterizados en este gen habían sido identificados con anterioridad como isoformas fisiológicas (Tabla 1, Artículo 3; Figura 2, Artículo 3; Figura 3, Artículo 3) [191,221]. *CHEK2* es, por tanto, el responsable de la mayor parte de los transcritos descritos (82/115; 71,3%). De hecho, en muestras de tumores de mama se había observado que este gen producía hasta 90 isoformas diferentes, resultados que estaban anticipando esta gran cantidad de eventos de *splicing* anómalos detectados en este gen [193]. En particular, algunas variantes *CHEK2* mostraban patrones de *splicing* extraordinariamente complejos con hasta 11 transcritos diferentes (c.319+5G>T), lo que dificultaba seriamente su interpretación clínica. La mayoría de estas variantes se ubican en los exones 5' del gen *CHEK2* (1 a 6) incluidos en el minigén mgChk2_ex1-7, que son los que exhibieron una mayor variabilidad con 52 transcritos diferentes, en gran parte debido

a que se trata de una región rica en eventos de *splicing* alternativo naturales, como $\Delta(E1q13)$, $\Delta(E4)\nabla(E4A38)$ o $\Delta(E5)$ [191,221]. Estos hallazgos podrían reflejar la naturaleza intrínsecamente compleja de la regulación del proceso de *splicing* en *CHEK2*.

De entre todos los transcritos descritos en *CHEK2*, la isoforma de ARNm más interesante fue la inducida por la variante c.684-2A>G, $\nabla(E6p1)$, pues implica el uso de un aceptor TG extremadamente raro en humanos ($\approx 0,02\%$) que, en general, es empleado como un sitio de *splicing* no constitutivo [25,232]. Este transcrito inesperado se explica, al menos en parte, por los nucleótidos circundantes que coinciden estrechamente con la secuencia consenso del aceptor TG, donde en 17 de los 23 nt aparece el primer o el segundo nucleótido más común (**Figura 4D, Artículo 3**). Mediante mapeo funcional, se descubrió que su reconocimiento dependía críticamente del intervalo c.685_698 (que anulaba por completo el uso de este aceptor TG cuando era delecionado), pero también de las regiones c.709_722 y c.707_743, (**Figura 4A-B, Artículo 3**). Es probable que el uso de este aceptor TG esté mediado por potenciadores y/o silenciadores de *splicing* o incluso por la estructura secundaria del ARN, que podría ocultar o exponer motivos de unión para ciertos factores de *splicing* esenciales. Según SpliceAid (<http://www.introni.it/splicing>), las deleciones c.685_698del, c.709_722del y c.707_743del eliminan potenciadores putativos del exón 6, como SRp55, SRp40, SC35 o Tra2 β , entre otros (**Figura 4C, Artículo 3**). En todo caso, el papel de cualquiera de estas proteínas SR en el reconocimiento de este aceptor atípico debe ser confirmado experimentalmente. Curiosamente, las mismas microdeleciones en el minigén WT no tuvieron ningún impacto sobre el *splicing*, lo cual sugiere que la correcta selección del aceptor canónico es independiente de las secuencias necesarias para el reconocimiento de este aceptor atípico TG.

Finalmente, la interpretación clínica de las variantes es un proceso especialmente difícil y laborioso, en el que se consideran múltiples evidencias de patogenicidad y/o benignidad, además de los transcritos producidos por las variantes testadas, como, por ejemplo, la ausencia o no de la mutación en cuestión en controles sanos. Asimismo, la enorme complejidad de los patrones de *splicing* de muchas de las variantes investigadas aumenta aún más la dificultad de este proceso. En consecuencia, para este propósito se colaboró con el Dr. Miguel de la Hoya (Hospital Clínico de San Carlos, Madrid) que propuso una estrategia de clasificación clínica tentativa basada en: i) las normas ACMG/AMP; ii) aspectos específicos de las recomendaciones del ClinGen SVI para interpretar los códigos

de evidencia PVS1 de pérdida de función y PS3/BS3 funcionales; iii) algunos enfoques no específicos de gen desarrollados por el panel de expertos de ClinGen *CDH1* [233], y iv) el juicio de expertos [59,179,180,186,234,235]. Un aspecto muy relevante en la interpretación de una variante es la evaluación del impacto de los transcritos anómalos sobre la funcionalidad del gen, tras la que se le otorgará a los mismos un nivel de patogenicidad muy fuerte, fuerte, moderado o leve o un grado de benignidad leve o fuerte (**Figura Suplementaria 3, Artículo 1; Métodos Suplementarios, Artículo 1; Figura Suplementaria 6, Artículo 3; Material Suplementario; Artículo 3**).

En *RAD51C*, a los transcritos PTC-NMD se les asignó un grado de patogenicidad muy fuerte. Los transcritos PTC ▼(E6q4)-a (p.Gly302SerFs*47), ▼(E6q4)-b, (p.Gly302ValFs*47), Δ(E7) (p.(Glu303Trpfs*41)), Δ(E7_E8) (p.(Gly302Alafs*5)), Δ(E8) (p.(Arg322Serfs*22)), ▼(E8q41) (p.(Pro343Valfs*4)) y ▼(E8q44) p.(Pro343Valfs*5)) que codifican productos proteicos carentes de la señal de localización nuclear, así como de las hebras β6 a la 9 en los 4 primeros, de las hebras β7 a la 9 en el transcrito siguiente y de la hebra β9 en los dos últimos, se consideraron inicialmente pruebas funcionales fuertes en pro de la patogenicidad. Sin embargo, en la clasificación de variantes ClinVar a estos transcritos se les otorgó un nivel muy fuerte de patogenicidad. Este cambio es debido a que las características estructurales de *RAD51C* predicen que las proteínas que carecen de una sola hebra β fallarían en la formación de la hoja β, resultando en el colapso del núcleo de la proteína y el mal plegamiento de la misma [131]. Además, se había demostrado también que la variante *missense* p.Arg312Trp (cadena β6) afectaba a la función de *RAD51C* [236]. En cuanto a las isoformas *in-frame*, a la hora de clasificar las variantes BRIDGES se consideró que la mayor parte de estos transcritos, Δ(E3q114) (p.(Gly153_Glu190del)), Δ(E5) (p.(Arg237_Val280del)) y Δ(E8q18) (p.(Val337_Lys342del)), los cuales se traducen en polipéptidos desprovistos, respectivamente, de la cadena β2, el dominio Walker-B (cadena β4) junto con la cadena β5 y la cadena β8, constituyen una evidencia fuerte en favor de la patogenicidad. Por el contrario, se desconocía el impacto de los transcritos *in-frame* que insertan 3 nt en la región aceptora del exón 8 de *RAD51C* (▼(E8p3)), es por ello por lo que no se fijó un nivel concreto de patogenicidad o benignidad para dichos eventos de *splicing* (**Figura Suplementaria 3, Artículo 1; Métodos Suplementarios, Artículo 1**). Posteriormente, en la clasificación de variantes ClinVar, al transcrito Δ(E2p3) (isoforma alternativa de *RAD51C*), se le atribuyó un grado de patogenicidad leve [230]. Por otro lado, al transcrito Δ(E5) se le concedió el nivel más alto de patogenicidad, dado que: i) deleciona

el dominio Walker-B (cadena $\beta 4$) y la cadena $\beta 5$; ii) elimina 44 aminoácidos, 26 de los cuales están fuertemente conservados en vertebrados (**Figura Suplementaria 2, Artículo 1; Figura Suplementaria 2, Artículo 2**), y iii) en esta región se había descrito una variante *missense* c.773G>A (p.Arg258His) clasificada como LP en ClinVar por haberse encontrado como mutación bialélica en pacientes con anemia de Fanconi [140]. Finalmente, la isoforma ▼(E8p3), que en la interpretación de variantes BRIDGES se consideró no informativa, más tarde se clasificó como una evidencia leve en pro de la patogenicidad. La única isoforma de ARN que se categorizó como no informativa en la interpretación clínica de las variantes ClinVar fue el transcrito *full-length* con el cambio puntual c.904G>C. Además de estas diferencias en los niveles de patogenicidad otorgados a cada uno de los transcritos, en las variantes ClinVar se estableció a mayores un sistema de puntuación que asignaba un valor numérico a cada una de las evidencias de patogenicidad o benignidad individuales evaluadas. En definitiva, gracias a esta estrategia de clasificación preliminar conseguimos catalogar 31 de las 40 variantes del gen *RAD51C* ensayadas como P/LP (**Tabla 2, Artículo 1; Tabla 2, Artículo 2**).

En el caso de *CHEK2*, el mayor grado de patogenicidad (muy fuerte) se le concedió a los transcritos PTC-NMD, a aquellos que incluían el *skipping* del exón 2 (eliminación del codón de inicio de la traducción), a los transcritos *in-frame* que suprimían los aminoácidos esenciales Ile157 (dominio FHA) o Glu273 (KD), y a los dos transcritos PTC que interrumpían la secuencia NLS (**Figura Suplementaria 6, Artículo 3; Material Suplementario, Artículo 3**). Las isoformas *in-frame* que no eliminaban aminoácidos críticos, como los mencionados anteriormente o los ubicados entre la histidina 371 y la leucina 391 (que conforman el *loop* de activación), pero suprimían regiones clave para la función de *CHEK2*, como parte del dominio quinasa del gen, se consideraron evidencias fuertes en favor de la patogenicidad. El resto de transcritos descritos adquirieron valores de patogenicidad bajos, se consideraron benignos o no se les atribuyó un nivel concreto de patogenicidad o benignidad (isoformas que afectan a la región 5'UTR). A partir de las evidencias, tanto predictivas como funcionales de cada uno de los transcritos, se calculó una puntuación agregada por variante, la cual permitió clasificar 5 mutaciones como LB y 27 como P/LP (**Tabla 2, Artículo 3**).

Por tanto, siguiendo las normas ACMG/AMP logramos clasificar 58 variantes de sitios de *splicing* de los genes *CHEK2* y *RAD51C* como P/LP, 29 como VUS y 5 como LB. De esta forma, conseguimos actualizar la clasificación clínica de 30 variantes reportadas en

la base de datos ClinVar, de las cuales 22 estaban catalogadas como variantes de significado clínico desconocido. Las 8 variantes restantes aparecían clasificadas como P/LP por el hecho de estar ubicadas en las posiciones altamente conservadas $\pm 1,2$. Estas mutaciones (c.146-4_146-2del -RAD51C-, c.793-1G>T -CHEK2-, c.793-1G>A -CHEK2-, c.-7+1G>T -CHEK2-, c.966-2A>G -RAD51C-, c.966-2A>T -RAD51C-, c.966-1G>C -RAD51C-, y c.1095+1G>A -CHEK2-) se clasificaron como VUS tras los ensayos funcionales de *splicing* porque generaban transcritos que no conllevan una pérdida de función. Esto pone de manifiesto que, incluso en este tipo de variantes $\pm 1,2$, el estudio funcional es altamente recomendable, puesto que, si bien es cierto que la mayoría de ellas afectarán al *splicing*, la alteración resultante no tiene por qué ser necesariamente patogénica.

Es importante hacer hincapié en el hecho de que las clasificaciones clínicas propuestas para los genes investigados se basan en decisiones no necesariamente compartidas por otros expertos en la materia, lo cual saca a la luz la falta de estandarización en este ámbito. Por otro lado, la aplicación de las normas ACMG/AMP para establecer estas estrategias de clasificación tentativas puso de relieve una serie de incoherencias en el sistema, como:

- I. Las variantes que modifican los nucleótidos $\pm 1,2$ y producen transcritos PTC-NMD se clasifican más fácilmente como patogénicas que las variantes *nonsense/frameshift* que introducen mutaciones PTC-NMD equivalentes. Esto es debido a que a ambos tipos de variantes se les otorga una evidencia muy fuerte de patogenicidad (código PVS1) porque se considera que interrumpen la función génica. Sin embargo, a las variantes $\pm 1,2$ se les asigna también otro criterio fuerte de patogenicidad basado en los resultados del estudio funcional (código PS3).
- II. Las variantes GT-AG $\pm 1,2$ son más fáciles de clasificar como patogénicas que otras variantes espliceogénicas ubicadas fuera de estas posiciones altamente conservadas que producen resultados de *splicing* idénticos, puesto que a estas últimas no se les asigna la evidencia muy fuerte de patogenicidad PVS1. Como máximo podrán recibir una evidencia fuerte, de acuerdo con los resultados del estudio funcional (código PS3), y una leve (código PP3), de acuerdo con las predicciones bioinformáticas. Es más, si a las variantes $\pm 1,2$ se les agrega un criterio de patogenicidad moderado, como el código PM2 otorgado a las variantes raras, se clasifican como probablemente patogénicas sin requerir un ensayo funcional de *splicing*.

De ahí que se propusiera abordar estas cuestiones mediante un enfoque un tanto radical: sustituir las predicciones *in silico* (PVS1, evidencia muy fuerte de patogenicidad, y PP3/BP4, las predicciones computacionales apoyan un efecto perjudicial/sugieren que no hay impacto) por pruebas funcionales, en lugar de combinar ambas (**Métodos Suplementarios, Artículo 1; Material Suplementario, Artículo 3**).

Por otra parte, el proceso de clasificación clínica de estos genes fue especialmente difícil porque: i) aún no estaba disponible la adaptación del panel de expertos de las reglas ACMG/AMP a los genes *CHEK2* y *RAD51C*; ii) la mayor parte de las lecturas de los minigenes eran muy complejas, con varios transcritos detectados (hasta ocho transcritos diferentes en la lectura del mgChk2_ex1-7 c.592+4A>G) y algunos de ellos no totalmente caracterizados; iii) varias variantes espliceogénicas presentaban efectos parciales (expresaban cantidades variables de transcrito *full-length*); iv) los minigenes WT (en particular, el minigén mgChk2_ex1-7) producían unos niveles importantes de transcritos alternativos; v) se disponía de muy pocas evidencias, aparte de los resultados obtenidos mediante minigenes híbridos, y vi) hasta siete variantes investigadas alteraban la región 5'UTR, una zona fuera del alcance de los criterios actuales ACMG/AMP. En consecuencia, nuestro proceso de clasificación dio lugar a una proporción relevante de VUS, sobre todo en el gen *CHEK2* ($\approx 40\%$).

En conclusión, se testaron funcionalmente 92 variantes ubicadas en las fronteras exón/intrón de los genes *RAD51C* y *CHEK2* mediante los minigenes mgR51C_ex2-8, mgChk2_ex1-7, mgChk2_ex6-10 y mgChk2_ex11-15, de las cuales 85 alteraron el proceso de *splicing* y 69 de ellas mostraron un impacto total. De acuerdo con las normas ACMG/AMP conseguimos clasificar 58 variantes como P/LP, mientras que 29 se catalogaron como VUS. Otros parámetros, como los datos de estudios caso-control, segregación familiar, rareza de la variante o asociación con otras enfermedades, incrementarían la significatividad del método de clasificación y proporcionarían estimaciones de riesgo más precisas para guiar el manejo clínico de las pacientes con CM. No obstante, mientras tanto, la metodología de minigenes, empleando el vector de *splicing* pSAD, es un enfoque versátil y adecuado para comprobar el efecto de las variantes potencialmente espliceogénicas, tanto en genes CM como en los vinculados a otras enfermedades genéticas. En definitiva, se puede aseverar que en esta tesis se ha efectuado el estudio más completo de variantes de sitios de *splicing* de los genes *RAD51C* y *CHEK2* presentado hasta la fecha.

2. ELEMENTOS REGULADORES DEL *SPLICING* Y VARIANTES SRE

Otro de los objetivos de esta tesis es el estudio de los elementos funcionales implicados en la regulación de exones que exhiben características especiales (exones alternativos, muy grandes o con sitios de *splicing* atípicos) y la búsqueda de variantes espliceogénicas en los intervalos SRE críticos mapeados. La comprensión de cómo se regula el *splicing* alternativo puede proporcionar información muy valiosa sobre la etiología de las enfermedades genéticas y ayudar al desarrollo de estrategias terapéuticas en patologías asociadas a alteraciones en este proceso [237–239]. Sin embargo, como hemos visto, el reconocimiento de un exón es un proceso altamente complejo que depende de un control combinatorio único de múltiples parámetros [27]. Así pues, estimar el impacto de las variantes sobre el *splicing* es una tarea especialmente compleja, salvo en aquellas que modifican los nucleótidos ubicados en los sitios consenso, para las que se han desarrollado algoritmos más precisos. En el caso de las variantes que afectan a los sitios SRE, la inexactitud de las predicciones *in silico* se explica por el hecho de que los factores reguladores del *splicing* pueden compartir sitios de unión, o bien por estar superpuestos, o bien por tratarse de sitios degenerados, y, además, los SRE muestran una actividad dependiente del contexto genético [29,240]. Un ejemplo claro de esto último son los SRE conocidos como CERES, los cuales pueden ejercer funciones tanto activadoras como represoras [30]. A lo anterior se suma la existencia de factores de *splicing* específicos de tejido como RBFOX1, característico de los tejidos cardíaco, muscular y neuronal [241]. Asimismo, el reconocimiento eficiente de los exones depende de efectos combinatorios o cooperativos/competitivos entre los distintos factores reguladores del *splicing* y sitios SRE, de modo que la destrucción de alguno de estos motivos por una variante puntual tiende a asociarse a alteraciones de *splicing* parciales/débiles [184]. Aun así, una variante SRE también puede mostrar un impacto total sobre el proceso de *splicing* cuando ejerce varios efectos a la vez, como la variante c.891C>T del gen *SPINK5* que simultáneamente debilita un ESE (Tra2 β) y fortalece un ESS (hnRNPA1) [242]. Por lo tanto, el testado funcional de cualquier variante SRE potencialmente espliceogénica es la mejor opción para comprobar su verdadero impacto sobre el *splicing*.

El gen *CHEK2*, como la mayoría de los genes humanos, está sujeto a fenómenos de *splicing* alternativo, pudiéndose detectar en condiciones fisiológicas múltiples transcritos a través de técnicas como el RNA-seq [191,221]. Dos de los eventos alternativos más comunes de *CHEK2* son el *skipping* de los exones 8 y 10 ($\Delta(E8)$ y $\Delta(E10)$), de ahí que se decidiera estudiar la regulación de ambos. También nos interesamos en la regulación del exón 4 de *BARD1*, un macroexón (950 pb) involucrado en 34 de los 62 eventos de *splicing* alternativos previamente descritos en este gen, entre los que la isoforma $\Delta(E4)$ es la más frecuente, sugiriendo un control preciso de su inclusión [199]. Por último, otro de nuestros genes diana es *BRIP1*, debido a su exón número 1, cuyo 5'ss canónico emplea un donador atípico GC (MES=3,10). Estos donadores atípicos solo aparecen en el 1% de los exones humanos [25] y se ha demostrado que están sujetos a una regulación específica [61,243].

En particular, nos centramos en la búsqueda de regiones enriquecidas en ESE esenciales para el reconocimiento de estos exones, en las que, consecuentemente, cualquier variante debe ser considerada candidata de alterar el *splicing* y, por tanto, incrementar el riesgo asociado a la enfermedad. El análisis funcional de *splicing* de microdeleciones exónicas solapantes en minigenes híbridos ha demostrado ser una estrategia poderosa para este propósito, frente a la precisión limitada de las predicciones *in silico* [26,182,244,245]. Dicho enfoque se empleó sistemáticamente en los exones de *BRCA2*, permitiendo la caracterización de 21 intervalos enriquecidos en SRE [61,181–183,185]. Actualmente, estamos priorizando exones con peculiaridades, como los exones sometidos a eventos de AS (p. ej., exón 3 de *RAD51D*), ya que su reconocimiento puede depender de la regulación ESE/ESS [186]. Notablemente, en los estudios citados se lograron identificar 71 variantes espliceogénicas ESE/ESS, 12 de ellas con impactos muy fuertes (expresan <10% de transcrito *full-length*), confirmando la idoneidad de este abordaje. Además, esta estrategia fue también empleada previamente para el estudio de la regulación de uno de los exones humanos más grandes, el exón 11 de *BRCA1* [246].

Mediante la incorporación de μ del en el minigén mgChk2_ex6-10, localizamos tres regiones ricas en ESE en los exones de interés del gen *CHEK2*: c.877_891 (exón 8), c.1073_1083 y c.1083_1092 (exón 10), las cuales constituyen puntos calientes para variantes SRE espliceogénicas (**Figura 16; Tabla 2**). También deben tenerse en cuenta otras dos μ del, c.849_863del y c.1011_1040del, puesto que tuvieron un impacto moderado en la inclusión de los exones 8 y 10, respectivamente. Los intervalos

espliceogénicos c.849_863, c.877_891, c.1073_1083 y c.1083_1092 son identificados por la herramienta *in silico* HEXplorer como regiones potenciadoras del *splicing*, lo que parece indicar que las predicciones proporcionadas por este algoritmo bioinformático son útiles para obtener una estimación inicial de los intervalos ricos en potenciadores o silenciadores (picos y valles, respectivamente), cuya existencia siempre debe ser corroborada funcionalmente. Un dato curioso es que las microdeleciones de 30 nt introducidas en el exón 8 no afectaron al reconocimiento de dicho exón porque, según HEXplorer, eliminaban simultáneamente intervalos ricos en ESE y ESS, contrarrestándose mutuamente. De hecho, las regiones ricas en ESE solo se revelaron tras el análisis de microdeleciones más pequeñas, de 15 nt.

El mapeo funcional del exón 4 de *BARD1* reveló tres intervalos ricos en ESE implicados en su reconocimiento (c.1232_1261, c.1257_1288 y c.1282_1311) (**Figura 21; Tabla 6**). La μ del c.1282_1311del tuvo el mayor impacto, dejando solo un 25% de transcrito mgFL, en cambio, en este caso HEXplorer predijo desacertadamente una región presuntamente rica en silenciadores, poniendo de nuevo de relieve la importancia de los ensayos funcionales (**Figura 25**). Los tres intervalos ricos en ESE podrían ser *hotspots* de variantes SRE espliceogénicas y clínicamente accionables. Es más, en ellos se acumulan 107 variantes ClinVar que podrían estar afectando al *splicing* de este gen (<https://www.ncbi.nlm.nih.gov/clinvar/?term=chr2%3A215%2C645%2C287-215%2C645%2C366>; último acceso 18 de septiembre de 2023). Cabe destacar que las microdeleciones internas de 12-nt de las 3 regiones ESE citadas no tuvieron un efecto patente en el reconocimiento del exón 4 (**Figura Suplementaria 10**), hecho que podría ser explicado por una composición heterogénea *enhancers*/silenciadores, similar al exón 8 de *CHEK2*, aunque este punto debería de ser confirmado experimentalmente.

En *BRIP1*, la secuencia de su donador GC atípico coincide con la previamente detectada en otros GC funcionales naturales o creados *de novo*, como los empleados por los exones 12 de *PALB2* y 17 de *BRCA2* o el originado por la mutación c.108+2T>C de *PALB2* (CAG|GCAAGT) [59,61]. En el caso del exón 17 de *BRCA2*, experimentos con siRNAs (*small interfering RNAs*) revelaron que su reconocimiento está mediado por los factores de *splicing* SC35, SF2/ASF y Tra2 β , mientras que 9G8, Tra2 β y SC35 participan en la selección del GC atípico del gen *BTK* [61,157]. Desafortunadamente, la búsqueda de los SRE implicados mediante μ del no reveló ningún intervalo crítico (**Figura 24**), sugiriendo

que este donador GC es reconocido de forma constitutiva por la maquinaria de *splicing* [221].



Figura 25. Perfiles de HEXplorer de las regiones A) 5' y B) 3' del exón 4 de *BARD1*. Se emplean recuadros de distintos colores para indicar la secuencia eliminada por las μ del de 30 nt ensayadas y su nivel de impacto sobre el proceso de *splicing* (en color rojo y naranja se marcan las μ del que exhiben un efecto alto y moderado, respectivamente, y en gris las que no muestran ningún impacto).

Respecto a la identificación de variantes SRE espliceogénicas, el estudio funcional exhaustivo de las 87 variantes ubicadas en los 3 intervalos clave de *CHEK2* (c.877_891, c.1073_1083 y c.1083_1092) desveló que una gran proporción de estas variantes (38/87, 44%) alteraban el reconocimiento de los exones 8 o 10 (**Figuras 17-18; Tabla 3**), lo que confirma la idoneidad de la estrategia empleada: i) selección de exones sujetos a *splicing* alternativo, y ii) análisis de microdeleciones solapantes para la detección de intervalos ricos en SRE, que constituyen *hotspots* de variantes espliceogénicas. De hecho, ya se contabilizan un total de 109 variantes ESE/ESS con impacto en el *splicing* reportadas por el grupo de Splicing y Susceptibilidad Genética a Cáncer del IBGM, dato que es importante remarcar al tratarse de un grupo de variantes especialmente difíciles de identificar debido a la imprecisión de los algoritmos de predicción. Por otro lado, estos resultados también ponen de manifiesto que, independientemente del impacto previsto de las variantes exónicas sobre la proteína, la alteración del *splicing* también debe considerarse un mecanismo deletéreo prevalente que debe ser investigado. De las 38 variantes espliceogénicas, 10 mostraron impactos fuertes o muy fuertes y 4 de ellas, c.883G>A,

c.883G>T, c.884A>T y c.1080G>T, produjeron cantidades insignificantes del transcrito mgFL (<5%) (**Figuras 17-18; Tabla 3**). Curiosamente, estas 10 variantes, a excepción de una, destruyen un posible motivo de unión del factor de *splicing* Tra2 β AGAAGA (**Tabla 9**) [247]. Tra2 β se expresa ubicuamente y participa en el control del *splicing*, por ejemplo, del exón 11 de *BRCA1* [246] o del gen *CHEK1* esencial en la respuesta a daño en el ADN, y está implicado en enfermedades como el cáncer, la atrofia muscular espinal, Alzheimer o demencia frontotemporal [248]. No obstante, según el programa SpliceAid, SC35 (*SRSF2*) y SRp30c (*SRSF9*) y, en menor medida, SRp40 (*SRSF5*) serían las proteínas SR claves implicadas en la regulación de los exones 8 y 10, respectivamente (**Figura Suplementaria 4**) [249]. En cambio, ESEfinder (<https://esefinder.ahc.umn.edu/cgi-bin/tools/ESE3/esefinder.cgi>) predice que el principal factor implicado en la regulación de ambos exones es SF2/ASF (*SRSF1*) [250]. En cualquier caso, la participación de Tra2 β o de cualquier otro factor de *splicing* en la regulación de los exones 8 y 10 de *CHEK2* debe ser estudiada mediante ensayos de unión proteína-ARN.

Tabla 9. Secuencias WT y mutantes de las 10 variantes SRE de *CHEK2* que exhiben impactos muy fuertes, y motivo putativo de interacción del factor de *splicing* Tra2 β .

Variante	Secuencia mutada ¹	Secuencia <i>wild type</i> ²
c.883G>A	GATGCAAAAGATT	GATGCAGAAGATT
c.883G>T	GATGCATAAGATT	GATGCAGAAGATT
c.884A>T	ATGCAGTAGATTA	ATGCAGAAGATTA
c.886G>A	GCAGAAAATTATT	GCAGAAGATTATT
c.886G>A	GCAGAAATTATT	GCAGAAGATTATT
c.1075G>A	TCTCAA ¹ AAAGAGG	TCTCAAGAAGAGG
c.1075G>T	TCTCAATAAGAGG	TCTCAAGAAGAGG
c.1076A>T	CTCAAGTAGAGGA	CTCAAGAAGAGGA
c.1078G>T	CAAGAA ¹ TAGGACT	CAAGAAGAGGACT
c.1080G>T ³	AGAAGATGACTGT	AGAAGAGGACTGT

¹La base mutada aparece en rojo. ²El motivo putativo de unión del factor de *splicing* Tra2 β está subrayado. ³La única mutación que no interrumpe un motivo putativo de unión de Tra2 β .

