



---

Universidad de Valladolid

PROGRAMA DE DOCTORADO EN INVESTIGACION  
BIOMEDICA

TESIS DOCTORAL:

**Desregulación del *splicing* del gen  
*PALB2* en cáncer de mama  
hereditario.**

Presentada por Alberto Valenzuela Palomo para  
optar al grado de Doctor por la Universidad de  
Valladolid

Dirigida por: Eladio A. Velasco Sampedro



## **Productividad científica.**

**Durante la duración de esta tesis el doctorando ha contribuido en las siguientes publicaciones científicas:**

Villate.O; Ibarlucea.N; Fraile-Bethencourt. E; **Valenzuela-Palomo.A**; Velasco.EA; Grozeva.D; Raymond.FL; Botella.MP; Tejada.MI. Functional analyses of a novel splice variant in the CHD7 gene, found by Next. *Front Genet.* 9, pp. 7. 08/01/2018.

Fraile-Bethencourt. E; **Valenzuela-Palomo.A**; Diez-Gomez.B; Acedo.A; Velasco.A. Identification of Eight Spliceogenic Variants in BRCA2 Exon 16 by Minigene Assays. *Front. Genet.* *Front. Genet.*, 24/05/2018.

Fraile-Bethencourt. E; **Valenzuela-Palomo.A**; Diez-Gomez.B; Marcos.G; Infante.M; Lastra.E; Duran.M; Gomez-Barrero.S; Velasco.EA. Genetic dissection of the BRCA2 promoter and transcriptional impact of DNA variants. *Breast Cancer Res Tr.* 15/05/2018.

Lopez-Perolio I; Leman R; Behar R; Lattimore V; Pearson JF; Castéra L; Martins A; Vaur D; Goardon N; DavyG; Garre P; García-Barberán V; Llovet P; Pérez-Segura P; Díaz-Rubio E; Caldés T; Hruska KS; Hsuan V; Wu S; Pesaran T; Karam R; Vallon-Christersson J; Borg A; Investigators K; **Valenzuela-Palomo A**; Velasco EA; SoutheyM; Vreeswijk MPG; Devilee P; Kvist A; Spurdle AB; Walker LC; Krieger S; de la Hoya M. Alternative Splicing and Based Classification of PALB2 Genetic Variants: an ENIGMA Report. *J Med Genet.* 56, pp. 453 - 460. 03/2019.

Fraile-Bethencourt. E; **Valenzuela-Palomo.A**; Diez-Gomez.B; Goina.E; Acedo.A; Buratti.E; Velasco.EA. Mis-splicing in breast cancer: identification of pathogenic BRCA2 variants by systematic minigene assays. *J Pathol.* 248, pp. 409 - 420. 18/03/2019.

Fraile-Bethencourt. E; **Valenzuela-Palomo.A**; Diez-Gomez.B; Caloca.MJ; Gomez-Barrero.S; ELADIO ANDRES VELASCO SAMPEDRO. Minigene Splicing Assays Identify 12 Spliceogenic Variants of BRCA2 Exons 14 and 15. *Front Genet.* 10, pp. 503. 20/05/2019.

**Valenzuela-Palomo. A**; Galite. L; Sanoguera-Miralles. L; Rots. D; Kreile. M; Velasco. EA. 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 Mar 6;11: 169.doi: 10.3389/fgene.2020.00169. eCollection 2020.

Sanoguera-Miralles. L; **Valenzuela-Palomo. A**; Bueno-Martínez E, Llovet P, Díez-Gómez B, Caloca MJ, Pérez-Segura P, Fraile-Bethencourt E, Colmena M; Carvalho S,

Allen J, Easton DF, Devilee P, Vreeswijk MPG, de la Hoya M, Velasco EA. Comprehensive Functional Characterization and Clinical Interpretation of 20 Splice-Site Variants of the RAD51C Gene. *Cancers*. 15/12/2020.

Bueno-Martínez.E; Sanoguera-Miralles.L; **Valenzuela-Palomo.A**; Lorca.V; Gómez-Sanz.A; Carvalho.S; Allen.J; Infante.M; Perez-Segura.P; Lazaro.C; Easton.DF; Devilee.P; Vreeswijk.MPG; de la Hoya.M; Velasco.EA. RAD51D Aberrant Splicing in Breast Cancer: Identification of Splicing Regulatory Elements and Minigene-Based Evaluation of 53 DNA Variants. *Cancers*. 07/06/2021.