Los datos recopilados en este estudio se aprovecharon también para evaluar los programas bioinformáticos diseñados para la detección de variantes SRE, como HEXplorer, HExoSplice y HOT-SKIP, y SpliceAI. En concreto, HEXplorer mostró la mayor sensibilidad (92,1%), HExoSplice exhibió la precisión más elevada (72,1%) y SpliceAI reveló el valor más alto de especificidad (85,7%). Según nuestros datos, HOT-SKIP no debería utilizarse debido a su menor sensibilidad (21%) y precisión (49%), puesto que solo detectó nueve de las 38 variantes espliceogénicas (**Tabla 4;**

Tabla Suplementaria 6). Cabe señalar que la precisión aquí calculada de las herramientas HEXplorer y HOT-SKIP coincide con nuestros hallazgos anteriores en el exón 3 de *RAD51D* [186]. Adicionalmente, las estimaciones de sensibilidad, especificidad y precisión de HExoSplice y HEXplorer están avaladas por resultados previos mostrados por otros autores [251,252]. Por ende, la combinación de HExoSplice y HEXplorer parece ser una buena opción para seleccionar variantes SRE candidatas. Sin embargo, ambos programas no detectaron cinco y tres variantes espliceogénicas, respectivamente, y, de hecho, la variante c.1080G>T, que produce una alteración casi completa del *splicing*, no sería seleccionable por ninguna de estas herramientas (**Figura 19; Tabla 3; Tabla Suplementaria 4**). Por ello, de momento, los ensayos funcionales de variantes no pueden ser sustituidos por análisis bioinformáticos.

Finalmente, en cuanto a la interpretación clínica de estas variantes exónicas del gen *CHEK2*, seguimos el mismo sistema que el aplicado en las variantes BRIDGES (capítulo anterior de Discusión). Si bien, en este caso, el proceso de clasificación fue bastante más sencillo, puesto que contábamos con una estrategia de clasificación tentativa ya establecida y el número de transcritos generados por las variantes ensayadas era bastante inferior. Específicamente, la electroforesis capilar de fragmentos fluorescentes detectó 16 transcritos diferentes (además de los distintos transcritos mgFL), 12 de los cuales eran isoformas menores no caracterizadas ($\leq 4\%$ de la expresión global). Las isoformas de ARN descritas, además del mgFL, fueron $\Delta(E8)$ (r.847_908del; p.Pro283Aspfs*8) y $\Delta(E8)\Delta(E10)$ (r.[847_908del; 1009_1095del]; p.Pro283Aspfs*8), transcritos PTC-NMD que adquieren un grado muy fuerte de patogenicidad (PVS1_(ARN), +8 puntos en el esquema de clasificación), y $\Delta(E10)$ (r.1009_1095del; p.Tyr337_Lys365del), una delección *in-frame* que se consideró una evidencia fuerte en favor de la patogenicidad (PVS1_S(ARN), +4 puntos). Basándonos en esta información, así como en otros criterios, tres variantes espliceogénicas se clasificaron como P/LP (c.883G>A, c.883G>T y c.884A>T), mientras que el resto quedaron catalogadas como VUS. El hecho de haber clasificado algunas de ellas como P/LP es algo bastante excepcional, pues normalmente las variantes SRE suelen mostrar efectos parciales. Dos de las variantes P/LP estaban registradas en la base de datos ClinVar, una categorizada como VUS, c.883G>A (*missense*, p.(Glu295Lys)), y la otra como patogénica, c.883G>T (*nonsense*, p.(Glu295*)), por ser una variante truncante de proteína, las cuales, junto con la variante c.884A>T (*missense*, p.(Glu295Val)), fueron reclasificadas como variantes espliceogénicas (**Tabla 5; Tabla Suplementaria Excel 1**). En referencia a las variantes no espliceogénicas, 3 fueron

catalogadas como P por ser variantes *nonsense* y *frameshift* (c.891T>G y c.891T>A, p.(Tyr297*); c.1089_1092del, p.(Ile364Argfs*17)) y 15 como B/LB (Tabla Suplementaria Excel 2). Aparentemente, la incorporación de los datos funcionales de los minigenes mgChk2_ex6-10 mutantes tiene un impacto menor del esperado en la clasificación clínica, ya que la mayoría de las variantes analizadas seguían siendo VUS. Existen dos factores que explican la alta tasa de VUS: (i) la mayoría de las variantes espliceogénicas testadas eran variantes con impacto parcial, produciendo una proporción significativa de transcritos canónicos (portadores de la variante investigada), y (ii) muchas variantes no espliceogénicas eran *missense*.

En conclusión, hemos ensayado un total de 38 microdeleciones en nuestros exones diana mediante los minigenes mgChk2_ex6-10, mgBARD1_ex3-5 y mgBRIP1_ex1-2, identificando 6 regiones enriquecidas en ESE, tres en *CHEK2* (c.877_891, c.1073_1083 y c.1083_1092) y tres en *BARD1* (c.1232_1261, c.1257_1288 y c.1282_1311). Además, en los intervalos ESE críticos de *CHEK2*, hemos testado 87 variantes exónicas, de las cuales 38 alteraron el *splicing*, 4 mostrando impactos muy fuertes. Finalmente, tres variantes espliceogénicas (c.883G>A, c.883G>T y c.884A>T) fueron clasificadas como P/LP. Este estudio proporciona algunos ejemplos que ilustran la relevancia clínica de realizar ensayos de *splicing* de variantes exónicas, independientemente de su anotación HGVS como sinónimas, *missense*, *in-frame*, *nonsense* o *frameshift*. Finalmente, incluyendo también las variantes BRIDGES testadas en este gen, se puede afirmar que en esta tesis se ha realizado el mayor estudio funcional de *splicing* del gen *CHEK2* (139 variantes) reportado hasta la fecha.

CONCLUSIONES

CONCLUSIONES

1. Se construyeron un total de 6 minigenes de los genes de susceptibilidad *RAD51C*, *CHEK2*, *BARD1* y *BRIP1*, los cuales permiten el estudio de cualquier variante candidata ubicada en los 27 exones clonados en los mismos.

Variantes de sitios de *splicing*

2. Una gran proporción de las variantes investigadas de los genes *RAD51C* y *CHEK2* alteraban el *splicing* (85/92; 92,4%), de las cuales 69 provocaban anomalías severas. Por tanto, el *splicing* aberrante de dichos genes es un mecanismo deletéreo prevalente asociado a variantes germinales en los mismos.
3. Las herramientas bioinformáticas, como MaxEntScan y SpliceAI, mostraron una alta precisión y son muy útiles para priorizar variantes potencialmente espliceogénicas.
4. Cualquier cambio de nucleótido en cualquier posición de la secuencia consenso de los sitios aceptor y donador es potencialmente deletéreo. De hecho, de las 85 variantes espliceogénicas, solo 42 están ubicadas en las posiciones intrónicas fuertemente conservadas $\pm 1,2$ y 7 variantes *missense* deben ser reclasificadas como espliceogénicas.
5. La técnica de análisis de fragmentos fluorescentes presenta una alta sensibilidad y resolución, permitiendo la detección de 142 transcritos diferentes. Entre otros, esta técnica posibilitó la identificación de un transcrito inducido por la variante c.684-2A>G de *CHEK2* que implica el uso de un sitio aceptor TG atípico altamente conservado (17/23 posiciones), cuyo reconocimiento depende de secuencias críticas ubicadas en el exón 6.
6. De acuerdo con las normas ACMG/AMP, 58 variantes fueron clasificadas como patogénicas/probablemente patogénicas. Notablemente, ocho variantes $\pm 1,2$ fueron reclasificadas como VUS debido a la generación de transcritos de dudoso impacto en la función génica. Por ello, el testado funcional de cualquier variante candidata es imprescindible para efectuar su interpretación clínica.

Elementos reguladores del *splicing* y variantes SRE

7. Tanto el reconocimiento de los exones alternativos 8 y 10 del gen *CHEK2* como la correcta inclusión del exón 4 de *BARD1* dependen de tres intervalos ricos en ESE. De las 87 variantes de *CHEK2* testadas, localizadas en dichos intervalos, 38 (43,7%) son espliceogénicas y 10 de ellas exhiben impactos fuertes o muy fuertes. Siguiendo las normas ACMG/AMP, tres de estas variantes fueron clasificadas como patogénicas/probablemente patogénicas. En conjunto, esto confirma la utilidad de la estrategia del mapeo funcional de SRE mediante microdeleciones solapantes.
8. Nueve de las 10 variantes exónicas con impactos fuertes o muy fuertes de los exones 8 y 10 de *CHEK2* eliminan un motivo putativo de unión del factor de *splicing* Tra2 β , que podría ser esencial para el reconocimiento de dichos exones.
9. Cualquier tipo de mutación es capaz de alterar el *splicing* de un gen. De hecho, 3 variantes *nonsense*, 33 *missense* y 2 sinónimas desregulan el *splicing* del gen *CHEK2*.

Conclusiones finales

- La alteración del *splicing* inducida por variantes germinales en genes de susceptibilidad es un mecanismo etiopatogénico prevalente en cáncer de mama hereditario.
- Los minigenes reporteros de *splicing* son una herramienta idónea para los estudios de regulación y la determinación del impacto que ejercen las variantes candidatas sobre el proceso de *splicing*.

BIBLIOGRAFÍA

BIBLIOGRAFÍA

- 1 Berk AJ. Discovery of RNA splicing and genes in pieces. *Proc Natl Acad Sci U S A* 2016; **113**: 801-805
- 2 Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* 1977; **74**: 3171-3175
- 3 Chow LT, Gelinas RE, Broker TR, *et al.* An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 1977; **12**: 1-8
- 4 Gilbert W. Why genes in pieces? *Nature* 1978; **271**: 501
- 5 Klessig DF. Two adenovirus mRNAs have a common 5' terminal leader sequence encoded at least 10 kb upstream from their main coding regions. *Cell* 1977; **12**: 9-21
- 6 Chow LT, Broker TR. The spliced structures of adenovirus 2 fiber message and the other late mRNAs. *Cell* 1978; **15**: 497-510
- 7 Lenasi T, Barboric M. Mutual relationships between transcription and pre-mRNA processing in the synthesis of mRNA. *Wiley Interdiscip Rev RNA* 2013; **4**: 139-154
- 8 Zhang P, Philippot Q, Ren W, *et al.* Genome-wide detection of human variants that disrupt intronic branchpoints. *Proc Natl Acad Sci U S A* 2022; **119**: e2211194119
- 9 Ule J, Blencowe BJ. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell* 2019; **76**: 329-345
- 10 Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* 2017; **18**: 437-451
- 11 Chao Y, Jiang Y, Zhong M, *et al.* Regulatory roles and mechanisms of alternative RNA splicing in adipogenesis and human metabolic health. *Cell Biosci* 2021; **11**: 66
- 12 Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. *Nat Rev Genet* 2016; **17**: 407-421
- 13 Pitolli C, Marini A, Sette C, *et al.* Non-Canonical Splicing and Its Implications in Brain Physiology and Cancer. *Int J Mol Sci* 2022; **23**: 2811

- 14 Will CL, Luhrmann R. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* 2011; **3**: a003707
- 15 Alioto TS. U12DB: a database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res* 2007; **35**: D110-D115
- 16 Akinyi M V., Frilander MJ. At the Intersection of Major and Minor Spliceosomes: Crosstalk Mechanisms and Their Impact on Gene Expression. *Front Genet* 2021; **12**: 700744
- 17 Rogalska ME, Vivori C, Valcárcel J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat Rev Genet* 2022; **24**: 251-269
- 18 Ding Z, Meng Y, Fan Y, *et al.* Roles of minor spliceosome in intron recognition and the convergence with the better understood major spliceosome. *Wiley Interdiscip Rev RNA* 2023; **14**: e1761
- 19 Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* 2014; **15**: 108-121
- 20 Turunen JJ, Niemelä EH, Verma B, *et al.* The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA* 2013; **4**: 61-76
- 21 Robberson BL, Cote GJ, Berget SM. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* 1990; **10**: 84-94
- 22 Li X, Liu S, Zhang L, *et al.* A unified mechanism for intron and exon definition and back-splicing. *Nature* 2019; **573**: 375-380
- 23 Roos D, de Boer M. Mutations in cis that affect mRNA synthesis, processing and translation. *Biochim Biophys Acta Mol Basis Dis* 2021; **1867**: 166166
- 24 Ward AJ, Cooper TA. The pathobiology of splicing. *J Pathol* 2010; **220**: 152-163
- 25 Parada GE, Munita R, Cerda CA, *et al.* A comprehensive survey of non-canonical splice sites in the human transcriptome. *Nucleic Acids Res* 2014; **42**: 10564-10578
- 26 Grodecká L, Buratti E, Freiburger T. Mutations of Pre-mRNA Splicing Regulatory Elements: Are Predictions Moving Forward to Clinical Diagnostics? *Int J Mol Sci* 2017; **18**: 1668

- 27 Shenasa H, Hertel KJ. Combinatorial regulation of alternative splicing. *Biochim Biophys Acta Gene Regul Mech* 2019; **1862**: 194392
- 28 Wang E, Aifantis I. RNA Splicing and Cancer. *Trends Cancer* 2020; **6**: 631-644
- 29 Fu XD, Ares M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* 2014; **15**: 689-701
- 30 Pagani F, Stuani C, Tzetis M, *et al.* New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum Mol Genet* 2003; **12**: 1111-1120
- 31 Dvinge H. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett* 2018; **592**: 2987-3006
- 32 Jeong S. SR Proteins: Binders, Regulators, and Connectors of RNA. *Mol Cells* 2017; **40**: 1-9
- 33 Slišković I, Eich H, Muller-McNicoll M. Exploring the multifunctionality of SR proteins. *Biochem Soc Trans* 2022; **50**: 187-198
- 34 Ge H, Manley JL. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* 1990; **62**: 25-34
- 35 Krainer AR, Conway GC, Kozak D. Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev* 1990; **4**: 1158-1171
- 36 Howard JM, Sanford JR. The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip Rev RNA* 2015; **6**: 93-110
- 37 Cowper AE, Cáceres JF, Mayeda A, *et al.* Serine-Arginine (SR) Protein-like Factors That Antagonize Authentic SR Proteins and Regulate Alternative Splicing. *Journal of Biological Chemistry* 2001; **276**: 48908-48914
- 38 Kumar K, Sinha SK, Maity U, *et al.* Insights into established and emerging roles of SR protein family in plants and animals. *Wiley Interdiscip Rev RNA* 2023; **14**: e1763
- 39 Bampton A, Gittings LM, Fratta P, *et al.* The role of hnRNPs in frontotemporal dementia and amyotrophic lateral sclerosis. *Acta Neuropathol* 2020; **140**: 599-623

- 40 Bhattarai K, Holcik M. Diverse roles of heterogeneous nuclear ribonucleoproteins in viral life cycle. *Front Virol* 2022; **2**: 1044652
- 41 Geuens T, Bouhy D, Timmerman V. The hnRNP family: insights into their role in health and disease. *Hum Genet* 2016; **135**: 851-867
- 42 Martinez-Contreras R, Cloutier P, Shkreta L, *et al.* hnRNP proteins and splicing control. *Adv Exp Med Biol* 2007; **623**: 123-147
- 43 Wang J, Sun D, Wang M, *et al.* Multiple functions of heterogeneous nuclear ribonucleoproteins in the positive single-stranded RNA virus life cycle. *Front Immunol* 2022; **13**: 989298
- 44 Wahl MC, Will CL, Lührmann R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* 2009; **136**: 701-718
- 45 Graveley BR. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* 2001; **17**: 100-107
- 46 Wright CJ, Smith CWJ, Jiggins CD. Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet* 2022; **23**: 697-710
- 47 Pan Q, Shai O, Lee LJ, *et al.* Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 2008; **40**: 1413-1415
- 48 Marasco LE, Kornblihtt AR. The physiology of alternative splicing. *Nat Rev Mol Cell Biol* 2023; **24**: 242-254
- 49 Lindeboom RGH, Supek F, Lehner B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet* 2016; **48**: 1112-1118
- 50 Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat Rev Genet* 2002; **3**: 285-298
- 51 Pagani F, Raponi M, Baralle FE. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc Natl Acad Sci U S A* 2005; **102**: 6368-6372

- 52 Soukariéh O, Gaildrat P, Hamieh M, *et al.* Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet* 2016; **12**: e1005756
- 53 Soemedi R, Cygan KJ, Rhine CL, *et al.* Pathogenic variants that alter protein code often disrupt splicing. *Nat Genet* 2017; **49**: 848-855
- 54 López-Bigas N, Audit B, Ouzounis C, *et al.* Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett* 2005; **579**: 1900-1903
- 55 Singh P, Ahi EP. The importance of alternative splicing in adaptive evolution. *Mol Ecol* 2022; **31**: 1928-1938
- 56 Abramowicz A, Gos M. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet* 2018; **59**: 253-268
- 57 Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet* 2016; **17**: 19-32
- 58 Lord J, Baralle D. Splicing in the Diagnosis of Rare Disease: Advances and Challenges. *Front Genet* 2021; **12**: 689892
- 59 Valenzuela-Palomo A, Bueno-Martínez E, Sanoguera-Miralles L, *et al.* Splicing predictions, minigene analyses, and ACMG-AMP clinical classification of 42 germline PALB2 splice-site variants. *J Pathol* 2022; **256**: 321-334
- 60 Nix P, Mundt E, Coffee B, *et al.* Interpretation of BRCA2 Splicing Variants: A Case Series of Challenging Variant Interpretations and the Importance of Functional RNA Analysis. *Fam Cancer* 2022; **21**: 7-19
- 61 Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V, *et al.* Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. *PLoS Genet* 2017; **13**: e1006691
- 62 Aten E, Sun Y, Almomani R, *et al.* Exome sequencing identifies a branch point variant in Aarskog-Scott syndrome. *Hum Mutat* 2013; **34**: 430-434
- 63 Cerasuolo A, Cammarota F, Duraturo F, *et al.* Implications of Splicing Alterations in the Onset and Phenotypic Variability of a Family with Subclinical Manifestation of

Peutz–Jeghers Syndrome: Bioinformatic and Molecular Evidence. *Int J Mol Sci* 2020; **21**: 8201

- 64 Palhais B, Dembic M, Sabaratnam R, *et al.* The prevalent deep intronic c. 639+919 G>A GLA mutation causes pseudoexon activation and Fabry disease by abolishing the binding of hnRNPA1 and hnRNP A2/B1 to a splicing silencer. *Mol Genet Metab* 2016; **119**: 258-269
- 65 Griffin C, Saint-Jeannet J. Spliceosomopathies: Diseases and mechanisms. *Dev Dyn* 2020; **249**: 1038-1046
- 66 Taylor J, Lee SC. Mutations in spliceosome genes and therapeutic opportunities in myeloid malignancies. *Genes Chromosomes Cancer* 2019; **58**: 889-902
- 67 Anczuków O, Krainer AR. Splicing-factor alterations in cancers. *RNA* 2016; **22**: 1285-1301
- 68 Dvinge H, Kim E, Abdel-Wahab O, *et al.* RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* 2016; **16**: 413-430
- 69 Moles-Fernández A, Duran-Lozano L, Montalban G, *et al.* Computational tools for splicing defect prediction in breast/ovarian cancer genes: How efficient are they at predicting RNA alterations? *Front Genet* 2018; **9**: 366
- 70 van der Klift HM, Jansen AML, van der Steenstraten N, *et al.* Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol Genet Genomic Med* 2015; **3**: 327-345
- 71 Tournier I, Vezain M, Martins A, *et al.* A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat* 2008; **29**: 1412-1424
- 72 Sanz DJ, Acedo A, Infante M, *et al.* A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin Cancer Res* 2010; **16**: 1957-1967
- 73 Cooper TA. Use of minigene systems to dissect alternative splicing elements. *Methods* 2005; **37**: 331-340

- 74 Cruz-Benítez L, Morales-Hernández E. Historia y estado actual sobre los tipos de procedimientos quirúrgicos realizados en cáncer de mama. *Gac Mex de Oncol* 2014; **13**: 124-133
- 75 Kapur BN. The fruits of long endeavors – 200 years of oncology. *Med J Armed Forces India* 2014; **70**: 95-97
- 76 Hajdu SI. Pathfinders in oncology from ancient times to the end of the Middle Ages. *Cancer* 2016; **122**: 1638-1646
- 77 Lukong KE. Understanding breast cancer – The long and winding road. *BBA Clin* 2017; **7**: 64-77
- 78 Ekmektzoglou KA, Xanthos T, German V, *et al*. Breast cancer: From the earliest times through to the end of the 20th century. *Eur J Obstet Gynecol Reprod Biol* 2009; **145**: 3-8
- 79 Lakhtakia R. A Brief History of Breast Cancer Part I: Surgical domination reinvented. *Sultan Qaboos Univ Med J* 2014; **14**: e166-e169
- 80 Barrón-Gallardo C, Jave-Suarez L, Aguilar-Lemarroy A. Historia del cáncer de mama [History of breast cancer]. *Rev Med Inst Mex Seguro Soc* 2020; **58**: S75-S82
- 81 Bianucci R, Perciaccante A, Charlier P, *et al*. Earliest evidence of malignant breast cancer in Renaissance paintings. *Lancet Oncol* 2018; **19**: 166-167
- 82 Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* 2022; **12**: 31-46
- 83 Łukasiewicz S, Czezelewski M, Forma A, *et al*. Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—An updated review. *Cancers (Basel)* 2021; **13**: 4287
- 84 Wojtyła C, Bertuccio P, Wojtyła A, *et al*. European trends in breast cancer mortality, 1980–2017 and predictions to 2025. *Eur J Cancer* 2021; **152**: 4-17
- 85 Rakha EA, Tse GM, Quinn CM. An update on the pathological classification of breast cancer. *Histopathology* 2023; **82**: 5-16
- 86 Tawfik O, Alallaf J. Current Approaches to Diagnosis and Treatment of Breast Cancer and Future Directions. *Women Health Care Issues* 2019; **2**

- 87 van Dooijeweert C, van Diest PJ, Ellis IO. Grading of invasive breast carcinoma: the way forward. *Virchows Archiv* 2022; **480**: 33-43
- 88 Koh J, Kim MJ. Introduction of a new staging system of breast cancer for radiologists: An emphasis on the prognostic stage. *Korean J Radiol* 2019; **20**: 69-82
- 89 Johnson KS, Conant EF, Soo MS. Molecular Subtypes of Breast Cancer: A Review for Breast Radiologists. *J Breast Imaging* 2021; **3**: 12-24
- 90 Perou CM, Sørile T, Eisen MB, *et al.* Molecular portraits of human breast tumours. *Nature* 2000; **406**: 747-752
- 91 Dalivandan ST, Plummer J, Gayther SA. Risks and Function of Breast Cancer Susceptibility Alleles. *Cancers (Basel)* 2021; **13**: 3953
- 92 Lynch HT, Silva E, Snyder C, *et al.* Hereditary breast cancer: part I. Diagnosing hereditary breast cancer syndromes. *Breast J* 2008; **14**: 3-13
- 93 Broca P. *Traité Des Tumeurs*. Paris: P. Asselin; 1866.
- 94 Krush A. Contributions of Pierre Paul Broca to Cancer Genetics. *Trans Nebr Acad Sci Affil Soc* 1979; **7**
- 95 Smithers DW. Family Histories of 459 Patients with Cancer of the Breast. *Br J Cancer* 1948; **2**: 163-167
- 96 Penrose LS, Mackenzie HJ, Karn MN. A genetical study of human mammary cancer. *Br J Cancer* 1948; **2**: 168-176
- 97 Martynova RP. Studies in the Genetics of Human Neoplasms: Cancer of the Breast, Based upon 201 Family Histories. *Am J Cancer* 1937; **29**: 530-540
- 98 Stephens F, Gardner E, Woolf C. A recheck of kindred 107, which has shown a high frequency of breast cancer. *Cancer* 1958; **11**: 967-972
- 99 Lynch HT, Krush AJ, Lemon HM, *et al.* Tumor Variation in Families With Breast Cancer. *JAMA* 1972; **222**: 1631-1635
- 100 Lynch HT, Guirgis HA, Albert S, *et al.* Familial association of carcinoma of the breast and ovary. *Surg Gynecol Obstet* 1974; **138**: 717-724

- 101 Newman B, Austin MA, Lee M, *et al.* Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proc Natl Acad Sci U S A* 1988; **85**: 3044-3048
- 102 Hall JM, Lee MK, Newman B, *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science (1979)* 1990; **250**: 1684-1689
- 103 Miki Y, Swensen J, Shattuck-Eidens D, *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science (1979)* 1994; **266**: 66-71
- 104 Wooster R, Bignell G, Lancaster J, *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* 1995; **378**: 789-792
- 105 Armstrong K, Weiner J, Weber B, *et al.* Early adoption of BRCA1/2 testing: who and why. *Genet Med* 2003; **5**: 92-98
- 106 Yoshida R. Hereditary breast and ovarian cancer (HBOC): review of its molecular characteristics, screening, treatment, and prognosis. *Breast Cancer* 2021; **28**: 1167-1180
- 107 Sokolova A, Johnstone KJ, McCart Reed AE, *et al.* Hereditary breast cancer: syndromes, tumour pathology and molecular testing. *Histopathology* 2023; **82**: 70-82
- 108 Dorling L, Carvalho S, Allen J, *et al.* Breast Cancer Risk Genes — Association Analysis in More than 113,000 Women. *N Engl J Med* 2021; **384**: 428-439
- 109 Hu C, Hart SN, Gnanaolivu R, *et al.* A Population-Based Study of Genes Previously Implicated in Breast Cancer. *N Engl J Med* 2021; **384**: 440-451
- 110 Graffeo R, Rana HQ, Conforti F, *et al.* Moderate penetrance genes complicate genetic testing for breast cancer diagnosis: ATM, CHEK2, BARD1 and RAD51D. *Breast* 2022; **65**: 32-40
- 111 Li S, Silvestri V, Leslie G, *et al.* Cancer Risks Associated With BRCA1 and BRCA2 Pathogenic Variants. *J Clin Oncol* 2022; **40**: 1529-1541
- 112 Yang X, Leslie G, Doroszuk A, *et al.* Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. *J Clin Oncol* 2020; **38**: 674-685

- 113 Stolarova L, Kleiblova P, Janatova M, *et al.* CHEK2 Germline Variants in Cancer Predisposition: Stalemate Rather than Checkmate. *Cells* 2020; **9**: 2675
- 114 Hall MJ, Bernhisel R, Hughes E, *et al.* Germline Pathogenic Variants in the Ataxia Telangiectasia Mutated (ATM) Gene are Associated with High and Moderate Risks for Multiple Cancers. *Cancer Prev Res (Phila)* 2021; **14**: 433-440
- 115 Yang X, Song H, Leslie G, *et al.* Ovarian and Breast Cancer Risks Associated With Pathogenic Variants in RAD51C and RAD51D. *Journal of the National Cancer Institute* 2020; **112**: 1242-1250
- 116 Hawsawi YM, Shams A, Theyab A, *et al.* BARD1 mystery: tumor suppressors are cancer susceptibility genes. *BMC Cancer* 2022; **22**: 599
- 117 Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 1971; **68**: 820-823
- 118 Reid S, Spalluto LB, Lang K, *et al.* An overview of genetic services delivery for hereditary breast cancer. *Breast Cancer Res Treat* 2022; **191**: 491-500
- 119 González-Santiago S, Ramón y Cajal T, Aguirre E, *et al.* SEOM clinical guidelines in hereditary breast and ovarian cancer (2019). *Clin Transl Oncol* 2020; **22**: 193-200
- 120 Hipp LE, Hulswit BB, Milliron KJ. Clinical tools and counseling considerations for breast cancer risk assessment and evaluation for hereditary cancer risk. *Best Pract Res Clin Obstet Gynaecol* 2022; **82**: 12-29
- 121 Castéra L, Krieger S, Rousselin A, *et al.* Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *Eur J Hum Genet* 2014; **22**: 1305-1313
- 122 Yadav S, Couch FJ. Germline Genetic Testing for Breast Cancer Risk: The Past, Present, and Future. *Am Soc Clin Oncol Educ Book* 2019; **39**: 61-74
- 123 Reid S, Pal T. Update on multi-gene panel testing and communication of genetic test results. *Breast J* 2020; **26**: 1513-1519
- 124 Richards S, Aziz N, Bale S, *et al.* Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of

- Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; **17**: 405-424
- 125 Smirnov D, Schlieben LD, Peymani F, *et al.* Guidelines for clinical interpretation of variant pathogenicity using RNA phenotypes. *Hum Mutat* 2022; **43**: 1056-1070
- 126 Pashayan N, Antoniou AC, Ivanus U, *et al.* Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat Rev Clin Oncol* 2020; **17**: 687-705
- 127 Desai N V., Yadav S, Batalini F, *et al.* Germline genetic testing in breast cancer: Rationale for the testing of all women diagnosed by the age of 60 years and for risk-based testing of those older than 60 years. *Cancer* 2021; **127**: 828-833
- 128 Makhnoon S, Shirts BH, Bowen DJ. Patients' perspectives of variants of uncertain significance and strategies for uncertainty management. *J Genet Couns* 2019; **28**: 313-325
- 129 Meindl A, Hellebrand H, Wiek C, *et al.* Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet* 2010; **42**: 410-414
- 130 Suszynska M, Ratajska M, Kozlowski P. BRIP1, RAD51C, and RAD51D mutations are associated with high susceptibility to ovarian cancer: mutation prevalence and precise risk estimates based on a pooled analysis of ~30,000 cases. *J Ovarian Res* 2020; **13**: 50
- 131 Miller KA, Sawicka D, Barsky D, *et al.* Domain mapping of the Rad51 paralog protein complexes. *Nucleic Acids Res* 2004; **32**: 169-178
- 132 Masson JY, Tarsounas MC, Stasiak AZ, *et al.* Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes Dev* 2001; **15**: 3296-3307
- 133 Liu T, Wan L, Wu Y, *et al.* hSWS1·SWSAP1 Is an Evolutionarily Conserved Complex Required for Efficient Homologous Recombination Repair. *J Biol Chem* 2011; **286**: 41758-41766
- 134 Rein HL, Bernstein KA, Baldock RA. RAD51 paralog function in replicative DNA damage and tolerance. *Curr Opin Genet Dev* 2021; **71**: 86-91

- 135 Park JY, Zhang F, Andreassen PR. PALB2: The hub of a network of tumor suppressors involved in DNA damage responses. *Biochim Biophys Acta Rev Cancer* 2014; **1846**: 263-275
- 136 Michl J, Zimmer J, Tarsounas M. Interplay between Fanconi anemia and homologous recombination pathways in genome integrity. *EMBO J* 2016; **35**: 909-923
- 137 Berti M, Teloni F, Mijic S, *et al.* Sequential role of RAD51 paralog complexes in replication fork remodeling and restart. *Nat Commun* 2020; **11**: 3531
- 138 Somyajit K, Subramanya S, Nagaraju G. RAD51C: A novel cancer susceptibility gene is linked to Fanconi anemia and breast cancer. *Carcinogenesis* 2010; **31**: 2031-2038
- 139 Suwaki N, Klare K, Tarsounas M. RAD51 paralogs: Roles in DNA damage signalling, recombinational repair and tumorigenesis. *Semin Cell Dev Biol* 2011; **22**: 898-905
- 140 Vaz F, Hanenberg H, Schuster B, *et al.* Mutation of the RAD51C gene in a Fanconi anemia-like disorder. *Nat Genet* 2010; **42**: 406-409
- 141 Boni J, Idani A, Roca C, *et al.* A decade of RAD51C and RAD51D germline variants in cancer. *Hum Mutat* 2022; **43**: 285-298
- 142 Pietragalla A, Arcieri M, Marchetti C, *et al.* Ovarian cancer predisposition beyond BRCA1 and BRCA2 genes. *Int J Gynecol Cancer* 2020; **30**: 1803-1810
- 143 Bell D, Varley J, Szydlo T, *et al.* Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science* 1999; **286**: 2528-2531
- 144 Meijers-Heijboer H, Van den Ouweland A, Klijn J, *et al.* Low-penetrance susceptibility to breast cancer due to CHEK2*1100delC in noncarriers of BRCA1 or BRCA2 mutations: The CHEK2-breast cancer consortium. *Nat Genet* 2002; **31**: 55-59
- 145 Tominaga K, Morisaki H, Kaneko Y, *et al.* Role of Human Cds1 (Chk2) Kinase in DNA Damage Checkpoint and Its Regulation by p53. *J Biol Chem* 1999; **274**: 31463-31467
- 146 Zannini L, Delia D, Buscemi G. CHK2 kinase in the DNA damage response and beyond. *J Mol Cell Biol* 2014; **6**: 442-457

- 147 Angeli D, Salvi S, Tedaldi G. Genetic Predisposition to Breast and Ovarian Cancers: How Many and Which Genes to Test? *Int J Mol Sci* 2020; **21**: 1128
- 148 Lowry KP, Geuzinge HA, Stout NK, *et al.* Breast Cancer Screening Strategies for Women With ATM, CHEK2, and PALB2 Pathogenic Variants. *JAMA Oncol* 2022; **8**: 587-596
- 149 Wu LC, Wang ZW, Tsan JT, *et al.* Identification of a RING protein that can interact in vivo with the BRCA1 gene product. *Nat Genet* 1996; **14**: 430-440
- 150 Movassat M, Forouzmand E, Reese F, *et al.* Exon size and sequence conservation improves identification of splice-altering nucleotides. *RNA* 2019; **25**: 1793-1805
- 151 Watters AK, Seltzer ES, MacKenzie D, *et al.* The Effects of Genetic and Epigenetic Alterations of BARD1 on the Development of Non-Breast and Non-Gynecological Cancers. *Genes (Basel)* 2020; **11**: 829
- 152 Irminger-Finger I, Ratajska M, Pilyugin M. New concepts on BARD1: Regulator of BRCA pathways and beyond. *Int J Biochem Cell Biol* 2016; **72**: 1-17
- 153 Bunch H, Jeong J, Kang K, *et al.* BRCA1-BARD1 regulates transcription through modulating topoisomerase II β . *Open Biol* 2021; **11**: 210221
- 154 Thai T, Du F, Tsan JT, *et al.* Mutations in the BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers. *Hum Mol Genet* 1998; **7**: 195-202
- 155 Alenezi WM, Fierheller CT, Recio N, *et al.* Literature Review of BARD1 as a Cancer Predisposing Gene with a Focus on Breast and Ovarian Cancers. *Genes (Basel)* 2020; **11**: 856
- 156 Weber-Lassalle N, Hauke J, Ramser J, *et al.* BRIP1 loss-of-function mutations confer high risk for familial ovarian cancer, but not familial breast cancer. *Breast Cancer Res* 2018; **20**: 7
- 157 Kralovicova J, Hwang G, Asplund AC, *et al.* Compensatory signals associated with the activation of human GC 5' splice sites. *Nucleic Acids Res* 2011; **39**: 7077-7091

- 158 Wu Y, Suhasini AN, Brosh RM. Welcome the Family of FANCI-like Helicases to the Block of Genome Stability Maintenance Proteins. *Cell Mol Life Sci* 2009; **66**: 1209-1222
- 159 Brosh RM, Wu Y. An emerging picture of FANCI's role in G4 resolution to facilitate DNA replication. *NAR Cancer* 2021; **3**: zcab034
- 160 Calvo JA, Fritchman B, Hernandez D, *et al.* Comprehensive Mutational Analysis of the BRCA1-Associated DNA Helicase and Tumor-Suppressor FANCI/BACH1/BRIP1. *Mol Cancer Res* 2021; **19**: 1015-1025
- 161 Ouhtit A, Gupta I, Shaikh Z. BRIP1, a potential candidate gene in development of non-BRCA1/2 breast cancer. *Front Biosci (Elite Ed)* 2016; **8**: 289-298
- 162 Cantor SB, Bell DW, Ganesan S, *et al.* BACH1, a Novel Helicase-like Protein, Interacts Directly with BRCA1 and Contributes to Its DNA Repair Function. *Cell* 2001; **105**: 149-160
- 163 Levrán O, Attwooll C, Henry RT, *et al.* The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. *Nat Genet* 2005; **37**: 931-933
- 164 Litman R, Peng M, Jin Z, *et al.* BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCI. *Cancer Cell* 2005; **8**: 255-265
- 165 Levitus M, Waisfisz Q, Godthelp BC, *et al.* The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. *Nat Genet* 2005; **37**: 934-935
- 166 Seal S, Thompson D, Renwick A, *et al.* Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* 2006; **38**: 1239-1241
- 167 Ramus SJ, Song H, Dicks E, *et al.* Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer. *J Natl Cancer Inst* 2015; **107**: djv214
- 168 Castillo-Guardiola V, Sarabia-Meseguer MD, Marín-Vera M, *et al.* New insights into the performance of multigene panel testing: Two novel nonsense variants in BRIP1 and TP53 in a young woman with breast cancer. *Cancer Genet* 2018; **228-229**: 1-4

- 169 Mikaeel RR, Young JP, Li Y, *et al.* Survey of germline variants in cancer-associated genes in young adults with colorectal cancer. *Genes Chromosomes Cancer* 2022; **61**: 105-113
- 170 Nero C, Pasciuto T, Cappuccio S, *et al.* Further refining 2020 ESGO/ESTRO/ESP molecular risk classes in patients with early-stage endometrial cancer: A propensity score-matched analysis. *Cancer* 2022; **128**: 2898-2907
- 171 He C, Zhou F, Zuo Z, *et al.* A Global View of Cancer-Specific Transcript Variants by Subtractive Transcriptome-Wide Analysis. *PLoS One* 2009; **4**: e4732
- 172 Kahles A, Lehmann K-V, Toussaint NC, *et al.* Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* 2018; **34**: 211-224
- 173 Yang Q, Zhao J, Zhang W, *et al.* Aberrant alternative splicing in breast cancer. *J Mol Cell Biol* 2019; **11**: 920-929
- 174 Lodomery M. Aberrant Alternative Splicing Is Another Hallmark of Cancer. *Int J Cell Biol* 2013; **2013**: 463786
- 175 Bradley RK, Anczuków O. RNA splicing dysregulation and the hallmarks of cancer. *Nat Rev Cancer* 2023; **23**: 135-155
- 176 Han P, Zhu J, Feng G, *et al.* Characterization of alternative splicing events and prognostic signatures in breast cancer. *BMC Cancer* 2021; **21**: 587
- 177 Gahete MD, Herman-Sanchez N, Fuentes-Fayos AC, *et al.* Dysregulation of splicing variants and spliceosome components in breast cancer. *Endocr Relat Cancer* 2022; **29**: R123-R142
- 178 Martínez-Montiel N, Anaya-Ruiz M, Pérez-Santos M, *et al.* Alternative Splicing in Breast Cancer and the Potential Development of Therapeutic Tools. *Genes (Basel)* 2017; **8**: 217
- 179 Valenzuela-Palomo A, Sanoguera-Miralles L, Bueno-Martínez E, *et al.* Splicing Analysis of 16 PALB2 ClinVar Variants by Minigene Assays: Identification of Six Likely Pathogenic Variants. *Cancers (Basel)* 2022; **14**: 4541

- 180 Bueno-Martínez E, Sanoguera-Miralles L, Valenzuela-Palomo A, *et al.* Minigene-based splicing analysis and ACMG/AMP-based tentative classification of 56 ATM variants. *J Pathol* 2022; **258**: 83-101
- 181 Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, *et al.* Mis-splicing in breast cancer: identification of pathogenic BRCA2 variants by systematic minigene assays. *J Pathol* 2019; **248**: 409-420
- 182 Acedo A, Hernández-Moro C, Curiel-García Á, *et al.* Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat* 2015; **36**: 210-221
- 183 Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, *et al.* Identification of eight spliceogenic variants in BRCA2 Exon 16 by minigene assays. *Front Genet* 2018; **9**: 188
- 184 Acedo A, Sanz DJ, Durán M, *et al.* Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res* 2012; **14**: R87
- 185 Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, *et al.* Minigene Splicing Assays Identify 12 Spliceogenic Variants of BRCA2 Exons 14 and 15. *Front Genet* 2019; **10**: 503
- 186 Bueno-Martínez E, Sanoguera-Miralles L, Valenzuela-Palomo A, *et al.* RAD51D Aberrant Splicing in Breast Cancer: Identification of Splicing Regulatory Elements and Minigene-Based Evaluation of 53 DNA Variants. *Cancers (Basel)* 2021; **13**: 2845
- 187 Lopez-Perolio I, Leman R, Behar R, *et al.* Alternative splicing and ACMG-AMP-2015-based classification of PALB2 genetic variants: An ENIGMA report. *J Med Genet* 2019; **56**: 453-460
- 188 de Garibay GR, Acedo A, García-Casado Z, *et al.* Capillary electrophoresis analysis of conventional splicing assays: IARC analytical and clinical classification of 31 BRCA2 genetic variants. *Hum Mutat* 2014; **35**: 53-57
- 189 Castéra L, Harter V, Muller E, *et al.* Landscape of pathogenic variations in a panel of 34 genes and cancer risk estimation from 5131 HBOC families. *Genet Med* 2018; **20**: 1677-1686

- 190 Rhine CL, Cygan KJ, Soemedi R, *et al.* Hereditary cancer genes are highly susceptible to splicing mutations. *PLoS Genet* 2018; **14**: e1007231
- 191 Davy G, Rousselin A, Goardon N, *et al.* Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. *Eur J Hum Genet* 2017; **25**: 1147-1154
- 192 Sanoguera-Miralles L, Valenzuela-Palomo A, Bueno-Martínez E, *et al.* Systematic minigene-based splicing analysis and tentative clinical classification of 52 CHEK2 splice-site variants. *Clin Chem* 2023
- 193 Staalesen V, Falck J, Geisler S, *et al.* Alternative splicing and mutation status of CHEK2 in stage III breast cancer. *Oncogene* 2004; **23**: 8535-8544
- 194 Canson D, Glubb D, Spurdle AB. Variant effect on splicing regulatory elements, branchpoint usage, and pseudoexonization: Strategies to enhance bioinformatic prediction using hereditary cancer genes as exemplars. *Hum Mutat* 2020; **41**: 1705-1721
- 195 Tavtigian S V., Harrison SM, Boucher KM, *et al.* Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat* 2020; **41**: 1734-1737
- 196 Walker LC, Hoya M de la, Wiggins GAR, *et al.* Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: Recommendations from the ClinGen SVI Splicing Subgroup. *Am J Hum Genet* 2023; **110**: 1046-1067
- 197 Stolarova L, Kleiblova P, Zemankova P, *et al.* ENIGMA CHEK2gether Project: A Comprehensive Study Identifies Functionally Impaired CHEK2 Germline Missense Variants Associated with Increased Breast Cancer Risk. *Clin Cancer Res* 2023; **29**: 3037-3050
- 198 Desmet FO, Hamroun D, Lalande M, *et al.* Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: e67
- 199 Walker LC, Lattimore VL, Kvist A, *et al.* Comprehensive Assessment of BARD1 Messenger Ribonucleic Acid Splicing With Implications for Variant Classification. *Front Genet* 2019; **10**: 1139

- 200 Shiovitz S, Korde LA. Genetics of breast cancer: a topic in evolution. *Ann Oncol* 2015; **26**: 1291-1299
- 201 van Marcke C, Collard A, Vikkula M, *et al*. Prevalence of pathogenic variants and variants of unknown significance in patients at high risk of breast cancer: A systematic review and meta-analysis of gene-panel data. *Crit Rev Oncol Hematol* 2018; **132**: 138-144
- 202 Slavin TP, Manjarrez S, Pritchard CC, *et al*. The effects of genomic germline variant reclassification on clinical cancer care. *Oncotarget* 2019; **10**: 417-423
- 203 Truty R, Ouyang K, Rojahn S, *et al*. Spectrum of splicing variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation. *The American Journal of Human Genetics* 2021; **108**: 696-708
- 204 Sterne-Weiler T, Sanford JR. Exon identity crisis: disease-causing mutations that disrupt the splicing code. *Genome Biol* 2014; **15**: 201
- 205 Smith A, Nelson RJ. Capillary Electrophoresis of DNA. *Curr Protoc Nucleic Acid Chem* 2003; **13**
- 206 Sanoguera-Miralles L, Valenzuela-Palomo A, Bueno-Martínez E, *et al*. Comprehensive Functional Characterization and Clinical Interpretation of 20 Splice-Site Variants of the RAD51C Gene. *Cancers (Basel)* 2020; **12**: 3771
- 207 Zhu L, Miao B, Dymerska D, *et al*. Germline Variants of CYBA and TRPM4 Predispose to Familial Colorectal Cancer. *Cancers (Basel)* 2022; **14**: 670
- 208 Lara B, Martínez MT, Blanco I, *et al*. Severe alpha-1 antitrypsin deficiency in composite heterozygotes inheriting a new splicing mutation QOMadrid. *Respir Res* 2014; **15**: 125
- 209 Villate O, Ibarluzea N, Fraile-Bethencourt E, *et al*. Functional Analyses of a Novel Splice Variant in the CHD7 Gene, Found by Next Generation Sequencing, Confirm Its Pathogenicity in a Spanish Patient and Diagnose Him with CHARGE Syndrome. *Front Genet* 2018; **9**: 7
- 210 Gailite L, Valenzuela-Palomo A, Sanoguera-Miralles L, *et al*. UGT1A1 Variants c.864+5G<T and c.996+2_996+5del of a Crigler-Najjar Patient Induce Aberrant Splicing in Minigene Assays. *Front Genet* 2020; **11**: 169

- 211 Suszynska M, Ratajska M, Kozlowski P. BRIP1, RAD51C, and RAD51D mutations are associated with high susceptibility to ovarian cancer: mutation prevalence and precise risk estimates based on a pooled analysis of ~30,000 cases. *J Ovarian Res* 2020; **13**: 50
- 212 Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; **11**: 377-394
- 213 Jaganathan K, Panagiotopoulou SK, McRae JF, *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 2019; **176**: 535-548
- 214 Shirts BH, Casadei S, Jacobson AL, *et al.* Improving performance of multigene panels for genomic analysis of cancer predisposition. *Genet Med* 2016; **18**: 974-981
- 215 Walsh T, Casadei S, Lee MK, *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A* 2011; **108**: 18032-18037
- 216 Lhota F, Zemankova P, Kleiblova P, *et al.* Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. *Clin Genet* 2016; **90**: 324-333
- 217 Golmard L, Caux-Moncoutier V, Davy G, *et al.* Germline mutation in the RAD51B gene confers predisposition to breast cancer. *BMC Cancer* 2013; **13**: 484
- 218 Pelttari LM, Heikkinen T, Thompson D, *et al.* RAD51C is a susceptibility gene for ovarian cancer. *Hum Mol Genet* 2011; **20**: 3278-3288
- 219 Coulet F, Fajac A, Colas C, *et al.* Germline RAD51C mutations in ovarian cancer susceptibility. *Clin Genet* 2013; **83**: 332-336
- 220 Neidhardt G, Becker A, Hauke J, *et al.* The RAD51C exonic splice-site mutations c.404G>C and c.404G>T are associated with familial breast and ovarian cancer. *Eur J Cancer Prev* 2017; **26**: 165-169
- 221 Landrith T, Li B, Cass AA, *et al.* Splicing profile by capture RNA-seq identifies pathogenic germline variants in tumor suppressor genes. *NPJ Precis Oncol* 2020; **4**: 4

- 222 Havranek O, Kleiblova P, Hojny J, *et al.* Association of Germline CHEK2 Gene Variants with Risk and Prognosis of Non-Hodgkin Lymphoma. *PLoS One* 2015; **10**: e0140819
- 223 Dong X, Wang L, Taniguchi K, *et al.* Mutations in CHEK2 associated with prostate cancer risk. *Am J Hum Genet* 2003; **72**: 270-280
- 224 Agiannitopoulos K, Papadopoulou E, Tsaousis GN, *et al.* Characterization of the c.793-1G > A splicing variant in CHEK2 gene as pathogenic: A case report. *BMC Med Genet* 2019; **20**: 131
- 225 Ryu JS, Lee HY, Cho EH, *et al.* Exon splicing analysis of intronic variants in multigene cancer panel testing for hereditary breast/ovarian cancer. *Cancer Sci* 2020; **111**: 3912-3925
- 226 Kleiblova P, Stolarova L, Krizova K, *et al.* Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. *Int J Cancer* 2019; **145**: 1782-1797
- 227 Kraus C, Hoyer J, Vasileiou G, *et al.* Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. *Int J Cancer* 2017; **140**: 95-102
- 228 Apostolou P, Dellatola V, Papadimitriou C, *et al.* CHEK2 Pathogenic Variants in Greek Breast Cancer Patients: Evidence for Strong Associations with Estrogen Receptor Positivity, Overuse of Risk-Reducing Procedures and Population Founder Effects. *Cancers (Basel)* 2021; **13**: 2106
- 229 Lin JH, Tang XY, Boulling A, *et al.* First estimate of the scale of canonical 5' splice site GT>GC variants capable of generating wild-type transcripts. *Hum Mutat* 2019; **40**: 1856-1873
- 230 Brandão RD, Mensaert K, López-Perolio I, *et al.* Targeted RNA-seq successfully identifies normal and pathogenic splicing events in breast/ovarian cancer susceptibility and Lynch syndrome genes. *Int J Cancer* 2019; **145**: 401-414
- 231 Fackenthal JD, Yoshimatsu T, Zhang B, *et al.* Naturally occurring BRCA2 alternative mRNA splicing events in clinically relevant samples. *J Med Genet* 2016; **53**: 548-558

- 232 Szafranski K, Schindler S, Taudien S, *et al.* Violating the splicing rules: TG dinucleotides function as alternative 3' splice sites in U2-dependent introns. *Genome Biol* 2007; **8**: R154
- 233 Lee K, Krempely K, Roberts ME, *et al.* Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Hum Mutat* 2018; **39**: 1553-1568
- 234 Brnich SE, Abou Tayoun AN, Couch FJ, *et al.* Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med* 2019; **12**: 3
- 235 Abou Tayoun AN, Pesaran T, DiStefano MT, *et al.* Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* 2018; **39**: 1517-1524
- 236 Gayarre J, Martín-Gimeno P, Osorio A, *et al.* Characterisation of the novel deleterious RAD51C p.Arg312Trp variant and prioritisation criteria for functional analysis of RAD51C missense changes. *Br J Cancer* 2017; **117**: 1048-1062
- 237 Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. *Nat Biotechnol* 2004; **22**: 535-546
- 238 Dhir A, Buratti E. Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. *FEBS J* 2010; **277**: 841-855
- 239 Jiang W, Chen L. Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Comput Struct Biotechnol J* 2020; **19**: 183-195
- 240 Wang Z, Burge CB. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* 2008; **14**: 802-813
- 241 David CJ, Manley JL. The search for alternative splicing regulators: new approaches offer a path to a splicing code. *Genes Dev* 2008; **22**: 279-285
- 242 Dal Mas A, Fortugno P, Donadon I, *et al.* Exon-Specific U1s Correct SPINK5 Exon 11 Skipping Caused by a Synonymous Substitution that Affects a Bifunctional Splicing Regulatory Element. *Hum Mutat* 2015; **36**: 504-512

- 243 Kralovicova J, Hwang G, Asplund AC, *et al.* Compensatory signals associated with the activation of human GC 5' splice sites. *Nucleic Acids Res* 2011; **39**: 7077-7091
- 244 Goina E, Skoko N, Pagani F. Binding of DAZAP1 and hnRNPA1/A2 to an Exonic Splicing Silencer in a Natural BRCA1 Exon 18 Mutant. *Mol Cell Biol* 2008; **28**: 3850-3860
- 245 Tang R, Prosser D, Love D. Evaluation of Bioinformatic Programmes for the Analysis of Variants within Splice Site Consensus Regions. *Adv Bioinformatics* 2016; **2016**: 5614058
- 246 Raponi M, Smith LD, Silipo M, *et al.* BRCA1 exon 11 a model of long exon splicing regulation. *RNA Biol* 2014; **11**: 351-359
- 247 Grellscheid S, Dalgliesh C, Storbeck M, *et al.* Identification of Evolutionarily Conserved Exons as Regulated Targets for the Splicing Activator Tra2 β in Development. *PLoS Genet* 2011; **7**: e1002390
- 248 Best A, James K, Dalgliesh C, *et al.* Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons. *Nat Commun* 2014; **5**: 4760
- 249 Piva F, Giulietti M, Nocchi L, *et al.* SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinformatics* 2009; **25**: 1211-1213
- 250 Cartegni L. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003; **31**: 3568-3571
- 251 Tubeuf H, Caputo SM, Sullivan T, *et al.* Calibration of Pathogenicity Due to Variant-Induced Leaky Splicing Defects by Using *BRCA2* Exon 3 as a Model System. *Cancer Res* 2020; **80**: 3593-3605
- 252 Tubeuf H, Charbonnier C, Soukarieh O, *et al.* Large-scale comparative evaluation of user-friendly tools for predicting variant-induced alterations of splicing regulatory elements. *Hum Mutat* 2020; **41**: 1811-1829

ANEXO

ANEXO

1. TABLAS

Tabla Suplementaria 1. Primers para incorporar en los exones 8 y 10 del mgChk2_ex6-10 las 12 μ del de 10, 11, 15 y 30 nt.

Microdelección	Exón	Primers (5'→3')
c.849_878del	Ex8	CTTTTTTCTCCCCCTCTTCTTAGCCTGCAGAAGATTATTATATTGTTTTG CAAAACAATATAATAATCTTCTGCAGGCTAAGAAGAGGGGGAGAAAAAG
c.849_863del	Ex8	CTTTTTTCTCCCCCTCTTCTTAGCCTAAAACTTTTTTGATGCAGAAGAT ATCTTCTGCATCAAAAAAGTTTTTAGGCTAAGAAGAGGGGGAGAAAAAG
c.863_877del	Ex8	CTTCTTAGCCTTGCATCATCAAGATTGCAGAAGATTATTATATTGTTTTG CAAAACAATATAATAATCTTCTGCAATCTTGATGATGCAAGGCTAAGAAG
c.876_905del	Ex8	GCATCATCAAGATTA AAAACTTTTTATTGTAAGTAGTAACTTTTTAAAA TTTTAAAAAGTTTACTACTTACAATAAAAAGTTTTTAATCTTGATGATGC
c.877_891del	Ex8	GCATCATCAAGATTA AAAACTTTTTTATATTGTTTTGGAATTGTAAGTA TACTTACAATTC CAAAACAATATAAAAAAAGTTTTTAATCTTGATGATGC
c.891_905del	Ex8	AAAAC TTTTTTGATGCAGAAGATTAATTGTAAGTAGTAACTTTTTAAAA TTTTAAAAAGTTTACTACTTACAATTAATCTTCTGCATCAAAAAAGTTTTT
c.1011_1040del	Ex10	GTCATGCCTGCCTTTCTGTGTAGTACTTAAAGCCAGAGAATGTTTTACTG CAGTAAAACATTCTCTGGCTTTAAGTACTACACAGAAAGGCAGGCATGAC
c.1037_1066del	Ex10	CTTCATGAAAACGGTATTATACACCCATCTCAAGAAGAGGACTGTCTTAT ATAAGACAGTCCTCTTCTTGAGATGGGTGTATAATACCGTTTTTCATGAAG
c.1063_1092del	Ex10	TGACTTAAAGCCAGAGAATGTTTTAAAGGTAAGAAATTGGCATTCTGTAT ATACAGAATGCCAATTTCTTACCTTTAAAACATTCTCTGGCTTTAAGTCA
c.1063_1073del	Ex10	TGACTTAAAGCCAGAGAATGTTTTAAGAAGAGGACTGTCTTATAAAGGTA TACCTTTATAAGACAGTCCTCTTCTTAAAACATTCTCTGGCTTTAAGTCA
c.1073_1083del	Ex10	CCAGAGAATGTTTTACTGTCATCTCTGTCTTATAAAGGTAAGAAATTGGC GCCAATTTCTTACCTTTATAAGACAGAGATGACAGTAAAACATTCTCTGG
c.1083_1092del	Ex10	TTTTACTGTCATCTCAAGAAGAGGAAAGGTAAGAAATTGGCATTCTGTAT ATACAGAATGCCAATTTCTTACCTTTCTCTTCTTGAGATGACAGTAAAA

Tabla Suplementaria 2. Primers para incorporar las 87 variantes exónicas candidatas de los exones 8 y 10 de *CHEK2* en el mgChk2_ex6-10.

Variante ¹	Exón	Primers (5'→3')
c.877G>M	Ex8	TCAAGATTAAAACTTTTTTTCATGCAGAAGATTATTATATT
		AATATAATAATCTTCTGCATKAAAAAAGTTTTTAATCTTGA
c.877G>T	Ex8	TCAAGATTAAAACTTTTTTTCATGCAGAAGATTATTATATT
		AATATAATAATCTTCTGCATAAAAAAAGTTTTTAATCTTGA
c.878A>S	Ex8	CAAGATTAAAACTTTTTTGSTGCAGAAGATTATTATATTG
		CAATATAATAATCTTCTGCASCAAAAAAAGTTTTTAATCTTG
c.878A>T	Ex8	CAAGATTAAAACTTTTTTGTTGCAGAAGATTATTATATTG
		CAATATAATAATCTTCTGCAACAAAAAAGTTTTTAATCTTG
c.879T>V	Ex8	AAGATTAAAACTTTTTTGAVGCAGAAGATTATTATATTGT
		ACAATATAATAATCTTCTGCBTCAAAAAAAGTTTTTAATCTT
c.880G>H	Ex8	AGATTAAAACTTTTTTGATHCAGAAGATTATTATATTGTT
		AACAATATAATAATCTTCTGDATCAAAAAAAGTTTTTAATCT
c.881C>R	Ex8	GATTAAAACTTTTTTGATGRAGAAGATTATTATATTGTTT
		AAACAATATAATAATCTTCTYCATCAAAAAAAGTTTTTAATC
c.881C>T	Ex8	GATTAAAACTTTTTTGATGTAGAAGATTATTATATTGTTT
		AAACAATATAATAATCTTCTACATCAAAAAAAGTTTTTAATC
c.882A>B	Ex8	ATTAAAACTTTTTTGATGCBGAAGATTATTATATTGTTTT
		AAAAACAATATAATAATCTTCTVGCATCAAAAAAAGTTTTTAAT
c.883G>M	Ex8	TTAAAACTTTTTTGATGCAMAAGATTATTATATTGTTTTG
		CAAAACAATATAATAATCTTKTGCATCAAAAAAAGTTTTTAA
c.883G>T	Ex8	TTAAAACTTTTTTGATGCATAAGATTATTATATTGTTTTG
		CAAAACAATATAATAATCTTATGCATCAAAAAAAGTTTTTAA
c.884A>S	Ex8	TAAAACTTTTTTGATGCAGSAGATTATTATATTGTTTTGG
		CCAAAACAATATAATAATCTSCTGCATCAAAAAAAGTTTTTA
c.884A>T	Ex8	TAAAACTTTTTTGATGCAGTAGATTATTATATTGTTTTGG
		CCAAAACAATATAATAATCTACTGCATCAAAAAAAGTTTTTA
c.885A>B	Ex8	AAAACTTTTTTGATGCAGABGATTATTATATTGTTTTGGA
		TCCAAAACAATATAATAATCVTCTGCATCAAAAAAAGTTTTT
c.886G>M	Ex8	AAAACTTTTTGATGCAGAAMATTATTATATTGTTTTGGAA
		TTCCAAAACAATATAATAATKTTCTGCATCAAAAAAAGTTTT

Variante ¹	Exón	Primers (5'→3')
c.886G>T	Ex8	AAAAC TTTT TGGATGCAGAATATTATTATATTG TTTTGGAA TTCCAAAACAATATAATAATATTCTGCATCAAAAAAGTTT
c.887A>S	Ex8	AAAAC TTTT TGGATGCAGAAGSTTATTATATTG TTTTGGAAAT ATTCCAAAACAATATAAASCTTCTGCATCAAAAAAGTTT
c.887A>T	Ex8	AAAAC TTTT TGGATGCAGAAGTTTATTATATTG TTTTGGAAAT ATTCCAAAACAATATAAATAA ACTTCTGCATCAAAAAAGTTT
c.888T>V	Ex8	AACTTTT TGGATGCAGAAGAVTATTATATTG TTTTGGAAATT AATTCAAAACAATATAATABTCTTCTGCATCAAAAAAGTT
c.889T>V	Ex8	ACTTTT TGGATGCAGAAGATVATTATATTG TTTTGGAAATTG CAATTCAAAACAATATAATBATCTTCTGCATCAAAAAAGT
c.890A>S	Ex8	CTTTT TGGATGCAGAAGATTSTTATATTG TTTTGGAAATTGT ACAATTCAAAACAATATAA SAATCTTCTGCATCAAAAAAG
c.890A>T	Ex8	CTTTT TGGATGCAGAAGATTTTATATTG TTTTGGAAATTGT ACAATTCAAAACAATATAA AATCTTCTGCATCAAAAAAG
c.891T>V	Ex8	TTTTT TGGATGCAGAAGATTA VTATATTG TTTTGGAAATTGTA TACAATTCAAAACAATATA BTAATCTTCTGCATCAAAAA
c.1073A>B	Ex10	GAATG TTTTACTGTCATCTCBAGAAGAGGACTGTCTTATAA TTATAAGACAGTCCTCTCTV GAGATGACAGTAAAACATTC
c.1074A>C	Ex10	AATGT TTTTACTGTCATCTCACGAAGAGGACTGTCTTATAAA TTTATAAGACAGTCCTCTT CGTGAGATGACAGTAAAACATT
c.1074A>K	Ex10	AATGT TTTTACTGTCATCTCAKGAAGAGGACTGTCTTATAAA TTTATAAGACAGTCCTCTT CMTGAGATGACAGTAAAACATT
c.1075G>H	Ex10	ATGT TTTTACTGTCATCTCAAHAAGAGGACTGTCTTATAAAG CTTTATAAGACAGTCCTCTT DTTGAGATGACAGTAAAACAT
c.1076A>Y	Ex10	TGTTT TACTGTCATCTCAAGYAGAGGACTGTCTTATAAAGG CCTTTATAAGACAGTCCTCT RCTTGAGATGACAGTAAAACA
c.1076A>G	Ex10	TGTTT TACTGTCATCTCAAGGAGAGGACTGTCTTATAAAGG CCTTTATAAGACAGTCCTCT CTTGAGATGACAGTAAAACA
c.1077A>B	Ex10	GTTT TACTGTCATCTCAAGABGAGGACTGTCTTATAAAGGT ACCTTTATAAGACAGTCCT CVTCTTGAGATGACAGTAAAAC

Variante ¹	Exón	Primers (5'→3')
c.1078G>M	Ex10	TTTACTGTCATCTCAAGAAMAGGACTGTCTTATAAAGGTA
		TACCTTTATAAGACAGTCCCTKTTCTTGAGATGACAGTAAAA
c.1078G>T	Ex10	TTTACTGTCATCTCAAGAATAGGACTGTCTTATAAAGGTA
		TACCTTTATAAGACAGTCCCTATTCTTGAGATGACAGTAAAA
c.1079A>S	Ex10	TTTACTGTCATCTCAAGAAGSGGACTGTCTTATAAAGGTAA
		TTACCTTTATAAGACAGTCCSCTTCTTGAGATGACAGTAAA
c.1079A>T	Ex10	TTACTGTCATCTCAAGAAGTGGACTGTCTTATAAAGGTA
		TACCTTTATAAGACAGTCCACTTCTTGAGATGACAGTAA
c.1080G>A	Ex10	TACTGTCATCTCAAGAAGAAGACTGTCTTATAAAGGTAA
		TTACCTTTATAAGACAGTCTTCTTCTTGAGATGACAGTA
c.1080G>Y	Ex10	TTACTGTCATCTCAAGAAGAYGACTGTCTTATAAAGGTAAG
		CTTACCTTTATAAGACAGTCRTCTTCTTGAGATGACAGTAA
c.1081G>A	Ex10	TACTGTCATCTCAAGAAGAGAAGTGTCTTATAAAGGTAAGA
		TCTTACCTTTATAAGACAGTTCTCTTCTTGAGATGACAGTA
c.1081G>C	Ex10	TACTGTCATCTCAAGAAGAGCACTGTCTTATAAAGGTAAGA
		TCTTACCTTTATAAGACAGTGCTCTTCTTGAGATGACAGTA
c.1081G>T	Ex10	ACTGTCATCTCAAGAAGAGTACTGTCTTATAAAGGTAAG
		CTTACCTTTATAAGACAGTACTCTTCTTGAGATGACAGT
c.1082A>B	Ex10	ACTGTCATCTCAAGAAGAGGBCTGTCTTATAAAGGTAAGAA
		TTCTTACCTTTATAAGACAGVCCCTTCTTGAGATGACAGT
c.1083C>A	Ex10	CTGTCATCTCAAGAAGAGGAATGTCTTATAAAGGTAAGAAA
		TTTCTTACCTTTATAAGACATTCCCTCTTCTTGAGATGACAG
c.1083C>K	Ex10	TGTCATCTCAAGAAGAGGAKTGTCTTATAAAGGTAAGAA
		TTCTTACCTTTATAAGACAMTCCCTCTTCTTGAGATGACA
c.1084T>C	Ex10	TGTCATCTCAAGAAGAGGACCGTCTTATAAAGGTAAGAAAT
		ATTTCTTACCTTTATAAGACGGTCCCTCTTCTTGAGATGACA
c.1085G>A	Ex10	TCATCTCAAGAAGAGGACTATCTTATAAAGGTAAGAAAT
		ATTTCTTACCTTTATAAGATAGTCCCTCTTCTTGAGATGA
c.1085G>T	Ex10	GTCATCTCAAGAAGAGGACTTTCTTATAAAGGTAAGAAATT
		AATTTCTTACCTTTATAAGAAAGTCCCTCTTCTTGAGATGAC
c.1086T>G	Ex10	CATCTCAAGAAGAGGACTGGCTTATAAAGGTAAGAAATT
		AATTTCTTACCTTTATAAGCCAGTCCCTCTTCTTGAGATG

Variante ¹	Exón	Primers (5'→3')
c.1087C>T	Ex10	ATCTCAAGAAGAGGACTGTTTTATAAAGGTAAGAAATTG
		CAATTTCTTACCTTTATAAAACAGTCCTCTTCTTGAGAT
c.1089T>C	Ex10	CTCAAGAAGAGGACTGTCTCATAAAGGTAAGAAATTGGC
		GCCAATTTCTTACCTTTATGAGACAGTCCTCTTCTTGAG
c.1089_1092del	Ex10	TGTCATCTCAAGAAGAGGACTGTCTAAGGTAAGAAATTGGCATTCTGTAT
		ATACAGAATGCCAATTTCTTACCTTAGACAGTCCTCTTCTTGAGATGACA
c.1091T>A	Ex10	TCAAGAAGAGGACTGTCTTAAAAAGGTAAGAAATTGGCATT
		AATGCCAATTTCTTACCTTTTTAAGACAGTCCTCTTCTTGA
c.1091T>C	Ex10	TCAAGAAGAGGACTGTCTTACAAAGGTAAGAAATTGGCATT
		AATGCCAATTTCTTACCTTTGTAAGACAGTCCTCTTCTTGA
¹ M= A/C; S= G/C; V= A/C/G; H= A/C/T; R= A/G; B= C/G/T; K=G/T; Y= C/T		

Tabla Suplementaria 3. Resultados de *splicing* de las variantes exónicas no espliceogénicas de los exones 8 y 10 de *CHEK2*.

Variante (HGVS)	Proteína (HGVS)	Transcritos ¹			
		Canónico	PTC	<i>In-frame</i>	No caracterizados
mgChk2_ex11-15 WT (6 ensayos)		91,5%± 1,5%	Δ(E8): 1,7%± 0,7%	Δ(E10): 6,8%± 0,9%	
<u>Ex8</u>					
c.877G>A	p.(Asp293Asn)	92,4%± 0,0%	Δ(E8): 1,0%± 0,1%	Δ(E10): 6,6%± 0,1%	
c.877G>C	p.(Asp293His)	90,9%± 2,2%	Δ(E8): 1,0%± 0,1%	Δ(E10): 5,4%± 0,4%	325-nt: 2,7%± 2,3%
c.877G>T	p.(Asp293Tyr)	84,2%± 0,5%	Δ(E8): 5,0%± 0,9%	Δ(E10): 7,6%± 0,2%	520-nt: 3,2%± 0,4%
c.878A>C	p.(Asp293Ala)	91,6%± 0,2%		Δ(E10): 7,2%± 0,3%	630-nt: 1,2%± 0,0%
c.878A>G	p.(Asp293Gly)	90,6%± 1,2%		Δ(E10): 8,4%± 0,2%	582-nt: 1,0%± 0,9%
c.878A>T	p.(Asp293Val)	86,9%± 0,7%	Δ(E8): 2,0%± 0,2%	Δ(E10): 9,7%± 0,5%	520-nt: 1,4%± 0,3%
c.879T>A	p.(Asp293Glu)	93,9%± 0,1%		Δ(E10): 6,1%± 0,1%	
c.879T>C	p.(Asp293=)	95,4%± 0,1%		Δ(E10): 4,6%± 0,1%	
c.879T>G	p.(Asp293Glu)	92,5%± 0,0%		Δ(E10): 7,5%± 0,0%	
c.880G>A	p.(Ala294Thr)	93,6%± 0,2%		Δ(E10): 6,4%± 0,2%	
c.880G>C	p.(Ala294Pro)	95,7%± 0,0%		Δ(E10): 4,3%± 0,0%	
c.880G>T	p.(Ala294Ser)	93,5%± 0,0%		Δ(E10): 6,5%± 0,0%	
c.881C>A	p.(Ala294Glu)	91,7%± 0,0%		Δ(E10): 8,3%± 0,0%	
c.881C>G	p.(Ala294Gly)	93,1%± 0,1%		Δ(E10): 6,9%± 0,1%	
c.881C>T	p.(Ala294Val)	86,8%± 0,4%	Δ(E8): 2,5%± 0,1%	Δ(E10): 9,0%± 0,4%	572-nt: 1,7%± 0,0%
c.882A>C	p.(Ala294=)	92,1%± 0,0%		Δ(E10): 7,9%± 0,0%	
c.882A>G	p.(Ala294=)	91,8%± 0,1%		Δ(E10): 8,2%± 0,1%	
c.882A>T	p.(Ala294=)	95,0%± 1,2%		Δ(E10): 5,0%± 1,2%	
c.885A>C	p.(Glu295Asp)	83,9%± 0,9%	Δ(E8): 4,7%± 1,6% Δ(E8) Δ(E10): 1,5%± 0,4%	Δ(E10): 8,5%± 0,3%	321-nt: 1,4%± 0,2%
c.887A>C	p.(Asp296Ala)	86,0%± 0,2%	Δ(E8): 11,7%± 0,4%	Δ(E10): 2,3%± 0,3%	
c.888T>A	p.(Asp296Glu)	90,3%± 1,9%	Δ(E8): 1,4%± 0,3%	Δ(E10): 8,3%± 1,6%	
c.888T>C	p.(Asp296=)	95,4%± 0,6%	Δ(E8): 1,8%± 0,2%	Δ(E10): 2,8%± 0,5%	
c.888T>G	p.(Asp296Glu)	95,7%± 0,2%	Δ(E8): 1,8%± 0,2%	Δ(E10): 2,5%± 0,1%	
c.889T>A	p.(Tyr297Asn)	91,9%± 1,2%	Δ(E8): 2,2%± 0,4%	Δ(E10): 4,7%± 0,8%	325-nt: 1,2%± 0,3%
c.889T>C	p.(Tyr297His)	92,6%± 1,5%		Δ(E10): 5,2%± 0,0%	325-nt: 2,2%± 1,6%
c.889T>G	p.(Tyr297Asp)	94,6%± 0,2%		Δ(E10): 5,4%± 0,2%	
c.890A>C	p.(Tyr297His)	92,8%± 0,4%		Δ(E10): 5,3%± 0,1%	325-nt: 1,9%± 0,2%
c.890A>G	p.(Tyr297Cys)	90,3%± 0,8%		Δ(E10): 8,4%± 0,8%	630-nt: 1,3%± 0,0%

Variante (HGVS)	Proteína (HGVS)	Transcritos ¹			
		Canónico	PTC	<i>In-frame</i>	No caracterizados
c.891T>A	p.(Tyr297*)	86,5%± 0,6%	Δ(E8): 1,6%± 0,3%	Δ(E10): 10,0%± 0,2%	572-nt: 1,9%± 0,2%
c.891T>C	p.(Tyr297=)	92,3%± 0,1%		Δ(E10): 7,7%± 0,1%	
c.891T>G	p.(Tyr297*)	84,6%± 0,2%	Δ(E8): 6,0%± 0,1%	Δ(E10): 9,4%± 0,4%	
Ex10					
c.1073A>C	p.(Gln358Pro)	87,6%± 0,5%	Δ(E8): 3,2%± 0,2%	Δ(E10): 9,2%± 0,4%	
c.1073A>G	p.(Gln358Arg)	93,5%± 0,4%	Δ(E8): 2,0%± 0,3%	Δ(E10): 4,5%± 0,3%	
c.1074A>C	p.(Gln358His)	88,2%± 0,5%	Δ(E8): 3,3%± 0,1%	Δ(E10): 7,4%± 0,3%	572-nt: 1,1%± 0,2%
c.1074A>G	p.(Gln358=)	82,4%± 0,5%	Δ(E8): 1,7%± 0,0%	Δ(E10): 15,9%± 0,5%	
c.1074A>T	p.(Gln358His)	90,5%± 1,4%	Δ(E8): 2,0%± 0,2%	Δ(E10): 7,5%± 1,4%	
c.1077A>G	p.(Glu359=)	88,2%± 0,9%	Δ(E8): 1,6%± 0,3%	Δ(E10): 10,2%± 0,6%	
c.1079A>C	p.(Glu360Ala)	83,6%± 0,7%	Δ(E8): 2,3%± 0,1%	Δ(E10): 14,1%± 0,7%	
c.1080G>A	p.(Glu360=)	83,0%± 0,3%	Δ(E8): 2,0%± 0,3%	Δ(E10): 15,0%± 0,2%	
c.1080G>C	p.(Glu360Asp)	90,7%± 1,3%	Δ(E8): 2,4%± 0,5%	Δ(E10): 6,9%± 0,8%	
c.1082A>C	p.(Asp361Ala)	81,8%± 1,8%	Δ(E8): 2,0%± 0,1%	Δ(E10): 16,2%± 1,7%	
c.1084T>C	p.(Cys362Arg)	92,4%± 0,8%	Δ(E8): 2,4%± 0,7%	Δ(E10): 4,1%± 0,1%	572-nt: 1,1%± 0,2%
c.1085G>A	p.(Cys362Tyr)	87,4%± 0,4%	Δ(E8): 7,7%± 0,2%	Δ(E10): 3,5%± 0,1%	572-nt: 1,4%± 0,2%
c.1085G>T	p.(Cys362Phe)	91,2%± 0,8%	Δ(E8): 3,3%± 0,1%	Δ(E10): 5,5%± 0,7%	
c.1087C>T	p.(Leu363Phe)	90,9%± 1,1%	Δ(E8): 2,3%± 0,5%	Δ(E10): 5,2%± 0,5%	572-nt: 1,6%± 0,6%
c.1089T>C	p.(Leu363=)	84,9%± 0,9%	Δ(E8): 1,8%± 0,1%	Δ(E10): 11,5%± 0,5%	572-nt: 1,8%± 0,4%
c.1089_1092del	p.(Ile364fs*)	82,2%± 0,7%	Δ(E8): 2,2%± 0,4%	Δ(E10): 15,6%± 0,3%	
c.1091T>A	p.(Ile364Lys)	89,2%± 0,1%	Δ(E8): 1,4%± 0,1%	Δ(E10): 8,4%± 0,1%	630-nt: 1,0%± 0,0%
c.1091T>C	p.(Ile364Thr)	87,9%± 1,3%	Δ(E8): 1,7%± 0,2%	Δ(E10): 8,7%± 1,0%	572-nt: 1,7%± 0,5%

¹Los transcritos de cada variante en la que no se suprime el exón afectado incorporarán la mutación puntual ensayada.

Tabla Suplementaria 4. Análisis bioinformático de las variantes exónicas de *CHEK2* candidatas empleando distintos *software* de predicción de SRE.

Variante de <i>CHEK2</i> ¹	HEXplorer (ΔHZ_{EI}) ^{2,4}	HOT-SKIP (ESS/ESE ratio) ^{3,4}	HEXoSplice ($\Delta\text{tESRseq}$) ^{2,4}	% Transcritos anómalos
c.877G>A	-112,56	13,00	-2,2383	7,6%
c.877G>C	-64,08	12,00	+0,0116	9,1%
c.877G>T	-153,61	4,67	-2,1134	15,8%
c.878A>C	4,27	1,50	-0,3675	8,4%
c.878A>G	-28,43	7,50	-1,6363	9,4%
c.878A>T	-95,96	8,00	-0,9938	13,1%
c.879T>A	19,69	0,11	+0,5330	6,1%
c.879T>C	-7,78	0,67	+1,1630	4,6%
c.879T>G	-62,41	0,50	-1,1780	7,5%
c.880G>A	-57,49	0,14	-0,9126	6,4%
c.880G>C	0,11	0,00	+0,4145	4,3%
c.880G>T	-66,86	0,08	-0,6869	6,5%
c.881C>A	74,42	0,04	+2,6008	8,3%
c.881C>G	18,87	1,25	+1,0482	6,9%
c.881C>T	-98,38	0,06	-2,4280	13,2%
c.882A>C	-36,64	0,00	+0,2017	7,9%
c.882A>G	-33,64	0,00	+0,9579	8,2%
c.882A>T	-4,05	0,00	-0,2922	5,0%
c.883G>A	-77,07	0,07	-3,2450	97,4%
c.883G>C	-63,13	0,00	-2,0770	42,2%
c.883G>T	-163,75	3,00	-3,5335	97,8%
c.884A>C	-34,76	0,00	-3,4520	24,3%
c.884A>G	-57,43	3,00	-2,7032	47,3%
c.884A>T	-137,08	0,10	-5,2654	96,0%
c.885A>C	-58,62	0,00	-1,5015	16,1%
c.885A>G	-64,04	0,43	-2,1503	18,5%
c.885A>T	-54,36	0,00	-2,9609	39,1%
c.886G>A	-96,43	0,17	-3,6765	51,0%
c.886G>C	-71,73	0,00	-2,7384	25,4%
c.886G>T	-130,62	0,67	-3,7661	79,4%
c.887A>C	-22,73	0,13	-2,4433	14,0%
c.887A>G	-51,6	1,57	-3,4783	47,3%
c.887A>T	-89,99	2,33	-2,2472	43,3%
c.888T>A	15,8	0,18	+0,1062	9,7%
c.888T>C	31,02	0,00	+0,1360	4,5%
c.888T>G	11,14	0,29	-1,0121	4,3%
c.889T>A	-13,69	0,60	-0,2196	8,1%
c.889T>C	64,87	0,11	+1,2495	7,4%
c.889T>G	92,06	0,36	+1,2706	5,4%
c.890A>C	24,77	8,00	+0,6860	7,2%
c.890A>G	21,7	13,00	+1,0965	9,7%

Variante de <i>CHEK2</i> ¹	HEXplorer (ΔHZ_{EI}) ^{2,4}	HOT-SKIP (ESS/ESE ratio) ^{3,4}	HEXoSplice ($\Delta tESRseq$) ^{2,4}	% Transcritos anómalos
c.890A>T	-83,0	10,00	+0,1206	21,7%
c.891T>A	-11,58	7,00	+0,0012	13,5%
c.891T>C	66,91	0,60	+1,1966	7,7%
c.891T>G	23,33	14,00	-0,8448	15,4%
c.1073A>C	-13,58	0,00	-1,7896	13,4%
c.1073A>G	-16,35	0,25	-0,2501	6,5%
c.1073A>T	-114,23	0,00	-2,6982	23,4%
c.1074A>C	-82,3	0,00	-1,8565	11,8%
c.1074A>G	-77,41	0,00	-2,6760	17,6%
c.1074A>T	-63,18	0,00	-1,3564	9,5%
c.1075G>A	-100,38	0,00	-4,2435	61,6%
c.1075G>C	-52,5	0,00	-2,3327	24,3%
c.1075G>T	-160,32	0,00	-4,4416	72,8%
c.1076A>C	-75,45	0,00	-3,1443	26,4%
c.1076A>G	-72,27	0,00	-2,8809	28,9%
c.1076A>T	-153,95	0,00	-4,8228	61,3%
c.1077A>C	-57,18	0,00	-0,3163	26,8%
c.1077A>G	-59	0,00	-2,1155	11,8%
c.1077A>T	-47,67	0,00	-2,2330	25,5%
c.1078G>A	-57,46	0,00	-1,6932	35,4%
c.1078G>C	-67,85	0,00	-2,6604	42,0%
c.1078G>T	-159,86	0,00	-3,6174	73,9%
c.1079A>C	-32,7	0,00	-0,1118	16,4%
c.1079A>G	-121,43	0,00	-3,5354	34,9%
c.1079A>T	-48,35	0,80	-1,4955	23,6%
c.1080G>A	57,25	0,00	+2,0714	17,0%
c.1080G>C	3,69	0,00	+1,6345	9,3%
c.1080G>T	17,55	0,06	+0,2733	95,2%
c.1081G>A	-0,6	0,00	+0,0731	24,8%
c.1081G>C	-24,15	0,00	-1,9522	24,0%
c.1081G>T	-50,3	0,33	-1,7542	49,5%
c.1082A>C	-46,58	0,13	-2,1431	18,2%
c.1082A>G	-87,2	3,00	-2,2549	31,8%
c.1082A>T	-79,08	1,50	-2,3888	23,3%
c.1083C>A	-7,8	0,86	-1,5070	27,4%
c.1083C>G	-6,97	1,67	-1,8781	23,7%
c.1083C>T	-22,62	0,38	-1,3504	23,0%
c.1084T>C	52,58	0,00	+0,2562	7,6%
c.1085G>A	13,71	0,75	-1,1462	12,6%
c.1085G>T	-47,86	5,00	-0,8028	8,8%
c.1086T>G	39,21	1,00	-0,3894	20,3%
c.1087C>T	-76,52	12,00	-0,2986	9,1%
c.1089T>C	24,14	2,00	+1,6830	15,1%

Variante de <i>CHEK2</i> ¹	HEXplorer ($\Delta HZEI$) ^{2,4}	HOT-SKIP (ESS/ESE ratio) ^{3,4}	HEXoSplice ($\Delta tESRseq$) ^{2,4}	% Transcritos anómalos
c.1089_1092del	52,97	N/A	N/A	17,8%
c.1091T>A	-6,3	0,50	+0,7316	10,8%
c.1091T>C	30,86	0,00	+1,7598	12,1%

¹Las variantes espliceogénicas se muestran en negrita. ²Los valores $\Delta tESRseq$ y $\Delta HZEI$ negativos indican una disminución de motivos ESE o un aumento de ESS. ³Cuanto mayor es el valor HOT-SKIP, más prevalecen los ESS sobre los ESE en la secuencia modificada. ⁴Los valores $\Delta HZEI$, HOT-SKIP y $\Delta tESRseq$ por debajo o por encima de sus correspondientes puntos de corte (≤ -5 , ≥ 1 y $\leq -0,75$, respectivamente) aparecen en negrita.