**Valenzuela-Palomo.A**; Bueno-Martinez.E; Sanoguera-Miralles.L; Lorca.V; Fraile-Bethencourt.E; Esteban-Sanchez.A; Gómez-Barrero.S; Carvalho.S; Allen.J; García-Álvarez.A; Pérez-Segura.P; Dorling.L; Easton.D; Devilee.P; Vreeswijk.M; de la Hoya.M; Velasco.EA. Splicing predictions, minigene analyses and ACMG/AMP clinical classification of 42 germline PALB2 splice-site variants. *J Pathol*. 2021 Nov 30. doi: 10.1002/path.5839. Online ahead of print.

**Esta tesis también ha dado lugar a las siguientes comunicaciones a congresos:**

Fraile-Bethencourt. E; Diez-Gomez.B; **Valenzuela-Palomo.A**; Acedo.A; Sanz.DJ; Goina.E; Buratti.E; Velasco.EA. "BRCA 2 Mis-Splicing: Regulación de los exones 17 y 18". Madrid 2017. Poster.

Villate.O; Fraile-Bethencourt. E; **Valenzuela-Palomo.A**; Velasco.EA; Grozeva.D; Raymond.FL; Botella.MP; Tejada.MI. Análisis funcionales confirman la patogenicidad de una variante de splicing en el gen chd7 hallada mediante secuenciación masiva. Madrid 2017. Poster.

**Valenzuela-Palomo.A**; Fraile-Bethencourt. E; Diez-Gomez.B; Estrala.C; Martinez-Delgado.B; Tejada.MI; Acedo.A; Velasco.EA. Caracterización funcional de variantes candidatas de splicing en genes de susceptibilidad mediante minigenes híbridos. Madrid 2017. Poster

Fraile-Bethencourt. E; Diez-Gomez.B; **Valenzuela-Palomo.A**; Acedo.A; Sanz.DJ; Goina.E; Buratti.E; Velasco.EA. "BRCA2 mis-splicing: exons 17 and 18 regulation". Copenhagen 2017. Poster.

Fraile-Bethencourt. E; **Valenzuela-Palomo.A**; Diez-Gomez.B; Acedo.A; Velasco.E. Clasification of 15 new brca2 exons 2-9 splicing variants by hybrid minigenes. Milan 2018. Poster.

**Valenzuela-Palomo.A**; Sanogera-Miralles.L; Fraile-Bethencourt.E; Diez-Gomez.B; Vreeswijk.M; Devilee.P; de la Hoya.M; Velasco.EA. Desregulación del splicing en Cáncer de Mama Hereditario (CMH)": Clasificación funcional y clínica de variantes de ADN del gen de susceptibilidad PALB2. Madrid 2019. Comunicación Oral.





## **TABLA DE CONTENIDOS**

<b>ABREVIATURAS</b> .....	15
<b>RESUMEN</b> .....	21
<b>INTRODUCCIÓN</b> .....	25
<b>1. Cáncer de mama.</b> .....	26
1.1. Cáncer de Mama Hereditario.....	28
1.2. Diagnóstico genético en cáncer de mama/ovario hereditario. ....	32
<b>2. El gen <i>PALB2</i> en Cáncer de Mama Hereditario</b> .....	33
2.1. Estructura y función de <i>PALB2</i> .....	33
2.2. <i>PALB2</i> y predisposición al cáncer.....	35
2.3. <i>PALB2</i> y Anemia de Fanconi. ....	38
2.4. Variantes de <i>PALB2</i> .....	39
<b>3. Dogma central de la Biología Molecular.</b> .....	43
3.1. Maduración del mRNA o <i>splicing</i> .....	44
<b>OBJETIVOS</b> .....	65
<b>MATERIALES Y MÉTODOS</b> .....	67
<b>1. Bioinformática.</b> .....	68
1.1. Bases de datos y programas:.....	68
1.2. Selección de variantes candidatas de <i>splicing</i> . ....	71
<b>2. Ensayos funcionales de <i>splicing</i> mediante minigenes híbridos.</b> .....	75
2.1. Minigenes en pMAD. ....	75
2.2. Minigenes de <i>PALB2</i> .....	77
2.3. Mutagénesis dirigida. ....	84