Tabla Suplementaria 5. Análisis comparativo de las predicciones de SpliceAI y las lecturas de los minigenes mgChk2_ex6-10 mutantes.

Variante de <i>CHEK2</i> ¹	AL ²	DL ²	SpliceAI ²	Lecturas del mgChk2_ex6-10 ³
Variantes del exón 8 de <i>CHEK2</i>				
c.877G>A	0,30 (30)	0,30 (-31)	Δ(E8)	no espliceogénica
c.877G>C	0,17 (30)	0,16 (-31)	no espliceogénica	no espliceogénica
c.877G>T	0,32 (30)	0,33 (-31)	Δ(E8)	no espliceogénica
c.878A>C			no espliceogénica	no espliceogénica
c.878A>G	0,05 (31)	0,03 (-30)	no espliceogénica	no espliceogénica
c.878A>T	0,29 (31)	0,27 (-30)	Δ(E8)	no espliceogénica
c.879T>A			no espliceogénica	no espliceogénica
c.879T>C			no espliceogénica	no espliceogénica
c.879T>G	0,07 (32)	0,08 (-29)	no espliceogénica	no espliceogénica
c.880G>A	0,19 (33)	0,17 (-28)	no espliceogénica	no espliceogénica
c.880G>C	0,06 (33)	0,03 (-28)	no espliceogénica	no espliceogénica
c.880G>T	0,22(33)	0,22 (-28)	Δ(E8)	no espliceogénica
c.881C>A			no espliceogénica	no espliceogénica
c.881C>G			no espliceogénica	no espliceogénica
c.881C>T	0,21(34)	0,20 (-27)	Δ(E8)	no espliceogénica
c.882A>C			no espliceogénica	no espliceogénica
c.882A>G			no espliceogénica	no espliceogénica
c.882A>T		0,01 (-26)	no espliceogénica	no espliceogénica
c.883G>A	0,52 (36)	0,50 (-25)	Δ(E8)	Δ(E8) (97%)
c.883G>C	0,29 (36)	0,27 (-25)	Δ(E8)	Δ(E8) (40%)
c.883G>T	0,79 (36)	0,76 (-25)	Δ(E8)	Δ(E8) (98%)
c.884A>C	0,27 (37)	0,25 (-24)	Δ(E8)	Δ(E8) (14%)
c.884A>G	0,35 (37)	0,31 (-24)	Δ(E8)	Δ(E8) (44%)
c.884A>T	0,71 (37)	0,71 (-24)	Δ(E8)	Δ(E8) (95%)
c.885A>C	0,19 (38)	0,18 (-23)	no espliceogénica	no espliceogénica
c.885A>G	0,10 (38)	0,12 (-23)	no espliceogénica	Δ(E8) (14%)
c.885A>T	0,39 (38)	0,40 (-23)	Δ(E8)	Δ(E8) (35%)
c.886G>A	0,56 (39)	0,53 (-22)	Δ(E8)	Δ(E8) (45%)
c.886G>C	0,47 (39)	0,46 (-22)	Δ(E8)	Δ(E8) (20%)
c.886G>T	0,64 (39)	0,63 (-22)	Δ(E8)	Δ(E8) (77%)
c.887A>C	0,22 (40)	0,21 (-21)	Δ(E8)	no espliceogénica
c.887A>G	0,30 (40)	0,34 (-21)	Δ(E8)	Δ(E8) (44%)
c.887A>T	0,29 (40)	0,29 (-21)	Δ(E8)	Δ(E8) (35%)
c.888T>A			no espliceogénica	no espliceogénica
c.888T>C			no espliceogénica	no espliceogénica
c.888T>G	0,08 (41)	0,06 (-20)	no espliceogénica	no espliceogénica
c.889T>A	0,01 (42)	0,02 (-19)	no espliceogénica	no espliceogénica
c.889T>C			no espliceogénica	no espliceogénica
c.889T>G			no espliceogénica	no espliceogénica
c.890A>C			no espliceogénica	no espliceogénica
c.890A>G			no espliceogénica	no espliceogénica
c.890A>T	0,04 (43)	0,02 (-18)	no espliceogénica	Δ(E8) (11%)
c.891T>A	0,19 (44)	0,15 (-17)	no espliceogénica	no espliceogénica
c.891T>C			no espliceogénica	no espliceogénica
c.891T>G	0,22 (44)	0,20 (-17)	Δ(E8)	no espliceogénica

Variante de CHEK2 ¹	AL ²	DL ²	SpliceAI ²	Lecturas del mgChk2_ex6-10 ³
Variantes del exón 10 de CHEK2				
c.1073A>C	0,01 (64)	0,00 (-22)	no espliceogénica	no espliceogénica
c.1073A>G			no espliceogénica	no espliceogénica
c.1073A>T	0,05 (64)	0,02 (-22)	no espliceogénica	Δ(E10) (21%)
c.1074A>C	0,01 (65)	0,01 (-21)	no espliceogénica	no espliceogénica
c.1074A>G	0,04 (65)	0,03 (-21)	no espliceogénica	no espliceogénica
c.1074A>T	0,01 (65)	0,00 (-21)	no espliceogénica	no espliceogénica
c.1075G>A	0,11 (66)	0,07 (-20)	no espliceogénica	Δ(E10) (62%)
c.1075G>C	0,02 (66)	0,01 (-20)	no espliceogénica	Δ(E10) (22%)
c.1075G>T	0,23 (66)	0,23 (-20)	Δ(E10)	Δ(E10) (73%)
c.1076A>C	0,05 (67)	0,03 (-19)	no espliceogénica	Δ(E10) (25%)
c.1076A>G	0,03 (67)	0,02 (-19)	no espliceogénica	Δ(E10) (27%)
c.1076A>T	0,14 (67)	0,10 (-19)	no espliceogénica	Δ(E10) (61%)
c.1077A>C	0,01 (68)	0,01 (-18)	no espliceogénica	Δ(E10) (25%)
c.1077A>G	0,01 (68)	0,01 (-18)	no espliceogénica	no espliceogénica
c.1077A>T	0,02 (68)	0,01 (-18)	no espliceogénica	Δ(E10) (23%)
c.1078G>A	0,03 (69)	0,02 (-17)	no espliceogénica	Δ(E10) (34%)
c.1078G>C	0,05 (69)	0,02 (-17)	no espliceogénica	Δ(E10) (41%)
c.1078G>T	0,21 (69)	0,16 (-17)	no espliceogénica	Δ(E10) (74%)
c.1079A>C		0,01 (-16)	no espliceogénica	no espliceogénica
c.1079A>G	0,06 (70)	0,05 (-16)	no espliceogénica	Δ(E10) (31%)
c.1079A>T	0,01 (70)	0,01 (-16)	no espliceogénica	Δ(E10) (20%)
c.1080G>A			no espliceogénica	no espliceogénica
c.1080G>C			no espliceogénica	no espliceogénica
c.1080G>T			no espliceogénica	Δ(E10) (95%)
c.1081G>A	0,01 (72)	0,01 (-14)	no espliceogénica	Δ(E10) (23%)
c.1081G>C	0,02 (72)	0,02 (-14)	no espliceogénica	Δ(E10) (21%)
c.1081G>T	0,04 (72)	0,02 (-14)	no espliceogénica	Δ(E10) (48%)
c.1082A>C	0,01 (73)	0,01 (-13)	no espliceogénica	no espliceogénica
c.1082A>G	0,03 (73)	0,03 (-13)	no espliceogénica	Δ(E10) (30%)
c.1082A>T	0,06 (73)	0,04 (-13)	no espliceogénica	Δ(E10) (22%)
c.1083C>A	0,00 (74)	0,00 (-12)	no espliceogénica	Δ(E10) (26%)
c.1083C>G		0,00 (-12)	no espliceogénica	Δ(E10) (21%)
c.1083C>T	0,00 (74)	0,00 (-12)	no espliceogénica	Δ(E10) (20%)
c.1084T>C			no espliceogénica	no espliceogénica
c.1085G>A	0,00 (76)		no espliceogénica	no espliceogénica
c.1085G>T			no espliceogénica	no espliceogénica
c.1086T>G			no espliceogénica	Δ(E10) (17%)
c.1087C>T	0,01 (78)	0,00 (-8)	no espliceogénica	no espliceogénica
c.1089T>C			no espliceogénica	no espliceogénica
c.1089_1092del			no espliceogénica	no espliceogénica
c.1091T>A			no espliceogénica	no espliceogénica
c.1091T>C			no espliceogénica	no espliceogénica

¹Nomenclatura HGVS utilizando como referencia la secuencia NM_007194.4. Se han subrayado y resaltado en negrita las 38 variantes espliceogénicas. ²Parámetros de SpliceAI empleados (*genome version: hg38, score type: raw, max distance: ±4,999*). Se muestra la puntuación Δ (y la posición en el pre-ARNm) para la pérdida del aceptor (AL) y la pérdida del donador (DL) de los exones 8 o 10. Para mayor claridad, las puntuaciones Δ de alta precisión (puntuación Δ>0,80), recomendadas (puntuación Δ>0,50) y de alta sensibilidad (puntuación Δ>0,20) se han indicado en rojo, amarillo y verde, respectivamente. A efectos meramente informativos, las puntuaciones Δ en el rango 0,10-0,20 se han marcado en gris. ³Para las variantes espliceogénicas (resaltadas en negrita), se muestra la contribución de los transcritos que contienen Δ(E8) (variantes del exón 8) o Δ(E10) (variantes del exón 10) a la señal de lectura global (%).

Tabla Suplementaria 6. Sensibilidad, especificidad y precisión de SpliceAI.				
Variantes <i>CHEK2</i> ubicadas en el exón 8				
	Verdadero positivo ²	Falso positivo ²	Falso negativo ²	Verdadero negativo ²
SpliceAI positivo ¹	12	7		
SpliceAI negativo ¹			2	24
Estadísticas del exón 8 ³				
	Valor	95% IC		
Sensibilidad	85,71%	57,19-98,22		
Especificidad	77,42%	58,90-90,41		
<i>Likelihood Ratio</i> positivo	3,80	1,91-7,54		
<i>Likelihood Ratio</i> negativo	0,18	0,05-0,68		
Prevalencia (% de variantes espliceogénicas) ²	31,11%	18,17-46,65		
Valor predictivo positivo	63,16%	46,33-77,29		
Valor predictivo negativo	92,31%	76,63-97,77		
Precisión	80,00%	65,40-90,42		
Variantes <i>CHEK2</i> ubicadas en el exón 10				
	Verdadero positivo ²	Falso positivo ²	Falso negativo ²	Verdadero negativo ²
SpliceAI positivo ¹	1	0		
SpliceAI negativo ¹			23	18
Estadísticas del exón 10 ³				
	Valor	95% IC		
Sensibilidad	4,17%			
Especificidad	100%			
<i>Likelihood Ratio</i> positivo	-			
<i>Likelihood Ratio</i> negativo	0,96	0,88-1,04		
Prevalencia (% de variantes espliceogénicas) ²	57,14%	40,96-72,28		
Valor predictivo positivo	100%	-		
Valor predictivo negativo	43,90%	41,86-45,97		
Precisión	45,24%	28,85-61,33		
¹ Consideramos positivas las predicciones de SpliceAI (predicción de <i>skipping</i> de exón) si la puntuación $\Delta \geq 20\%$ tanto para la pérdida del aceptor como para la pérdida del donador (exón 8 o exón 10). De lo contrario, consideramos las predicciones de SpliceAI negativas (véase la Tabla Suplementaria 5). ² Consideramos que una lectura del minigén <i>mgChk2_ex8-10</i> mutante es positiva (alteración del <i>splicing</i> detectada) si la contribución del transcrito canónico a la expresión global se reduce un 10% (si se compara con la lectura del minigén WT). De lo contrario, consideramos la lectura del minigén <i>mgChk2_ex6-ex10</i> mutante negativa. ³ Las estadísticas se realizaron <i>online</i> empleando la evaluación de pruebas diagnósticas de MedCalc (www.medcalc.org/calc/).				

Tabla Suplementaria 7. Primers empleados para el clonaje consecutivo y la estabilización del minigén mgBARD1_ex3-5.

Primers cloning	Secuencia (5'→3')
BARD1_ex4_BAMHI_FW	TATATAGGATCCGGATCCTATTTGGAATATAAGTTTCA
BARD1_ex4_KPNI_RV	TATATAGGTACCGGTACCACAAAATGATAAAAACTGAA
BARD1_ex5_Ins-FW	CTATTTTACCTTCAGTTTTTATCATTTTGTTAATAATTTATTAACATGAACAGTAGCC
BARD1_ex5_Ins-RV	CCTCACAAAAGGGAACAAAAGCTGGGTACCCAAAAGGTAACCAATTTCTTGCCTGTCACA
BARD1_ex3_SACII_FW	TATATACCGCGGCAGGCTAAGTAAAGTCTCTT
BARD1_ex3_BAMH1_FW	TATATAGGATCCTGATCTTCCAAAAACCACC
BARD1_delivs3_250pb_FW	AATACGTATTTTACAGATTTTAAAATCATCTGCTATCAAAGAGTTCACACTC
BARD1_delivs3_250pb_RV	GAGTGTGAACTCTTTGATAGCAGATGATTTTAAAATCTGAAATACGTATT

Tabla Suplementaria 8. Primers para incorporar en el exón 4 del mgBARD1_ex3-5 las μ del de 30 y 12 nt.

Microdelección	Primers (5' →3')
c.367del30	TTCTTTTTCTTTTTCTGTCAGATTTTAATGATGCAGGAAACAAGAAGA TCTTCTGTTTCTGCATCATTAATAATCTGACAGAAAAAAGAAAAAGAA
c.392del30	TTGAAAGAAGATAAACCTAGGAAAAATCAATTAATAATGTGGTTTAGCCC GGGCTAAACCACATTTTAATGAATTTTTCCTAGGTTTATCTTCTTTCAA
c.417del30	GTTTGTTAATGATGCAGGAAACAATCGAAGTAAGAAAAGTCAGATATGTT AACATATCTGACTTCTTACTTCGATTGTTTTCCTGCATCATTAACAACAA
c.442del30	GAAGAATCAATTAATAATGTGGTTTGTGAGTAAAGCTTCAGTGCAAACCC GGGTTTGCCTGAAGCTTACTCACAACCACATTTTAATGAATCTTTC
c.467del30	AGCCCTCGAAGTAAGAAAAGTCAGATAGCCTGCAATAAAAAAAGATGCAAG CTTGCACTTTTTTATTGCAGGCTATCTGACTTTTCTTACTTCGAGGGCT
c.581del104	AAGTCCCTGCAGATGTTTCTGAGATTTGACTCCAAAGAGGAATCTAAG CTTAGATTCTCTTTGGAGTCAAACTCAGAAACATCTGCAGGAGGACTT
c.1182del30	AATTCATTAGTCTTTCACCAGGTACCAGGCGAGTGATGTCTAGTCCCTCA TGAGGGACTAGACATCACTCGCCTGGTACCTGGTGAAAGACTAATGAATT
c.1207del30	ACCACCTTCTACATTAAGTAGTTTCAGCAATGAAGCTGTTGCCCAATATGG CCATATTGGGCAACAGCTTCATTGCTGAACTACTTAATGTAGAAGGTGGT
c.1230del12	CAAGTTACAGGCGAGTGATGCTAGGAAGCTGTTGCCCAATATGGCTGTG CACAGCCATATTGGGCAACAGCTTCCCTAGACATCACTCGCCTGTAACCTG
c.1232del30	AGTTACAGGCGAGTGATGCTAGTCCGTGAAAAGAAATCATAGAGGAGA TCTCCTCTATGATTTCTTTTTACAGGACTAGACATCACTCGCCTGTAAC
c.1240del12	GCGAGTGATGCTAGTCCCTCAGCACCCAATATGGCTGTGAAAAGAAATC GATTTCTTTTTACAGCCATATTGGGTGCTGAGGGACTAGACATCACTCGC
c.1250del12	TCTAGTCCCTCAGCAATGAAGCTGTCTGTGAAAAGAAATCATAGAGGAGA TCTCCTCTATGATTTCTTTTTACAGACAGCTTCATTGCTGAGGGACTAGA
c.1255del12	TCCCTCAGCAATGAAGCTGTTGCCCAAAAAGAAATCATAGAGGAGAGACTT AAGTCTCTCCTCTATGATTTCTTTTTGGGCAACAGCTTCATTGCTGAGGGA
c.1257del30	CCTCAGCAATGAAGCTGTTGCCCAAGACTTTGCTCCATATTGCTTCTATT AATAGAAGCAATATGGAGCAAAGTCTTTGGGCAACAGCTTCATTGCTGAGG
c.1265del12	ATGAAGCTGTTGCCCAATATGGCTGATAGAGGAGAGACTTTGCTCCATAT ATATGGAGCAAAGTCTCTCCTCTATCAGCCATATTGGGCAACAGCTTCAT
c.1275del12	TGCCCAATATGGCTGTGAAAAGAAAGACTTTGCTCCATATTGCTTCTATT AATAGAAGCAATATGGAGCAAAGTCTTTCTTTTTACAGCCATATTGGGCA
c.1280del12	AATATGGCTGTGAAAAGAAATCATATGCTCCATATTGCTTCTATTAAGGT ACCTAATAGAAGCAATATGGAGCATATGATTTCTTTTTACAGCCATATT
c.1282del30	TATGGCTGTGAAAAGAAATCATAGAAAGGTAGGATGCTTACTCTGAAATA TATTTACAGAGTAAGCATCCTACCTTTCTATGATTTCTTTTTACAGCCATA
c.1290del12	TGAAAAGAAATCATAGAGGAGAGACTGCTTCTATTAAGGTAGGATGCTTA TAAGCATCCTACCTAATAGAAGCAGTCTCTCCTCTATGATTTCTTTTTCA
c.1300del12	TCATAGAGGAGAGACTTTGCTCCATAAGGTAGGATGCTTACTCTGAAATA TATTTACAGAGTAAGCATCCTACCTTATGGAGCAAAGTCTCTCCTCTATGA

Tabla Suplementaria 9. Sitios de unión del *enhancer* Tra2 β predichos por la herramienta HSF.

HGVS	Motivo	Score
c.371_375	AAGAA	100,00
c.374_378	AAGAT	86,88
c.412_416	AACAA	94,14
c.415_419	AAGAA	100,00
c.418_422	AAGAA	100,00
c.454_458	AAGAA	100,00
c.504_508	AATAA	98,53
c.507_511	AAAAA	94,14
c.508_512	AAAAA	94,14
c.509_513	AAAAA	94,14
c.512_516	AAGAT	86,88
c.530_534	AAGAC	86,88
c.586_590	AAAAA	94,14
c.600_604	AAGAT	86,88
c.609_613	AAAAA	94,14
c.610_614	AAAAA	94,14
c.617_621	AAAAA	94,14
c.618_622	AAAAA	94,14
c.619_623	AAAAA	94,14
c.622_626	AAGAA	100,00
c.647_651	AAAAA	94,14
c.668_672	AAAAA	94,14
c.671_675	AAGAA	100,00
c.674_678	AAGAT	86,88
c.695_699	AAGAG	89,07
c.771_775	AATAG	87,60
c.855_859	AATAG	87,60
c.874_878	AAGAG	89,07
c.913_917	AAAAA	94,14
c.931_935	AAGAA	100,00
c.952_956	AATAA	98,53
c.976_980	AATAG	87,60
c.1000_1004	AAGAG	89,07
c.1066_1070	AATAT	85,42
c.1135_1139	AAAAA	94,14
c.1255_1259	AATAT	85,42
c.1269_1273	AAGAA	100,00

En negrita se marcan los motivos que muestran un mayor *score* y en azul se resalta la región rica en sitios de unión para este factor regulador del *splicing*.

Tabla Suplementaria 10. Primers para destruir los motivos silenciadores predichos alrededor de los donadores crípticos del exón 4 de *BARD1*.

Sitios crípticos ¹	Primers (5'→3') ²
SC2_5'	TTAAAATGTGGTTTAAACCCCTCGAAGTAACAAAGTCAGATCTGTTGTGAGTAAAGC
	GCTTTACTCACAAACAGATCTGACTTTTGTTACTTTCGAGGGTTAAACCACATTTTAA
SC2_3'	AGATCTGTTGTGAGTCAAGCTTCAATGCAAACCCAGCCTG
	CAGGCTGGGTTTGCAATTGAAGCTTGACTCACAAACAGATCT
SC4_5'	GCTGAGCAAATTGAGTCTCCAGACACTCACAGCAGCAATGAAGCAGTTACTCCTGAGA
	TCTCAGGAGTAACTGCTTCATTGCTGCTGTGAGTGTCTGGAGACTCAATTTGCTCAGC
SC4_3'	GTTACTCCTGACAAGGTCTGCAAAAATTATCTTACATCAAAGAAATCTTTGCCATTGGAAAATAAT
	ATTATTTTCCAATGGCAAAGATTTCTTTGATGTAAGATAATTTTTGCAGACCTTGTCAGGAGTAAC
SC6_5'	TGGTGGTACATCATTTGACGAAAAACAGTAACATGTCCGATGAATTCATTGGTCTTTCACC
	GGTGAAAGACCAATGAATTCATCGGACATGTTACTGTTTTTCGTCAATGATGTACCACCA
SC6_3'_1	CACCTTCTACATTAACCTAGTTCAAGTTAAAGGCGAGTGATGTCAAGTCCCTCAGCAATG
	CATTGCTGAGGGACTTGACATCACTCGCCTTAACTTGAAGTAAATGTAGAAGGTG
SC6_3'_2	CACCTTCTACATTAACCTGGTTCAAGTTAACGGCGAATGATGTCAAGTCCCTCAGCAATG
	CATTGCTGAGGGACTTGACATCATTCGCCGTTAACTTGAACCAGTTAATGTAGAAGGTG

¹SC2, sitio críptico 2 (c.469_477) localizado 843 nt *upstream* del donador canónico del exón 4 de *BARD1*; SC4, sitio críptico 4 (c.904_912) ubicado 408 nt *upstream* del 5'ss del exón largo del gen *BARD1*; SC6, sitio críptico 6 (c.1175_1183) el cual se sitúa 137 nt *upstream* del donador tradicional del exón 4 de *BARD1*. ²En rojo aparecen marcados los nucleótidos que hemos mutado respecto a la secuencia de referencia NM_000465.4.

Tabla Suplementaria 11. Primers empleados para el clonaje del minigén mgBRIP1_ex1-2 y su estabilización.

Primers <i>cloning</i>	Secuencia (5'→3')
BRIP1_ex2_XBAI_FW	TATATATCTAGATGATTTATAAGACTAGGAAATTATT
BRIP1_ex2_BAMHI_RV	TATATAGGATCCGTTTTTTCATTAAGTATAGAAATT
BRIP1_ex1_FW	GGTCTGAGTCACCTGGACAACCTCAAAGGCGGAATTCGTCTCGGGTTGTG
BRIP1_ex1_RV	AACAAAATAATTTTCTAGTCTTATAAATCACAAACATTGCATAAAGGATAAACAC
BRIP1_delex1_52pb_FW	GAGTCACCTGGACAACCTCAAAGGCAACGGCGGCCCCAGTCCTGCACACA
BRIP1_delex1_52pb_RV	TGTGTGCAGGACTGGGGCCGCCGTTGCCTTTGAGGTTGTCCAGGTGACTC

Tabla Suplementaria 12. Primers para incorporar en los exones 1 y 2 del mgBRIP1_ex1-2 las μ del de 16 y 30 nt.

Microdelección	Primers (5' →3')
c.-88del30	TATTTGCTCTCAGAAGTCGGTTTCCAGCGTTGCTTCGGTTTCCCTCCAGA
	TCTGGAGGGAAACCGAAGCAACGCTGGAAACCGACTTCTGAGAGCAAATA
c.-63del30	TTTCCTTTTCTTCAGTGAATCGGAGCAGGCAAGTCGGAGGGAAAGTCCACG
	CGTGGACTTCCCTCCGACTTGCCTGCTCCGATTCACTGAAGAAAAGGAAA
c.-63del16	TTTCCTTTTCTTCAGTGAATCGGAGGGTTTCCCTCCAGACAGGCAAGTCG
	CGACTTGCCTGTCTGGAGGGAAACCTCCGATTCACTGAAGAAAAGGAAA
c.-49del16	GTGAATCGGAGCTCAGAGCGTTGCTCAGGCAAGTCGGAGGGAAAGTCCACG
	CGTGGACTTCCCTCCGACTTGCCTGAGCAACGCTCTGAGCTCCGATTCAC
c.-28del30	TGACATTTTCATTTCTGTTGTAGTTGTCTTCAATGTGGTCTGAATATACA
	TGTATATTCAGACCACATTGAAGACAACACTACAACAGAAATGAAATGTCA
c.-3del30	AGGAATCTGAAATAAACAGGAAAGCATTGGTGGGGTGAAGATTTACTTTC
	GAAAGTAAATCTTCACCCACCAATGCTTTCCTGTTTATTTTCAGATTCCT

2. FIGURAS

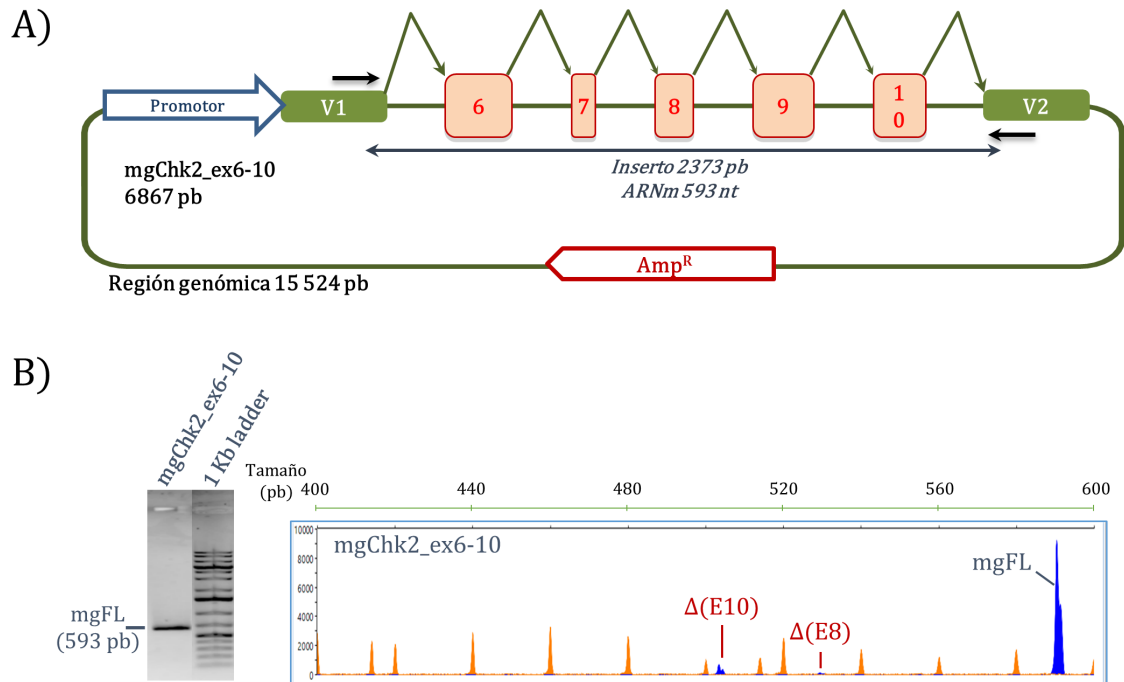


Figura Suplementaria 1. Estructura y validación funcional del minigén mgChk2_ex6 10. A) Representación gráfica del minigén mgChk2_ex6 10. Los exones están dentro de recuadros de color rojizo; las flechas negras sobre los exones V1 y V2 indican los primers de RT-PCR específicos del vector. B) Validación funcional del minigén: los productos RT-PCR se analizaron mediante electroforesis en gel de agarosa (izquierda) y electroforesis capilar de fragmentos fluorescentes (derecha). En el electroferograma del análisis de fragmentos fluorescentes, los picos azules se corresponden con los transcritos producidos por el minigén, mientras que los naranjas representan el patrón de tamaño LIZ600. Por otro lado, el eje x refleja el tamaño en pb y el eje y representa las RFU.