2.4.	Transfección.....	89
2.5.	Retrotranscripción.....	90
2.6.	Análisis de transcritos.....	91
<b>3.</b>	<b>Interpretación clínica de variantes.....</b>	<b>92</b>
3.1.	Recomendaciones propias en la clasificación de variantes de sitios de <i>splicing</i> .....	97
<b>RESULTADOS.....</b>		<b>99</b>
<b>1.</b>	<b>Análisis de variantes detectadas en el proyecto BRIDGES.....</b>	<b>100</b>
1.1	Análisis Bioinformático.....	100
1.2	Construcción de los minigenes.....	104
1.3	Ensayos funcionales.....	108
<b>2.</b>	<b>Análisis <i>in silico</i> y funcional de variantes de <i>PALB2</i> reportadas en la base de datos ClinVar.....</b>	<b>137</b>
2.1.	Estudio bioinformático de variantes ClinVar.....	137
2.2.	Estudio funcional de variantes ClinVar.....	146
<b>3.</b>	<b>Mapeo funcional de elementos reguladores del <i>splicing</i> de los exones 11 y 12 de <i>PALB2</i>.....</b>	<b>152</b>
3.1.	Estudio de microdeleciones del exón 11.....	153
3.2.	Estudio de microdeleciones del exón 12.....	154
3.3.	Selección de variantes ClinVar de los exones 11 y 12 de <i>PALB2</i> situadas en las microdeleciones c.3143del29, c.3229del30 y c.3293del30.....	155
	c.3157G>T p.(Asp1053Tyr).....	158
<b>DISCUSIÓN.....</b>		<b>162</b>
<b>1.</b>	<b>Variantes BRIDGES.....</b>	<b>166</b>

1.1. Clasificaciones Clínica de variantes BRIDGES.....	169
1.2. Transcritos de <i>PALB2</i> biológicamente relevantes. Regiones codificantes críticas de <i>PALB2</i> .....	170
<b>2. Variantes ClinVar.....</b>	<b>175</b>
<b>3. Mapeo Funcional de elementos reguladores del <i>splicing</i> de los exones 11 y 12 de <i>PALB2</i>.....</b>	<b>178</b>
<b>CONCLUSIONES.....</b>	<b>182</b>
<b>BIBLIOGRAFÍA.....</b>	<b>186</b>
<b>ANEXOS.....</b>	<b>216</b>

## **LISTA DE FIGURAS**

<b>Figura 1.</b> Incidencia (arriba) y número de muertes (abajo) estimadas por tipo de cáncer a nivel mundial ( <a href="https://gco.iarc.fr/today">https://gco.iarc.fr/today</a> ) .....	27
<b>Figura 2.</b> Modelo de oncogénesis en 2 pasos de Knudson.....	31
<b>Figura 3.</b> Distribución de las variantes de <i>PALB2</i> registradas en la base de datos ClinVar.....	40
<b>Figura 4.</b> Distribución de variantes de significado clínico desconocido de <i>PALB2</i> en ClinVar....	41
<b>Figura 5 A y B.</b> Esquema representativo del dogma central de la biología molecular.....	43
<b>Figura 6.</b> Secuencias básicas de reconocimiento de intrones por el <i>espliceosoma</i> . (Ma et al., 2015).....	46
<b>Figura 7.</b> Secuencias básicas de reconocimiento de intrones por el <i>espliceosoma</i> menor. (Ma et al., 2015).....	46
<b>Figura 8.</b> Ensamblaje del <i>espliceosoma</i> y reacción de <i>splicing</i> (Adaptacion de Will and Luhrmann, 2011).....	47
<b>Figura 9.</b> Patrones de <i>splicing</i> alternativo .....	51
<b>Figura 10.</b> Regulación del proceso de <i>splicing</i> por elementos en <i>cis</i> y factores trans de <i>splicing</i> (Proteínas SR, hnRNP, etc). .....	53
<b>Figura 11.</b> Estructura de las proteínas SR y hnRNPs. Lista y dominios de proteínas SR. ....	54
<b>Figura 12.</b> Efecto de la desregulación del <i>splicing</i> alternativo sobre la progresión del cáncer.....	62
<b>Figura 13.</b> Esquema del vector pMAD.....	76
<b>Figura 14.</b> Estructura del minigén de los exones 1 a 3 de <i>PALB2</i> .....	77
<b>Figura 15.</b> Secuencia del inserto del minigén mgPALB2_ex1-3. ....	78
<b>Figura 16.</b> Estructura del minigén de los exones 4 a 6 de <i>PALB2</i> .....	79
<b>Figura 17.</b> Secuencia del inserto del minigén mgPALB2_ex4-6. ....	79
<b>Figura 18.</b> Estructura del minigén de los exones 5 a 12 de <i>PALB2</i> .....	80
<b>Figura 19.</b> Secuencia del inserto en el minigén mgPALB2_ex5-12.....	80
<b>Figura 20.</b> El minigén mgPALB2_ex1-3.. .....	106

<b>Figura 21.</b> El minigén mgPALB2_ex4-6..	107
<b>Figura 22.</b> El minigén mgPALB2_ex5-12.....	108
<b>Figura 23.</b> Ensayos funcionales de <i>splicing</i> de variantes <i>del</i> minigén mgPALB2_ex1-3.....	110
<b>Figura 24.</b> Electroforesis capilar de c.49-2A>T en los minigenes mgPALB2_ex1-3 y mgPALB2_ex2-3.....	112
<b>Figura 25.</b> Electroforesis capilar de c.48+4C>T y c.48+7G>C, c.109-6_109-4del y c.211+5del en el minigén mgPALB2_ex1-3.....	121
<b>Figura 26.</b> Reproducibilidad de los eventos de <i>splicing</i> en células de cáncer de mama triple negativas MDA-MB-231 en el minigén mgPALB2_ex1-3.....	122
<b>Figura 27.</b> Transcrito canónico V1-ex4-ex5-ex6-V2 con un tamaño de 2.556 nt y transcritos generados por las variantes: c.1684+4A>G, c.1685-2A>C, c.1685-2A>G, c.2513A>C y c.2515-2A>G.....	123
<b>Figura 28.</b> Análisis de la variante c.1684+4A>G, mediante electroforesis en gel de agarosa con los oligonucleótidos: RT-PSPL3-Fw y RT-PALB2-ex6-rv.....	124
<b>Figura 29.</b> Electroforesis capilar de c.1685-2A>C, c.1685-2A>G, c.2513A>C y 2515-2A>G en el minigén mgPALB2_ex4-6.....	125
<b>Figura 30.</b> Electroforesis capilar de c.2587-2A>G, c.2748+1G>T, c.2748+2dup y c.2748+1G>A en el minigén mgPALB2_ex5-12.....	128
<b>Figura 31.</b> Electroforesis capilar de c.2749-1G>T, c.2834G>C y c.2834+1G>A en el minigén mgPALB2_ex5-12.....	129
<b>Figura 32.</b> Electroforesis capilar de c.2834+2T>C c.2834+3A>G y c.2834+5G>A en el minigén mgPALB2_ex5-12.....	130
<b>Figura 33.</b> Electroforesis capilar de c.2586+4A>T, c.2748+4A>T, c.2750T>C y 2834+6T>C en el minigén mgPALB2_ex5-12.....	131
<b>Figura 34.</b> Electroforesis capilar de c.2996 + 4A>G en el minigen mgPALB2_ex5-12.....	132
<b>Figura 35.</b> Electroforesis capilar de c.2997-1G>A y c.2997-2del en el minigén mgPALB2_ex5-12.....	133