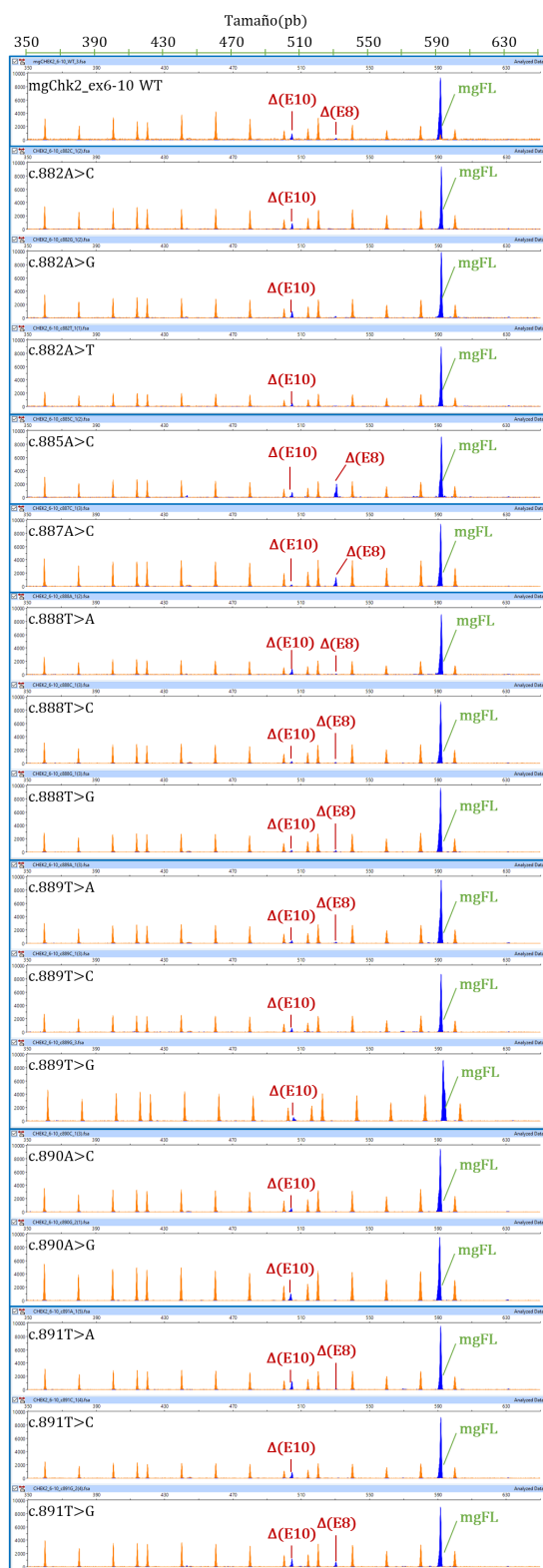
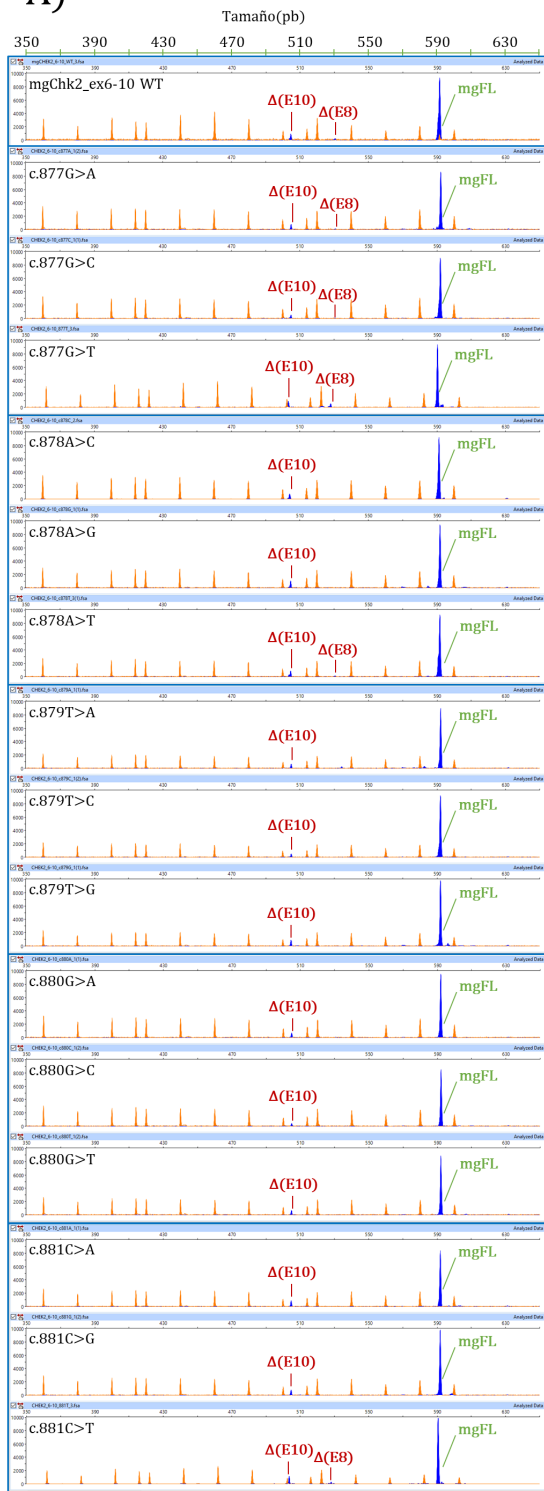
EcoRI

GAATTCggctacagtaagacccatgtttgattgctgggctcagaaatggaatttggagctgacttagtgttt
ttggctggttaagcccttgacattttactggtgaagtttactgtgctcctctgcaaacagggacaatggaaa
ggctgttatatggaatgtctgtaaatcttaacttgaagtggaccaggagtggttaggtctcataatataaa
acattcagacaatcactatctttgttttccctctagTGGTGCCTGTGGAGAGGTAAGCTGGCTTCGAGAG Ex6
GAAACATGTAAGAAAGTAGCCATAAAGATCATCAGCAAAAGGAAGTTTGCTATTGGTTCAGCAAGAGAGGCA
gtaagtaccaataaaaaggctgatcaccttttattaacacgtcttcataacttcccaaaataatcttagattga
tgatttttaaaactcaggaataaacccctgcttagatggatgaatttgtgagatcaaaagaacatggaaaattt
tggggcatctttgattaatctagaacttttactttagataatttgggtccgcttactatttgccaagcatt
ttacaagttattttagtgaggattttaagtcattgacctgggctttgatttctttgtataagtagagctgg
gtttggaactcaggcagccttgagtcactgagtttaactgtaaatgttttttaaatattggcaattattat
tactaataaactctgttattctgtttatcaaagGACCCAGCTCTCAATGTTGAAACAGAAATAGAAATTTTGA Ex7
AAAAGCTAAATCATgtaagtattattatagctcacttagatttatctatctcaaatatgagaagagta
taaagcctttcagtataataaagataaatatggccaggtgtggtggctgatgcctgtaaccctagcatttgg
gagccgaggtgggtgaatcacctgaggtcaggagtttgagaccagcctggccaacatggtgaaacccaaatt
tcttgcattgttctcactgaggagttggctagattctgtctgccttctccccctgattgaaacttccaagcctc
aggcagggggccagatgcctgttcttggcctccctgtccctctgggcagatgttctaaagctcttgggtttc
ctcttggagtttctcactactttcccttttttctccccctcttcttagCCTTGCATCATCAAGATTAAAAAC Ex8
TTTTTGTGTCAGAAAGATTATTATATTGTTTTGGAATTgtaagtagtaaaactttttaaataattctgtttcca
ttcaaatgctttataaagccaccaagaatctaagttaggcttgcctttctcatctggggggcctcaacgta
tgtatgtagtagtggggccacaaagtctgtgtataggggttaatatcaaccaaaaaacctccaatgttt
ccaaaaataggtggatgttttccaactgggacatctgattgtgaataaaccttattagttttaaaaatgaga
gtcaaaagtgaataaaaaatgtcaactgtttgctgtcttaataatgatttggtttggttttatgtcttctgt
ccaagtgcgttttccccaggaatgaacccttgccttgccttgggttgcagGATGGAAGGGGGAGAGCTGT Ex9
TTGACAAAGTGGTGGGAATAAACGCCTGAAAGAAGCTACTGCAAGCTCTATTTTTACCAGATGCTCTTGGC
TGTGCGAGgtaaaagaagccctcaattcttttgacagctgcctttcttttagttgaattactcctgaattagg
cagaaatcaggggaccctggattaaaactcagaagagaatctgtgtgttggctgttcttcttggcaaaaat
gtcctactccttctcactgagatgaaaaagctctcacttcatttccctgctttttaaagatcttctcaacaaa
tttgtaaaccttaccccaacagtttctatgattttggagaataatggttgtgtgtatcacggcttacggttt
accagctgtgagatgtgtgtgttgtaactttgattctacttaagtatctactgcatgaatctgagggatcat
gcctgcctttctgtgtagTACCTTCATGAAAACGGTATTATACCCGTGACTTAAAGCCAGAGAATGTTTTAC Ex10
TGTCATCTCAAGAAGAGGACTGTCTTATAAAGgtaagaaattggcattctgtatgtggctattcctgtagatt
cttgttcagaaacttttaattcctcaataaaaagtgaggagcgtaaaaggatgagaatctgacctttcact
gttaccattattttagagatcaggtgatactgctcttatccaAAGCTT

HindIII

Figura Suplementaria 2. Secuencia del inserto del minigén mgChk2_ex6-10. Los exones se indican en naranja y en mayúsculas, los intrones en minúsculas y en negro y los sitios de clonación (*EcoRI* e *HindIII*) están subrayados y en mayúsculas. Estructura del inserto (2373 pb): *EcoRI* - *ivs5* (250 pb) - Ex6 (109 pb) - *ivs6-1* (198 pb) // *ivs6-2* (200 pb) - Ex7 (54 pb) - *ivs7-1* (200 pb) // *ivs7-2* (200 pb) - Ex8 (62 pb) - *ivs8-1* (200 pb) // *ivs8-2* (180 pb) - Ex9 (100 pb) - *ivs9-1* (176 pb) // *ivs9-2* (200 pb) - Ex10 (87 pb) - *ivs10-1* (157 pb) - *HindIII* (los intrones acortados se indican con una doble barra //).

A)



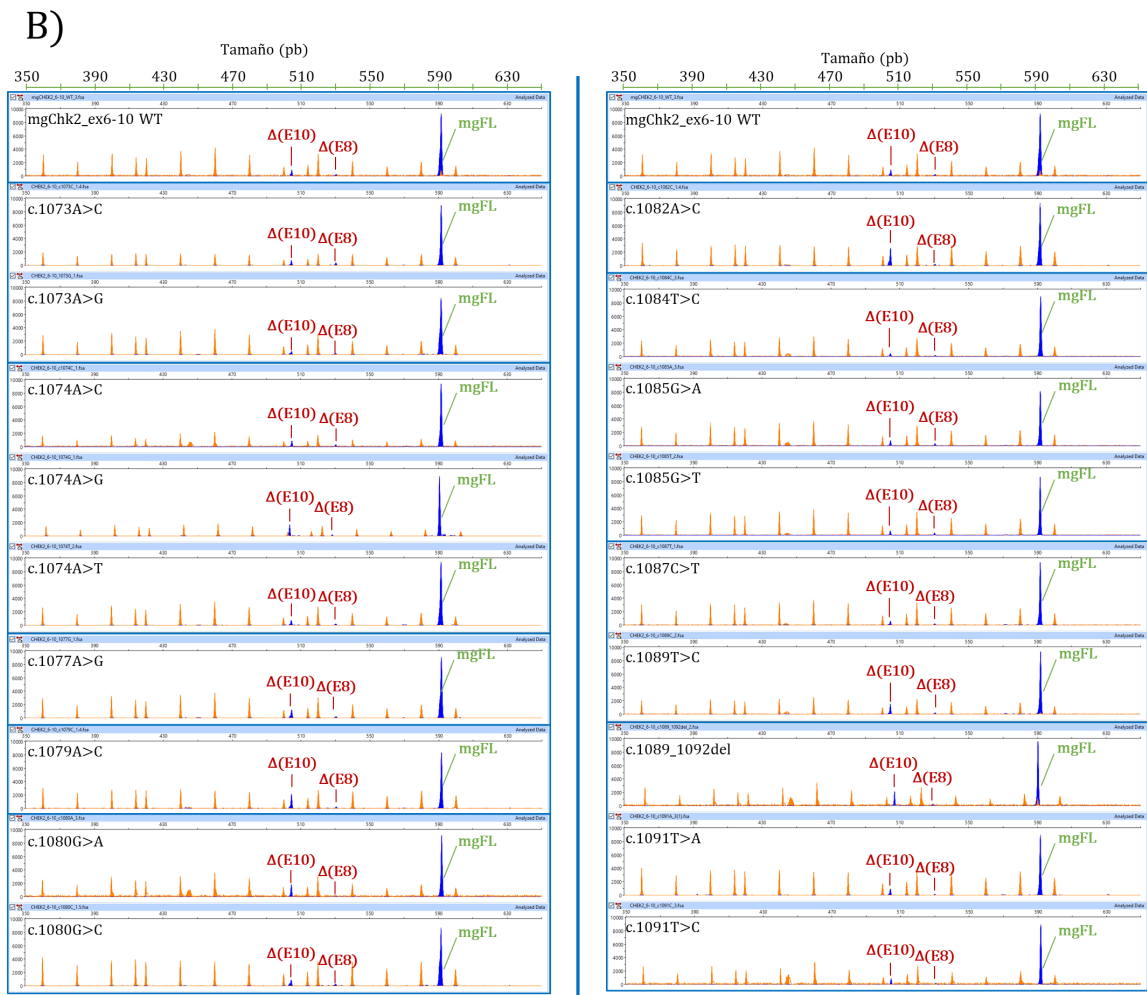
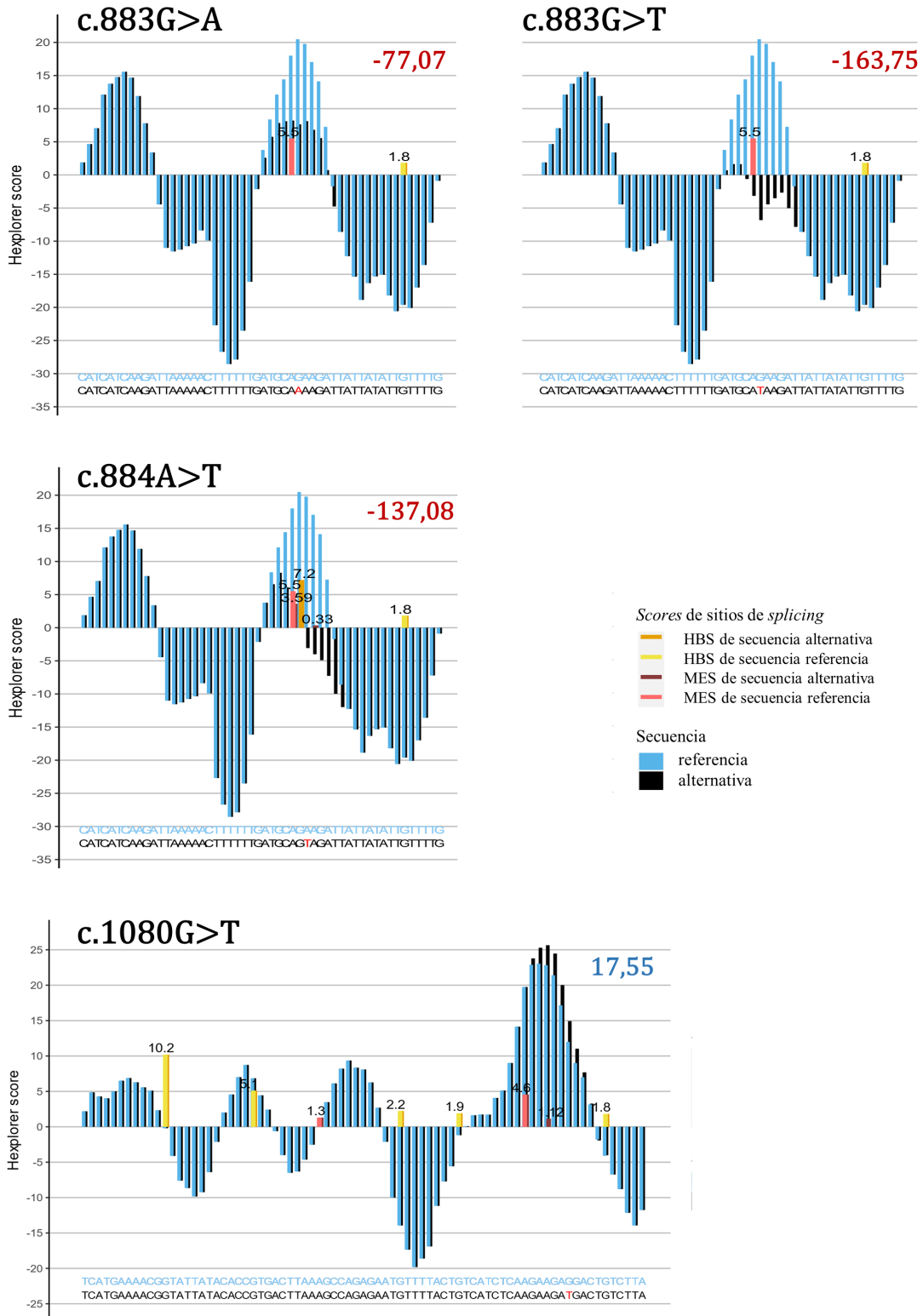
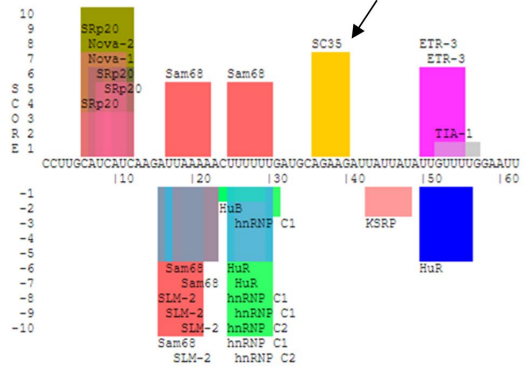


Figura Suplementaria 3. Análisis de fragmentos fluorescentes de las variantes no espliceogénicas de los exones A) 8 y B) 10 de *CHEK2*. Los productos marcados con FAM (picos azules) se analizaron con LIZ600 (picos naranjas) como estándar de tamaño. El eje x indica el tamaño en pb (electroferograma superior) y el eje y representa las RFU.

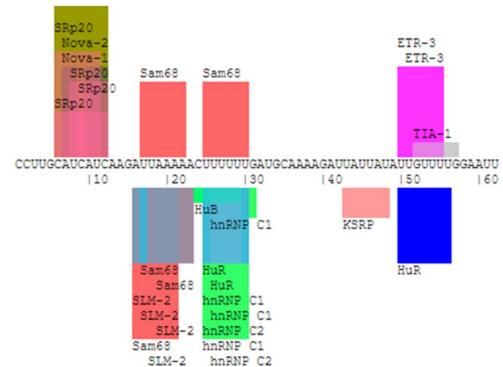
A)



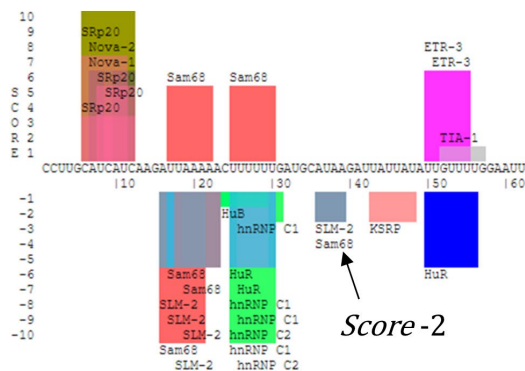
B)
Exon 8 WT



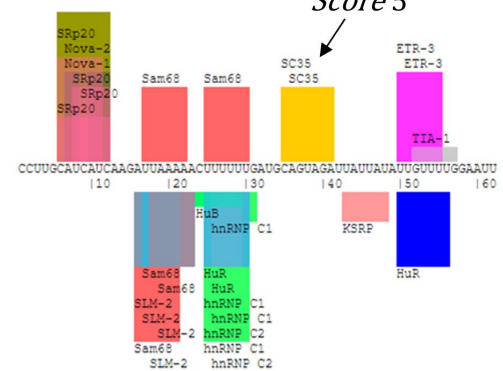
c.883G>A



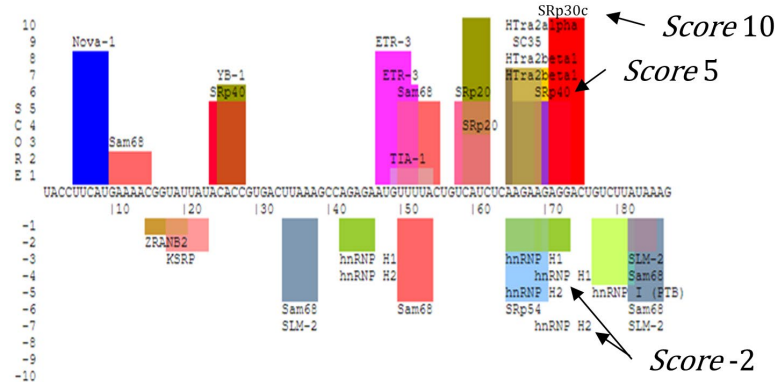
c.883G>T



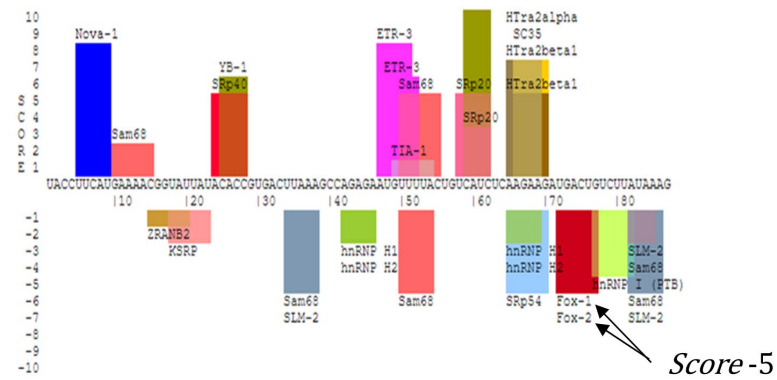
c.884A>T



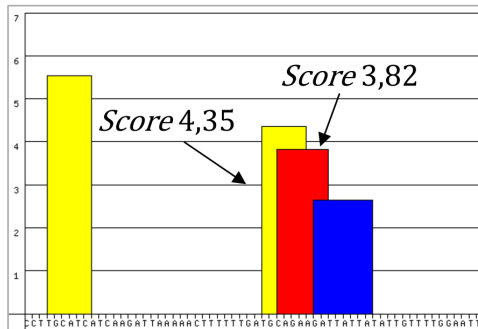
Exon 10 WT



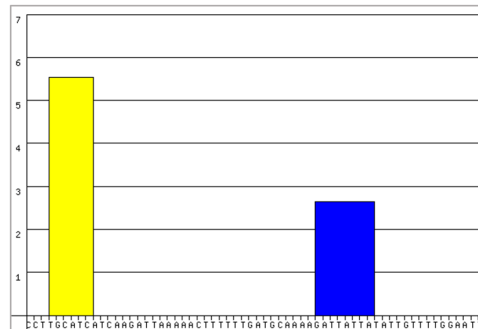
c.1080G>T



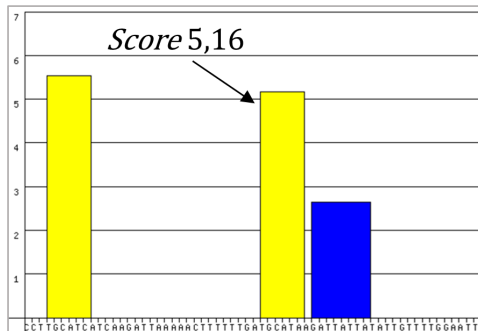
C)
Exon 8 WT



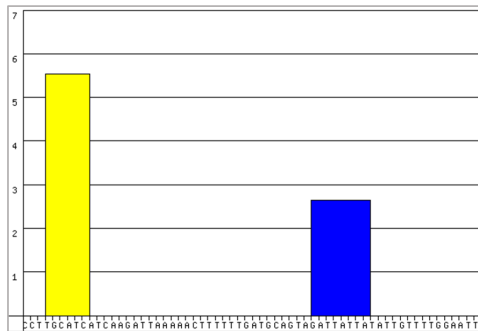
c.883G>A



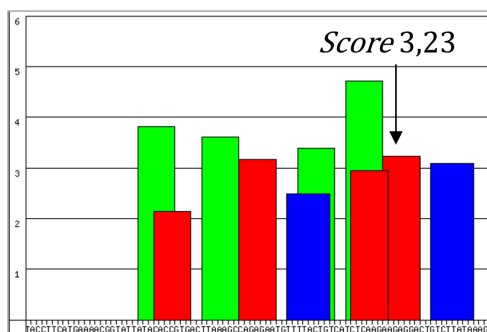
c.883G>T



c.884A>T



Exon 10 WT



c.1080G>T

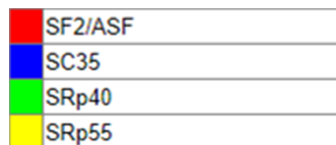
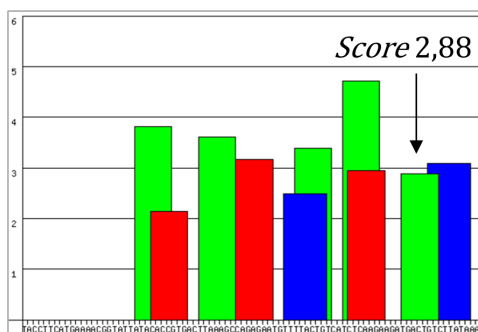


Figura Suplementaria 4. Análisis bioinformático de las variantes espliceogénicas de los exones 8 y 10 de *CHEK2*. Análisis empleando las herramientas bioinformáticas A) HEXplorer, B) SpliceAid y C) ESEFinder de las variantes exónicas altamente espliceogénicas, c.883G>A, c.883G>T, c.884A>T y c.1080G>T, así como de las secuencias WT de los exones 8 y 10 de *CHEK2*. En el caso de SpliceAid y ESEFinder, se indican las puntuaciones de las proteínas de unión al ARN que se ven alteradas en presencia de las distintas mutaciones.