<b>Figura 36.</b> Electroforesis capilar de c.3113G>A, c.3113+3A>G y c.3113+5G>C en el minigén mgPALB2_ex5-12.....	133
<b>Figura 37.</b> Electroforesis capilar de c.3201+1G>A, c.3201+3_3201+4insTG y c.3201+6T>A en el minigén mgPALB2_ex5-12.....	134
<b>Figura 38.</b> Electroforesis capilar de c.3348C>T, c.3350+4A>G y c.3350+5G>A en el minigén mgPALB2_ex5-12.....	135
<b>Figura 39.</b> Reproducibilidad de los eventos de <i>splicing</i> en células de cáncer de mama triple negativas MDA-MB-231 en el minigén mgPALB2_ex5-12.....	136
<b>Figura 40.</b> Ensayos funcionales de <i>splicing</i> de variantes <i>PALB2</i> seleccionadas de la base de datos ClinVar. (A) Electroforesis capilar de fluorescencia de productos RT-PCR de los minigén <i>wild type</i> mgPALB2_ex1-3 y mutantes .....	148
<b>Figura 41.</b> Electroforesis capilar de las microdeleciones c.3116del30, c.3143del29 y c.3169del30 en el minigén mgPALB2_ex5-12.....	154
<b>Figura 42.</b> Electroforesis capilar de las microdeleciones c.3229del30, c.3254del44, y c.3293del30 en el minigén mgPALB2_ex5-12.....	155
<b>Figura 43.</b> Electroforesis capilar de las variantes c.3155A>T, c.3157G>T y c.3161C>T en el minigén mgPALB2_ex5-12.....	157
<b>Figura 44.</b> Electroforesis capilar de las variantes c.3242A>G, c.3244A>G y c.3251C>T en el minigén mgPALB2_ex5-12.....	157
<b>Figura 45.</b> Electroforesis capilar de las variantes c.3294_3298del, c.3306C>T y c.3306C>G en el minigén mgPALB2_ex5-12.....	159
<b>Figura 46.</b> Mapa de microdeleciones del exón 11 Y 12 de <i>PALB2</i> y variantes exónicas ensayadas. ....	160
<b>Figura 47.</b> Secuencias del sitio de <i>splicing</i> GT canónico y los sitios GC atípicos de las variantes c.48+2T>C, c.108+2T>C, y c.2834+2T>C, del exón 12 de <i>PALB2</i> , del exón 17 de <i>BRCA2</i> exón 17 y de los transcritos Δ(E10q41) y Δ(E10q65).....	169

## **LISTA DE TABLAS**

<b>Tabla 1.</b> Criterios de selección de pacientes en el plan de prevención para CMOH. ....	32
<b>Tabla 2.</b> Numero de acceso de la secuencia del gen <i>PALB2</i> de Homo Sapiens. ....	69
<b>Tabla 3.</b> Clasificación por estrellas en la base de datos ClinVar.....	70
<b>Tabla 4.</b> Secuencia de oligonucleótidos utilizados para la construcción de los minigenes: mgPALB2_ex1-3, mgPALB2_ex4-6 and mgPALB2_ex5-12. ....	81
<b>Tabla 5.</b> Mezcla de amplificación en el rastreo de colonias por PCR. ....	83
<b>Tabla 6.</b> Mezcla de amplificación PCR de mutagénesis, PCR inserción y delección. ....	84
<b>Tabla 7.</b> Mutaciones y secuencia de oligonucleótidos utilizados para la mutagénesis dirigida de las variantes BRIDGES. ....	85
<b>Tabla 8.</b> Variantes y secuencia de oligonucleótidos utilizados para la mutagénesis dirigida de las variantes procedentes de ClinVar.....	87
<b>Tabla 9.</b> Variantes y secuencia de oligonucleótidos utilizados para la mutagénesis dirigida de las variantes exónicas de los exones 11 y 12 de <i>PALB2</i> , procedentes de ClinVar. ....	88
<b>Tabla 10.</b> Variantes y secuencia de oligonucleótidos utilizados para la mutagénesis dirigida de las microdelecciones de los exones 11 y 12 de <i>PALB2</i> , procedentes de ClinVar.....	88
<b>Tabla 11.</b> RT-PCR MIX .....	90
<b>Tabla 12.</b> Criterios de clasificación de variantes patogénicas .....	93
<b>Tabla 13.</b> Criterios de clasificación de variantes benignas.....	96
<b>Tabla 14.</b> Variantes de <i>PALB2</i> estudiadas in silico del proyecto BRIDGES en el minigén mgPALB2_ex1-3.....	101
<b>Tabla 15.</b> Variantes de <i>PALB2</i> estudiadas in silico del proyecto BRIDGES en el minigén mgPALB2_ex4-6.....	102
<b>Tabla 16.</b> Variantes de <i>PALB2</i> estudiadas in silico del proyecto BRIDGES en el minigén mgPALB2_ex5-12.....	103