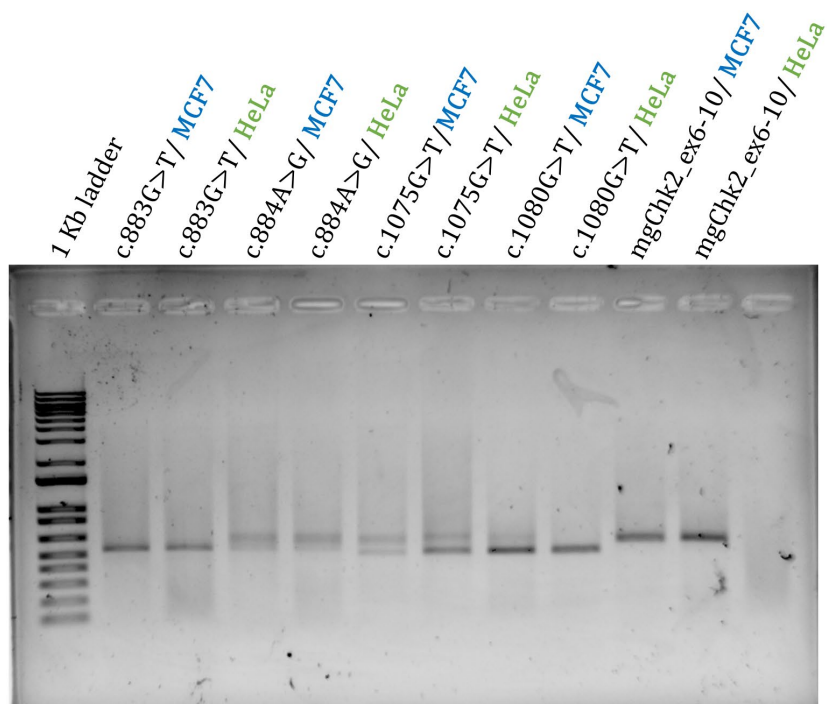
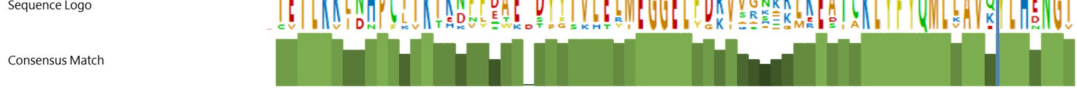
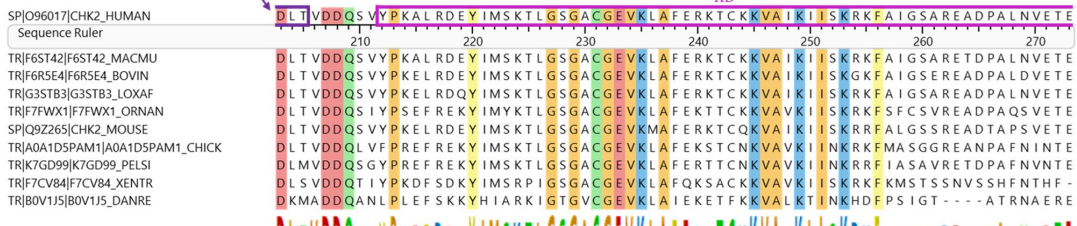
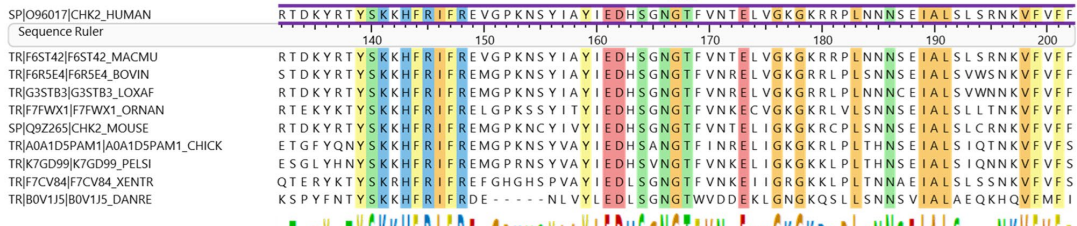
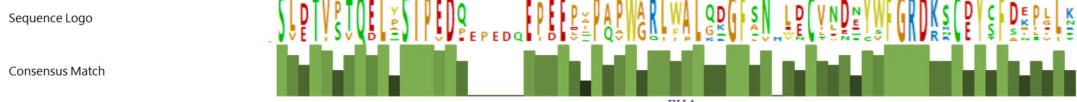
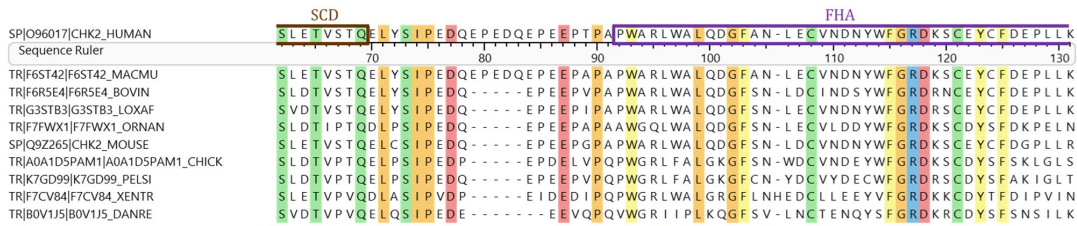
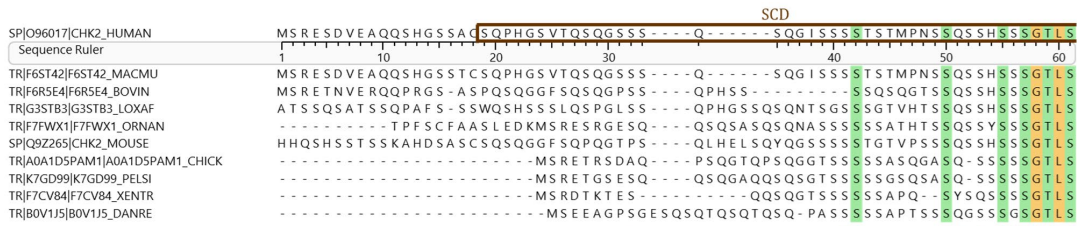


Figura Suplementaria 5. Reproducibilidad de los resultados de *splicing* del minigén *mgChk2_ex6-10*. Electroforesis en gel de agarosa (1,0%) de los resultados de *splicing* de cuatro variantes ubicadas en los exones 8 (c.883G>T y c.884G>T) y 10 (c.1075G>T y c.1080G>T) de *CHEK2* testadas en células HeLa y MCF-7.



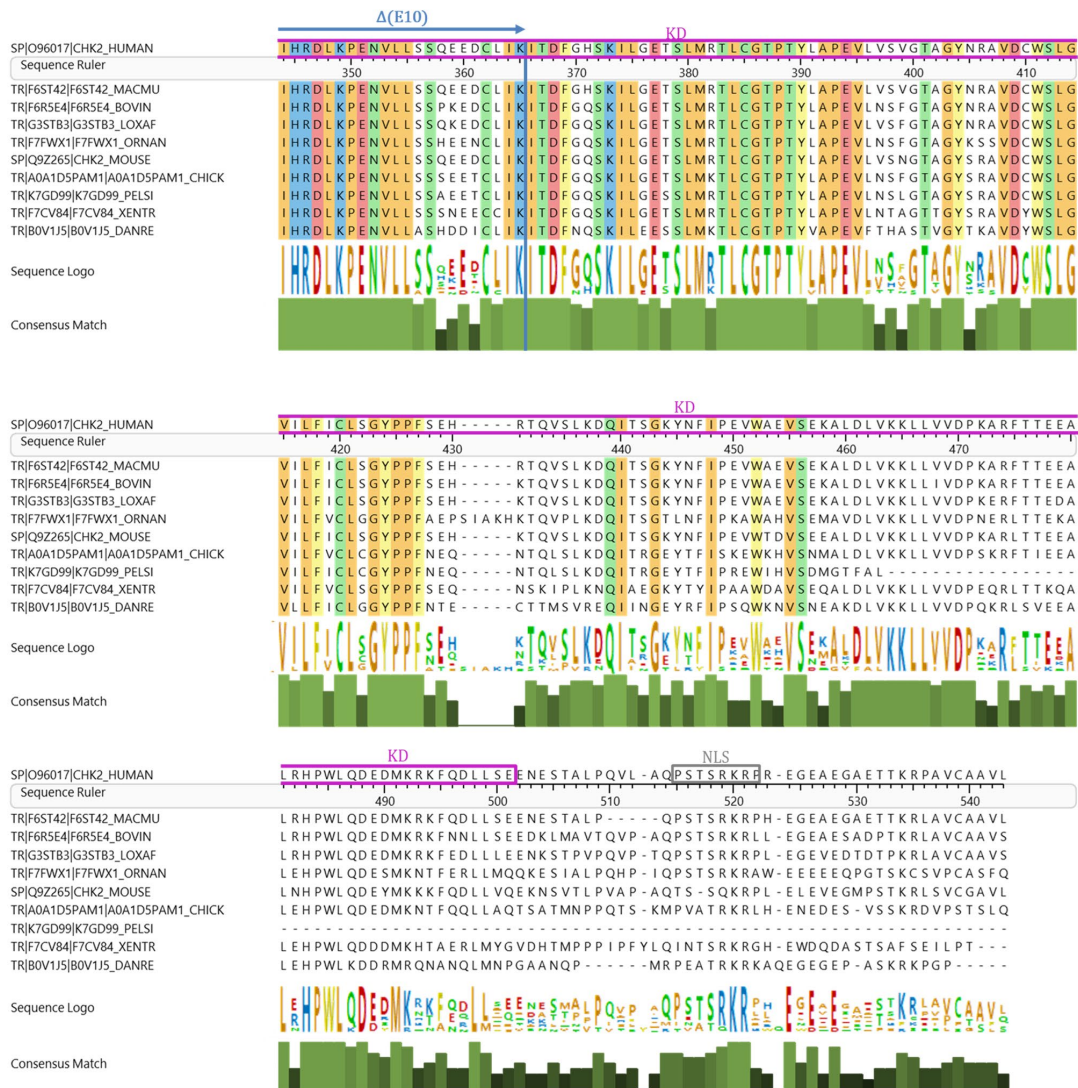


Figura Suplementaria 6. Conservación de la secuencia proteica de CHK2 en vertebrados. Conservación de los aminoácidos ubicados en la región deletada por el transcrito *in-frame* $\Delta(E10)$. Dicha región está marcada con rayas y flechas azules, y los dominios funcionales clave de CHK2 basados en los datos publicados por Stolarova y colaboradores en el año 2020 se indican como sigue: el dominio SCD se muestra en un recuadro marrón, el dominio FHA se indica en un recuadro morado, el dominio KD se indica en un recuadro fucsia y la señal NLS se muestra en un recuadro gris. Las secuencias proteicas se alinearon con la herramienta Align de la base de datos Uniprot (<https://www.uniprot.org/align/>). El archivo de alineación se visualizó con MegAlign Pro versión 15.0.0 del *software* Lasergene de DNASTAR. Se resaltan los residuos conservados (aminoácidos no polares en naranja, aminoácidos polares en verde, aminoácidos básicos polares en azul y aminoácidos ácidos polares en rojo), y la intensidad del color de las barras verdes indica el grado de conservación de cada aminoácido. Organismos: humano (*Homo sapiens*); macaco Rhesus (*Macaca mulatta*: MACMU); elefante africano (*Loxodonta africana*: LOXAF); vaca (*Bos taurus*: BOVIN); ornitorrinco (*Ornithorhynchus anatinus*: ORNAN); ratón (*Mus musculus*: MOUSE); gallina (*Gallus gallus*: CHICK); tortuga china de caparazón blando (*Pelodiscus sinensis*: PELSI); rana africana (*Xenopus tropicalis*: XENTR); pez cebra (*Danio rerio*: DANRE).

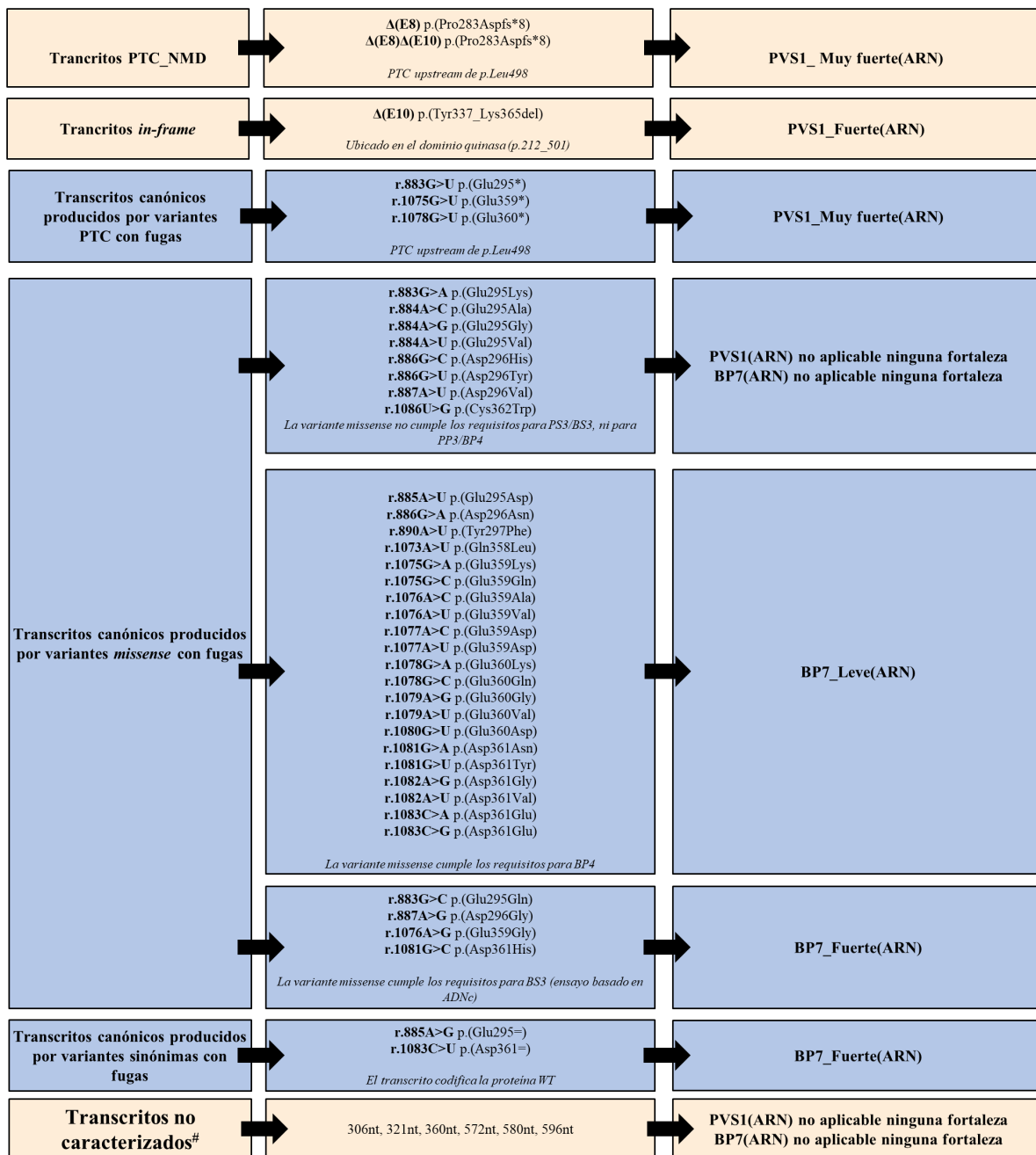


Figura Suplementaria 7A. Fortalezas de código patogénico/benigno aplicables a transcritos individuales del gen *CHEK2*. Hemos adaptado el razonamiento del árbol de decisión PVS1 propuesto por el ClinGen SVI [235] con el propósito específico de determinar la fortaleza de las evidencias patogénicas o benignas aportadas por los 41 transcritos del *mgChk2_ex6-10* únicos (incluyendo transcritos canónicos portadores de variantes PTC, *missense* o sinónimas, como los observados en variantes espliceogénicas con fugas, resaltadas en oscuro). # En aras de la exhaustividad, hemos añadido al diagrama de flujo los 6 transcritos *mgChk2_ex6-10* no caracterizados, asumiendo que los transcritos *mgChk2_ex6-10* no caracterizados de idéntico tamaño son realmente idénticos (en cualquier caso, no afecta al proceso de clasificación, como se muestra en la **Figura suplementaria 7B**).

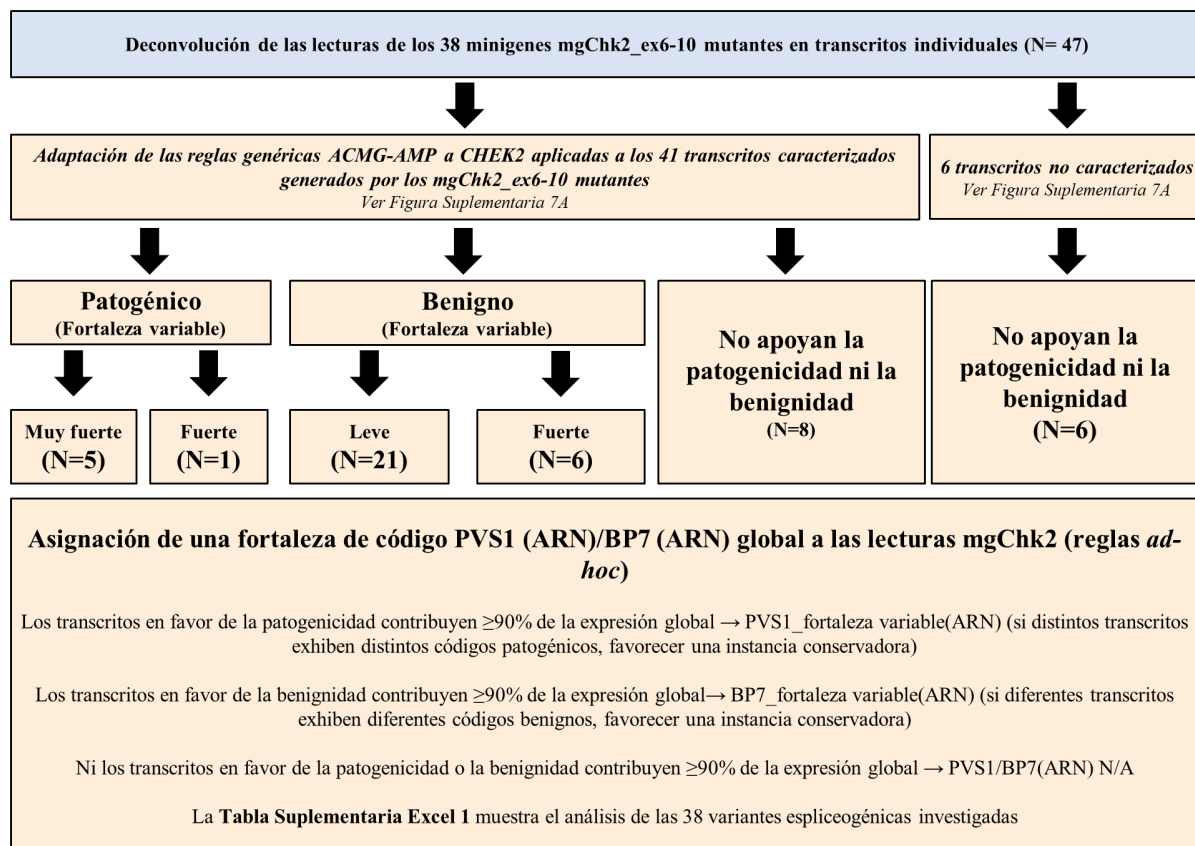


Figura Suplementaria 7B. Árbol de decisión propuesto, asignando un código PVS1_O/BP7_O a las lecturas de los minigenes mutantes de CHEK2. La mayoría de las variantes genéticas de *CHEK2* analizadas en este estudio producen dos transcritos aberrantes, y/o un componente mgFL residual sustancial (variantes *leaky*). Para abordar esta cuestión, proponemos el siguiente enfoque: (i) deconvolución de la señal de lectura en componentes individuales (transcritos), (ii) asignación de una intensidad de código patogénica o benigna a cada transcrito individual (de acuerdo con las directrices ACMG-AMP, véase la **Figura Suplementaria 7A** para más detalles), y (iii) aplicación de reglas combinatorias *ad hoc* (teniendo en cuenta la contribución relativa de las intensidades de código patogénicas y benignas a la expresión global) para producir una intensidad de código PVS1_O/BP7_O global. Obsérvese que los transcritos no caracterizados (detectados en 9 variantes) se han incorporado al esquema como transcritos que no apoyan ni la patogenicidad ni la benignidad. Estos transcritos no caracterizados representan $<5\%$ de la señal de lectura global en todos los casos. Sin embargo, en el caso hipotético de que las variantes produjeran transcritos no caracterizados que representen $>10\%$ de la señal de lectura global, el árbol de decisión propuesto terminaría anotando las lecturas de dichos minigenes como PVS1/BP7(ARN) N/A, incluso si el 100% de la señal caracterizada da su apoyo a la patogenicidad (o benignidad).

A)

SacI

CCGCGGcaggctaagtaaagtctcttcatagcaaattacatgagcaaccttagatttgatgtatgattttac
tctttaaaacagtattcaacaaggatattacaattgaccattgtatgttagaataacctctgctccatttatt
tctgttcaaacgttttagttttggaattaaattctgctgaatgggttgcttttttttttttttaattattt
taaagTAATTGTGTAAGTGACTGCATTGGAAGTGTCCAGTGTGTACACCCGGCCTGGATACAAGACT
TGAGATAAATAGACAACTGGACAGCATGATTCAACTTTGTAGTAAGCTTCGAAATTTGCTACATGACAATGA
GCTGTCAgtaagaactatccctatctctcttagttaaattcatcagttaaaaactgatgaattcatattcat
aaagtataaaaacatctatctggagttctggaatacgtatttcagattttaaaatctgtaggttttttttt
tttttttaaatagccattgagtctctctatgttgtccaagccggacttgaactctgcactcaagggattccc
cccacctcagcctccctggatcatgtgacacaccgggtggttttttggaagatca)

Ex3

*Bam*HI

GGATCCatttggaaatataagtttcatgttttctctgctatattcctcagtccttaggatatggcacaaggt
cggattcagtaagattgaatgagtgaaatctgctatcaaaaggttcacactctaggagctgagaaagaagt
acataattaaagatgatacactttaggggaactgtaaacaaaattctcgggagctccatgtgggagcaata
aatttcatgtaacagatttctttttctttttctgtcagATTGAAAGAAGATAAACCTAGGAAAAGTTTGT
TTAATGATGCAGGAAACAAGAAGAAATCAATTAATAATGTGGTTTAGCCCTCGAAGTAAGAAAGTCAGATATGT
TGTGAGTAAAGCTTCAGTGCAAACCCAGCCTGCAATAAAAAAAGATGCAAGTGCTCAGCAAGACTCATATGAA
TTTGTTTCCCCAAGTCTCTGCGAGATGTTCTGAGAGGGCTAAAAAGGCTTCGCAAGATCGAAAAAAGC
AAAAAAGAAAACTTTAGCTGAAATCAACCAAAAAATGGAATTTAGAGGCAGAAAAAGAAGATGGTGAATTTGA
CTCCAAAGAGGAATCTAAGCAAAGCTGGTATCCTTCTGTAGCCAACCATCTGTTATCTCCAGTCTCAGATA
AATGGTGAATAGACTTACTAGCAAGTGGCTCCTTGACAGAATCTGAATGTTTTGGAAGTTTAACTGAAGTCT
CTTTACCATTGGCTGAGCAAATAGAGTCTCCAGACACTAAGAGCAGGAATGAAGTAGTGACTCCTGAGAAGGT
CTGCAAAAATTATCTTACATCTAAGAAATCTTTGCCATTAGAAAAATAATGGAAAACGTGGCCATCACAAATAGA
CTTTCCAGTCCCATTCTAAGAGATGTAGAACCAGCATTCTGAGCACCAGTGGAGATTTTGTAAAGCAAACCG
TGCCCTCAGAAAAATACCATTGCCTGAATGTTCTTACCACCTTCATGCAAAACGTAAAGTTGGTGGTACATC
AGGGAGGAAAAACAGTAAACATGTCCGATGAATTCATTAGTCTTTACCAGGTACACCACCTTCTACATTAAGT
AGTTCAAGTTACAGGCGAGTGATGTCTAGTCCCTCAGCAATGAAGCTGTTGCCAATATGGCTGTAAAAAGAA
ATCATAGAGGAGAGACTTTGCTCCATATTGCTTCTATTAAGgtaggatgcttactctgaaataccatctcaga
atgaggccaactataaagcaatttctttgagttttgaaaaatggcataggattactaggataattaacctt
tcacagacatgatacttctctgaaccagagaagccagattcatagggagagcatcttacttcagttggagc
agtggcccctgagtcctgggcgcatgatctttaggagaaaaccaatattgaaatattcggcttttattttgc
caagtgttttgcctttgtctattttaccttcagttttatcattttgtGGTACC

Ex4

*Kpn*I

(taataatttattaactcatgaacagtagcctaagagaaaacgatttaagttttactttatattgagaaggc
agcatttaaaaagctcaatattttctttctttcttaatgcttttttaatttccgttttgttcatttttcta
GGGGACATACCTTCTGTTGAATACCTTTTACAAAATGGAAGTGTCCAAATGTTAAAGACCATGCTGGATGG
ACACCATTGtagttgtctggtttttattctcattctttctgtgttttacagttcttatagtttatagttatg
tagttgtctatatatcatcctctgccacataactcttttagtctgaagaacttatgttttcatcaagtatg
agaacatgattactttcttctagcttttcattttgtgacaggcaagaaattggttaccctttg

Ex5

B)

SacII

CCGCGGgaggctaagtaaagtctcttcatagcaaattacatgagcaaccttagatttgatgtatgtattttac
tctttaaaacagatttcaacaaggatattacaattgaccattgtatggttagaataacctctgctccatttatt
tctgttcaaacgttttagtttttggaattaaattctgctgaatgggttgcttttttttttttttaattattt
taaagTAATTGTGTAAGTACTGCATTGGAAGTGGATGTCCAGTGTGTTACACCCCGGCTGGATACAAGACT Ex3
TGAAGATAAATAGACAACCTGGACAGCATGATTCAACTTTGTAGTAAGCTTCGAAATTTGCTACATGACAATGA
GCTGTCAGgtaagaactatccctatctctcttagttaaattcatcagttaaaaactgatgaattcatattcat
aaagtataaaaaacatctatctggagttcttggaaatcgtatttcagattttaaatcatctgctatcaaagag
ttcacactctaggagctgagaaagaagtacataattaaagaatgatacacttttaggggaactgtaaacaaaat
tcttcgggagctccatgtgggagcaataaatttcattgtaaacagatttctttttctttttctgtcagATTTG
AAAGAAGATAAACCTAGGAAAAGTTTGTAAATGATGCAGGAAAACAAGAAGAAATTCATTAATAATGTGGTTA
GCCCTCGAAGTAAGAAAGTCAGATATGTTGTGAGTAAAGCTTCAGTGCAAACCCAGCCTGCAATAAAAAAAGA
TGCAAGTGCTCAGCAAGACTCATATGAATTTGTTCCCAAGTCTCCTGCAGATGTTTCTGAGAGGGCTAAA
AAGGCTTCTGCAAGATCTGGAAAAAGCAAAAAAGAAAACCTTAGCTGAAATCAACCAAAAAATGGAATTTAG
AGGCAGAAAAAGAGATGGTGAATTTGACTCAAAGAGGAATCTAAGCAAAGCTGGTATCTCTTGAGCCCA
ACCATTGTTTCTCCAGTCTCAGATAAATGGTGAATAGACTTACTAGCAAAGTGGCTCTTGACAGAATCT
GAATGTTTTGGAAGTTAACTGAAGTCTCTTTACCATTGGCTGAGCAAATAGAGTCTCCAGACACTAAGAGCA Ex4
GGAATGAAGTAGTACTCCTGAGAAGGTCTGCAAAAAATATCTTACATCTAAGAAATCTTTGCCATTAGAAAA
TAATGGAAAACGTGGCCATCACAATAGACTTTCCAGTCCCATTCTAAGAGATGTAGAACCAGCATTCTGAGC
ACCAGTGGAGATTTTGTAAAGCAAACCGTCCCTCAGAAAATATACCATTGCCTGAATGTTCTTACCACCTT
CATGCAACAGTAAAGTTGGTACATCAGGGAGGAAAAACAGTAACATGTCGATGAATTCATTAGTCTTTT
ACCAGTACACCACCTTCTACATTAAGTAGTTCAAGTTACAGGCGAGTGTCTAGTCCCTCAGCAATGAAG
CTGTTGCCAATATGGCTGTGAAAAGAAATCATAGAGGAGAGACTTTGCTCCATATTGCTTCTATTAAGgtag
gatgcttactctgaaataccatctcagaatgaggccaactataaagcaatttctttgagttttgaaaaatg
gcataggattactaggataattaacctttcacagacatgatacttctctgtaaccagagagaagccagattcata
gggagagcatcttacttctcagttggagcagtgccctgagctggtggcagatgcttgttaggagaaaaccaa
tatttgaatatttcggttttattttgccaagtgcctttgctttgtctattttaccttcagttttatcatt
ttgttaataattttaactcatgaacagtagccttaagagaaaacgatttaagttttactttatattgaaga
aggcagcatttaaaaagctcaatattttctttctttccttaagcttttttaatttccgtttgttcatttt
tctagGGCGACATACCTTCTGTTGAATACCTTTACAAAATGGAAGTGATCCAAATGTTAAAGACCATGCTGG Ex5
ATGGACACCATTTgtagttgtctggtttttatttctatttcttctgtgttttacagttcttatagtttatagt
tatgtagttgtctatataatcatctctgcccacataactcttttagtctgaagaacttatgttttcatcaag
tatgagaacatgattacttctcttagcttttcatttgtgacaggcaagaattggttaccttttGGTACC
KpnI

Figura Suplementaria 8. Secuencia de los insertos empleados para construir los distintos minigenes de *BARD1*. Los exones se indican en morado y en mayúsculas, los intrones en minúsculas y en negro y los sitios de clonación (*Bam*HI, *Kpn*I, *Sac*II) están subrayados y en mayúsculas.

A) Los recuadros de distintos colores engloban los insertos necesarios para confeccionar los distintos minigenes iniciales de *BARD1*: azul→ mgBARD1_ex4, verde→ mgBARD1_ex4-5 y amarillo → mgBARD1_ex3-5.

- Estructura del inserto incorporado en el mgBARD1_ex4 (1503 pb): *Bam*HI – ivs3 (253 pb) – Ex4 (950 pb) – ivs4 (300 pb)– *Kpn*I
- Estructura del inserto que configura el mgBARD1_ex4-5 (1931 pb): *Bam*HI – ivs3 (253 pb) – Ex4 (950 pb) – ivs4.1 (300 pb)//ivs4.2 (147 pb) – Ex5 (81 pb) – ivs5.1(200 pb) – *Kpn*I (los intrones acertados se indican con una doble barra //)
- Estructura del inserto que configura el mgBARD1_ex3-5 (2574 pb): *Sac*II – ivs2 (218 pb) – Ex3 (149 pb) – ivs3.1 (270 pb) – *Bam*HI – ivs3.2 (253 pb) – Ex4 (950 pb) – ivs4.1 (300 pb)//ivs4.2 (147 pb) – Ex5 (81 pb) – ivs5.1(200 pb) – *Kpn*I (los intrones acertados se indican con una doble barra //)

B) Estructura del inserto que conforma el mgBARD1_ex3-5 final estable (2324 pb): *Sac*II – ivs2 (218 pb) – Ex3 (149 pb) – ivs3.1 (122 pb)//ivs3.2 (157 pb) – Ex4 (950 pb) – ivs4.1 (300 pb)//ivs4.2 (147 pb) – Ex5 (81 pb) – ivs5.1(200 pb) – *Kpn*I (los intrones acertados se indican con una doble barra //)

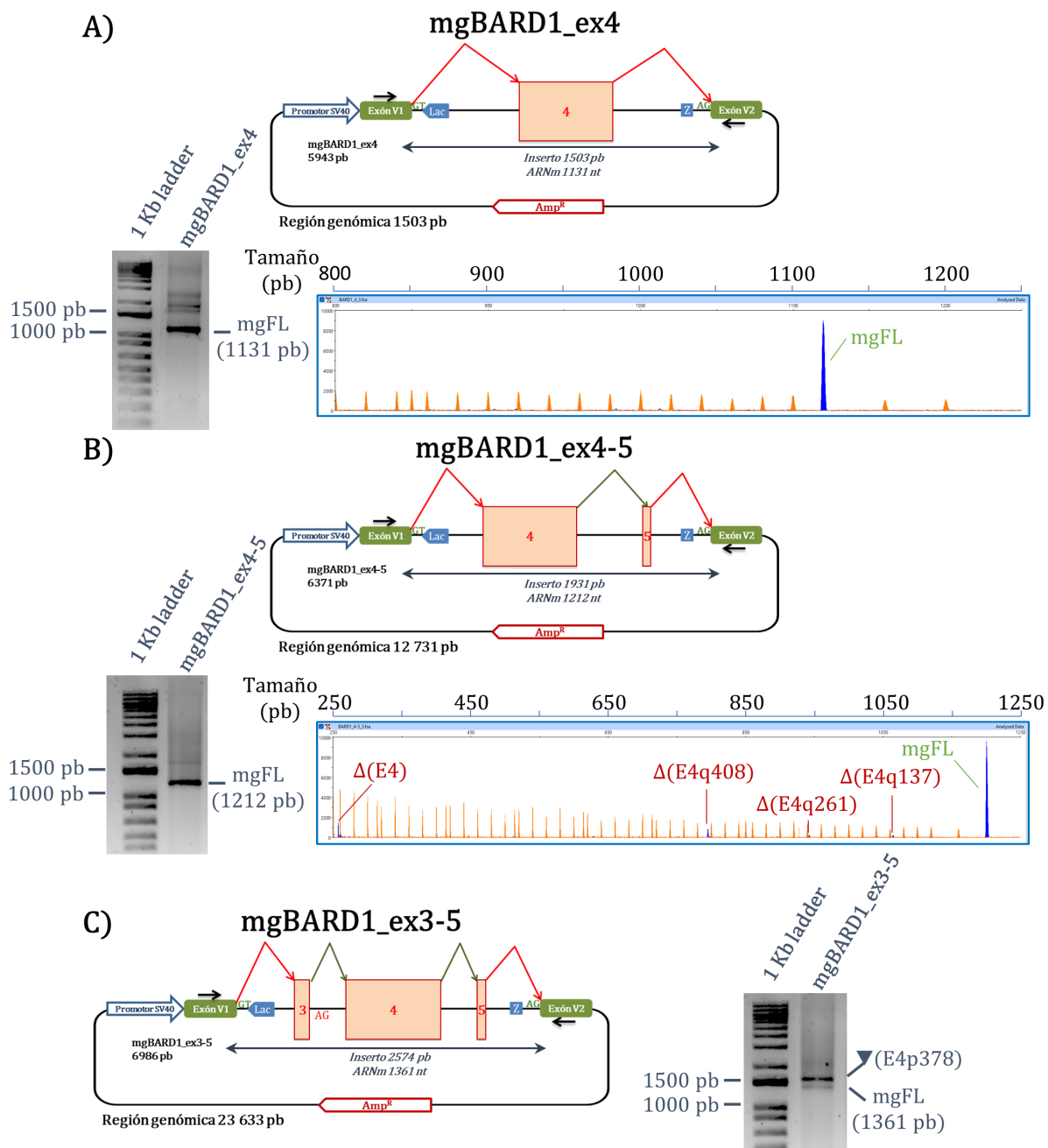


Figura Suplementaria 9. Estructura y validación funcional de los minigenes intermediarios de *BARD1*. Representación gráfica de los minigenes intermediarios de *BARD1* A) mgBARD1_ex4, B) mgBARD1_ex4-5 y C) mgBARD1_ex3-5. Los exones están dentro de recuadros de color rojizo; las flechas negras sobre los exones V1 y V2 indican los primers de RT-PCR específicos del vector. En A) y B) debajo de la representación gráfica aparece la validación funcional de los minigenes: los productos RT-PCR se analizaron mediante electroforesis en gel de agarosa (izquierda) y electroforesis capilar de fragmentos fluorescentes (derecha). En el electroferograma del análisis de fragmentos fluorescentes, los picos azules se corresponden con los transcritos producidos por el minigén, mientras que los naranjas representan el patrón de tamaño LIZ1200. El eje x refleja el tamaño en pb y el eje y representa las RFU. En C) solo se visualizaron los transcritos generados por el minigén mgBARD1_ex3-5 por electroforesis en gel de agarosa (derecha de la representación gráfica).

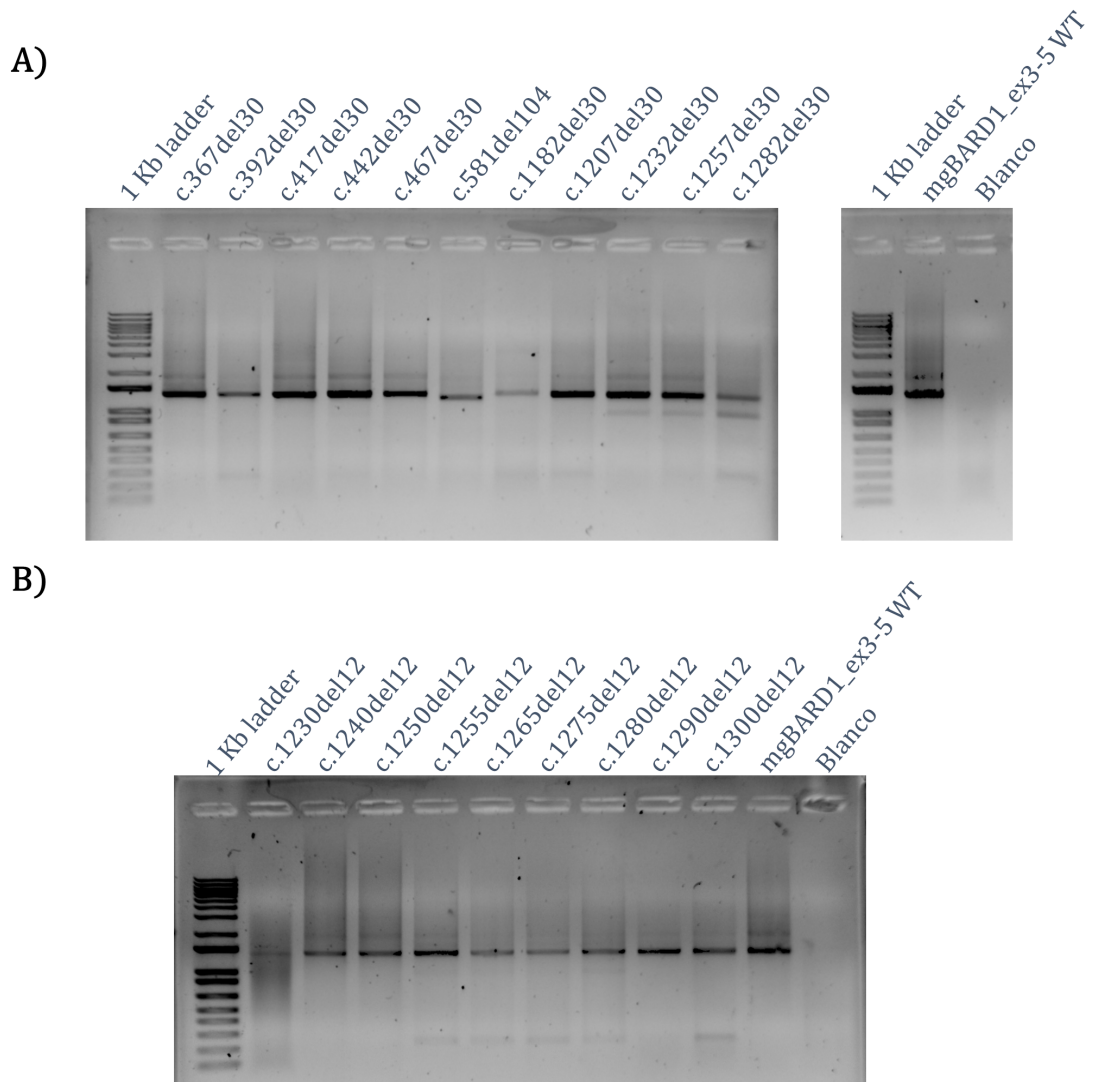
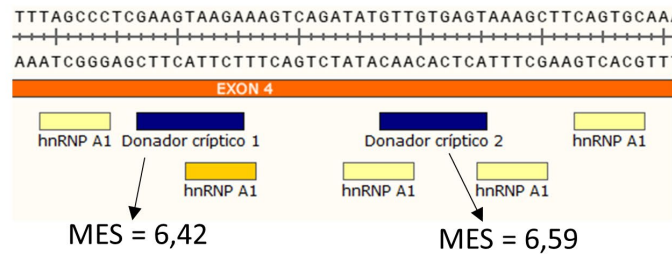
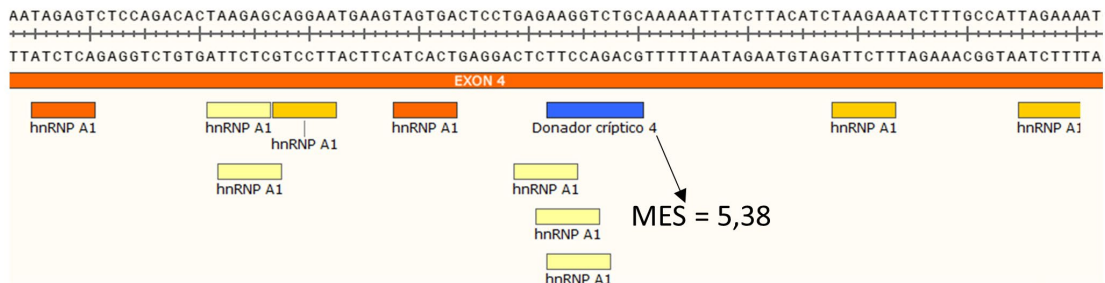


Figura Suplementaria 10. Electroforesis en gel de agarosa de los transcritos generados por las μ del incorporadas en mgBARD1_ex3-5. A) Resultados de las diez μ del de 30 nt incorporadas en las regiones exónicas próximas a los sitios 5'ss y 3'ss y de una microdelección de mayor tamaño, c.581del104del, que elimina una región rica en motivos putativos de unión del *enhancer* Tra2 β (para más información ver la **Tabla Suplementaria 9**). B) Resultados de las nueve μ del de 12 nt ubicadas en las regiones de 30 nt que al ser delecionadas alteraban el proceso de *splicing*.

A)



B)



C)

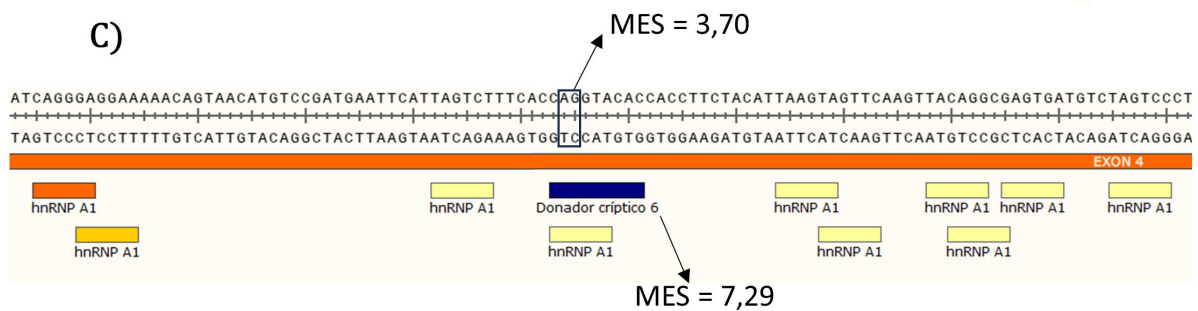


Figura Suplementaria 11. Representación de los donadores crípticos del exón 4 de *BARD1* y los sitios hnRNPA1 predichos que los rodean. En esta imagen vemos los sitios 5'ss crípticos A) 1, 2, B) 4 y C) 6 ubicados 863 (c.449_457), 843 (c.469_477), 408 (c.904_912) y 137 (c.1175_1183) pb *upstream* del donador canónico del exón 4, junto con su puntuación MES y los motivos de unión para el factor de *splicing* hnRNPA1 identificados empleando la herramienta HSF. Estos motivos aparecen en cajas de distintos colores, desde un amarillo claro hasta un naranja fuerte, los cuales concuerdan con su fortaleza (amarillo claro + débil; naranja oscuro + fuerte). Adicionalmente, en el apartado C) remarcamos el sitio aceptor empleado por la maquinaria de *splicing* cuando se destruyen los silenciadores próximos al donador críptico 6.

A)

XbaI

TCTAGATgatttataagactaggaattatTTTTGTTCCATTGTTTTGTTTATACTTTTATAATTCCAGTT
tctgtttccagatttctcccttagttttgggtctctggttttcttgaaggcgtgtcctcaatatttgacatt
ttcatttctgtttagTTAGGAATCTGAAATAAACAGGAAAGCACTATGTCTTCAATGTGGTCTGAATATACA
ATTGGTGGGGTGAAGATTTACTTTCTTATAAAGCTTACCCGTCACAGCTTGCTATGATGAATTCTgtaagta
ttttcagcagtaagtttttattaagatatagatacttatgaccataaagtaacattgagatttgactgca
ttcatattctcagtaaatggagaatattttctgggttcacttggagaatatctaacctctgggaaactactg
attgcagctaaactgtatatatctgtggttaaggcaagaagaacattctgaaaaatgttctcagctcatttg
gaatttctatacttttaatagaaaaacGGATCC

Ex2

B)

BamHI

GGAATTCGTCTCGGGTTGTGTGGTTGAGGGGTCTGGTGGGTCGAGGAAAGGTAACGGCGGCCCCAGTCTGCA
CACAAAGCCGGGAAGTAGCAGCACCCCCAGGAAGAGGGAGGAGGAAGGGCTCGTGCCCTTTCTTCTTCCA
GGGCTCCGCTTTATTTGCTCTCAGAAGTCGGTTTCCTTTCTTTCTTTCAGTGAATCGGAGCTCAGAGCGTTG
CTTCGGTTTCCCTCCAGACAGgcaagtcggagggaagtcacggaaaggccttgagttcttctgcacagcctg
cgggcccattgggtggtgactggatcaccagggctcggggatctcaactggctgggggaggaggagatacgg
ggcagaaaacctgaggaggagcctatccttacctccttgagcgcaggcctgttcttagccgggcgctgcg
ttaattcttgccgggagcagaggggtgtgtttatcctttatgcaatggttgattataagactaggaat
tattttgttccattgtttgtttatacttttattaattccagtttctgtttccagatttctcccttagttt
tgggtctctggttttcttgaaggcgtgtcctcaatatttgacatttctcatttctgtttagTTAGGAATCTG
AAATAAACAGGAAAGCACTATGTCTTCAATGTGGTCTGAATATACAATTGGTGGGGTGAAGATTTACTTTCT
TATAAAGCTTACCCGTCACAGCTTGCTATGATGAATTCTgtaagtatTTTTcagcagtaagtttttattaag
atatagatacttatgaccataaagtaacattgagatttgactgcattcatttctcagtaaatggagaat
ttctgggttcacttggagaatatctaacctctgggaaactactgattgcagctaaactgtatatatctgtg
gttaaggcaagaagaacatttctgaaaaatgttctcagctcatttggatttctatacttttaatagaaaaacG
GATCC

Ex1

Ex2

BamHI

C)

AACGGCGGCCCCAGTCTGCACACAAGGCCGGGAAGTAGCAGCACCCCCAGGAAGAGGGAGGAGGAAGGGCT
CGTGCCCTTTCTTCTTCCAGGGCTCCGCTTTATTTGCTCTCAGAAGTCGGTTTCCTTTCTTTCTTTCAGT
GAATCGGAGCTCAGAGCGTTGCTTCGGTTTCCCTCCAGACAGgcaagtcggagggaagtcacggaaaggcct
tgagttcttctgcacagcctgccccatgggtggtgactggatcaccagggctcggggatctcaactggc
tgggggaggaggagatacggggcagaaaacctgaggaggagcctatccttacctccttgagcgcaggcct
gttcttagccccggcgctgcttaattcttggcgggagcagaggggtgtgtttatcctttatgcaatggttg
tgattataagactaggaattatTTTgtttccattgtttgtttatacttttattaattccagtttctgtt
tccagatttctcccttagtttgggtctctggttttcttgaaggcgtgtcctcaatatttgacatttctcatt
tctgtttagTTAGGAATCTGAAATAAACAGGAAAGCACTATGTCTTCAATGTGGTCTGAATATACAATTGGT
GGGGTGAAGATTTACTTTCTTATAAAGCTTACCCGTCACAGCTTGCTATGATGAATTCTgtaagtatTTTTc
agcagtaagtttttattaagatatagatacttatgaccataaagtaacattgagatttgactgcattcata
ttctcagtaaatggagaatattttctgggttcacttggagaatatctaacctctgggaaactactgattgca
gctaaactgtatatatctgtggttaaggcaagaagaacatttctgaaaaatgttctcagctcatttggattt
ctatacttttaatagaaaaacGGATCC

Ex1

Ex2

BamHI

Figura Suplementaria 12. Secuencia de los insertos empleados para construir los distintos minigenes de BRIP1. Los exones se indican en naranja y en mayúsculas, los intrones en minúsculas y en negro y los sitios de clonación (*XbaI*, *BamHI*) están subrayados y en mayúsculas.

A). Estructura del inserto contenido por el mgBRIP1_ex2 (531 pb): *XbaI* – ivs1 (156 pb) – Ex2 (123 pb) – ivs2 (252 pb) – *BamHI*

B) Estructura del inserto que configura el mgBRIP1_ex1-2 (1021 pb): Ex1 (240 pb) – ivs1.1(250 pb) // ivs1.2 (156 pb) – Ex2 (123 pb) – ivs2 (252 pb) – *BamHI* (los intrones acortados se indican con una doble barra //)

C) Estructura del inserto que conforma el minigén mgBRIP1_ex1-2 estable (969 pb): Ex1 (188 pb) – ivs1.1(250 pb) // ivs1.2 (156 pb) – Ex2 (123 pb) – ivs2 (252 pb) – *BamHI* (los intrones acortados se indican con una doble barra //)

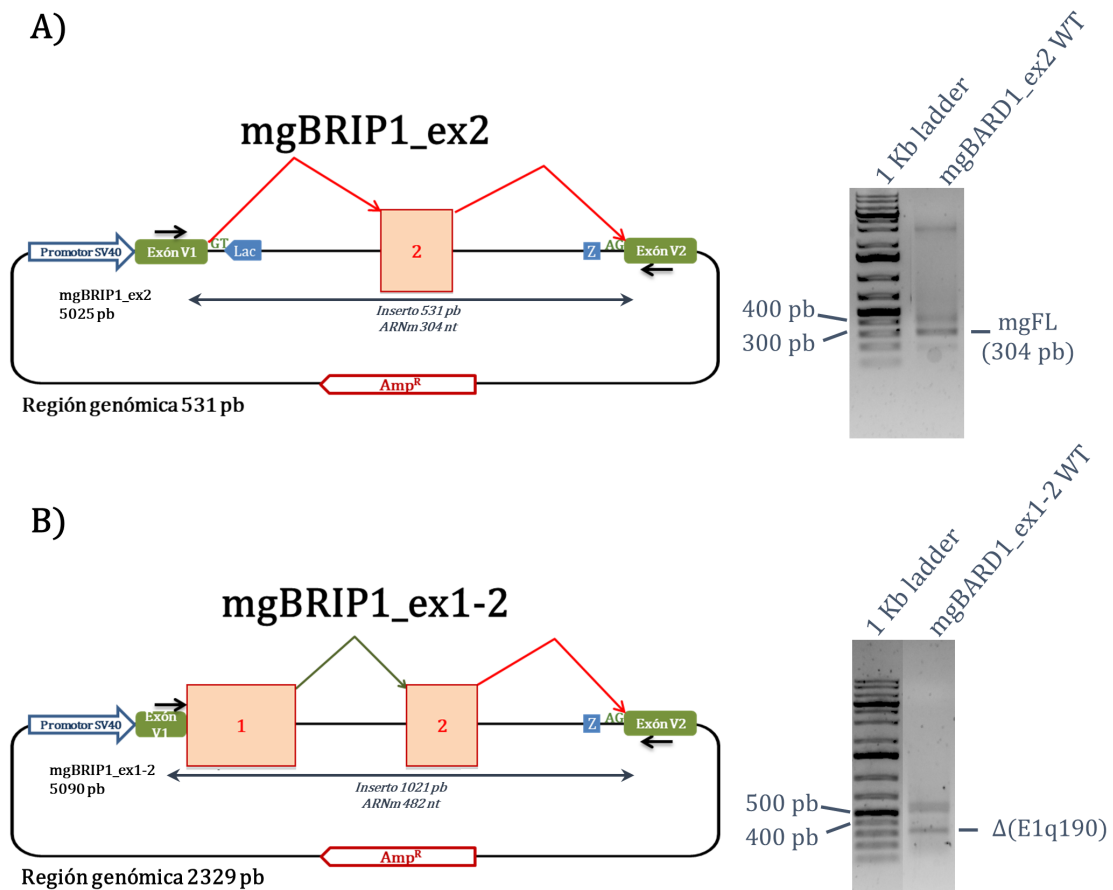


Figura Suplementaria 13. Estructura y validación funcional de los minigenes intermediarios de *BRIP1*. A la izquierda tenemos la representación gráfica de los minigenes intermediarios de *BRIP1* A) mgBRIP1_ex2 y B) mgBRIP1_ex1-2. Los exones están dentro de recuadros de color rojizo; las flechas negras sobre los exones V1 y V2 indican los primers de RT-PCR específicos del vector. A la derecha se muestra la validación funcional de los minigenes por electroforesis en gel de agarosa.