<b>Tabla 17.</b> Eventos de <i>splicing</i> producidos por variantes encontradas en pacientes BRIDGES del gen <i>PALB2</i> .....	113
<b>Tabla 18.</b> Anotación de transcritos en el mgPALB2_ex1-3, mgPALB2_ex4-6 y mgPALB2_ex5-12 de acuerdo a la secuencia de <i>PALB2</i> NM_024675.4.....	116
<b>Tabla 19.</b> Variantes de <i>PALB2</i> estudiadas in silico de la base de datos ClinVar en el minigén mgPALB2_ex1-3.....	138
<b>Tabla 20.</b> Variantes de <i>PALB2</i> estudiadas in silico de la base de datos ClinVar en el minigén mgPALB2_ex4-6.....	140
<b>Tabla 21.</b> Variantes de <i>PALB2</i> estudiadas in silico de la base de datos ClinVar en el minigén mgPALB2_ex5-12.....	142
<b>Tabla 22.</b> Resultados de los estudios de <i>splicing</i> de variantes ClinVar de los exones 1 a 3 de <i>PALB2</i> . .....	149
<b>Tabla 23.</b> Anotación de los transcritos de acuerdo con la secuencia de <i>PALB2</i> NM_024675.4.	150
<b>Tabla 24.</b> Eventos de <i>splicing</i> producidos a causa de las microdeleciones seleccionadas en los exones 11 y 12 de <i>PALB2</i> .....	153
<b>Tabla 25.</b> Eventos de <i>splicing</i> producidos por variantes exónicas ClinVar y situadas en nuestras deleciones positivas.....	158
<b>Tabla 26.</b> Clasificación ACMG/AMP de las variantes BRIDGES ensayadas. ....	175

# **ABREVIATURAS**

**3'ss:** Sitio 3' de *splicing*.

**5'ss:** Sitio 5' de *splicing*.

**aa:** aminoácido.

**ACMG:** American College of Medical Genetic and Genomics.

**AME:** atrofia muscular espinal.

**AMP:** Association for Molecular Pathology.

**ATM:** *Ataxia Telangiectasia Mutated gene*.

**ATP:** Adenosine Triphosphate.

**BARD1:** *BRCA1-associated RING domain 1*.

**BC:** Breast Cancer.

**BINI:** *Bridging Integrador 1 gene*.

**BRCA1:** *Breast Cancer type 1 gene*.

**BRCA2:** *Breast Cancer type 2 gene*.

**BRIDGES:** Breast Cancer After Diagnostic Gene Sequencing.

**BRIP1:** *BRCA1 interacting Protein C-Terminal Helicase 1 gene*.

**CC:** *Coiled Coil* (del castellado, superenrollado).

**CDH1:** *Cadherin 1 gene*.

**cDNA:** complementary DNA.

**CFTR:** *Cystic Fibrosis Transmembrane conductance Regulator gene*.

**CHAM:** chromatin-association motif

**CHEK2:** *Checkpoint Kinase 2 gene*.

**circRNA:** RNA circulares.

**CMHO:** Cáncer de Mama y Ovario Hereditario.

**Da/KDa:** Dalton/kiloDalton.

**DMPK:** *DM1 protein kinase gene*.

**DNA:** Ácido dexosirribonucléico.

**ENCODE:** Encyclopedia of DNA Elements.

**ENIGMA:** Evidence-based Network for the Interpretation of Germline Mutant Alleles.

**ER:** *Estrogen receptor.*

**ERBB2 o HER2:** *Human Epidermal growth factor gene.*

**ESE:** *Exonic Splicing Enhancer.*

**ESS:** *Exonic Splicing Silencer.*

**FA:** *Fanconi Anemia.*

**Fw:** Forward.

**GTR:** Genetic Testing Registry.

**HGVS:** The Human Variation Society.

**Hi-Fi PCR:** High-fidelity PCR.

**HnRNP:** *Ribonucleoproteínas heterogéneas nucleares*

**HR:** *Homologous Recombination.*

**IARC:** International agency for research on cancer.

**ISE:** *Intronic splicing enhancer.*

**ISS:** *Intronic splicing silencer.M*

**Kb:** kilobase

**KEAPI:** *Kelch Like ECH Associated Protein 1.*

**KH:** *The K Homology.*

**LB:** Luria Bertani.

**LKBI-STK11:** *Serine/Threonine Kinase 11 gene.*

**LM:** *Low Melt.*

**LncRNAs:** *Long non-coding RNA.*

**LoF:** Loss of Function.

**MES:** Max Ent Scan.

**MLH1:** *MutL Homolog 1 gene.*

**MMC:** *intravesical chemotherapy using mitomycin C.*

**MRG15:** *Morf-Related gene 15.*

**mRNA:** *RNA mensajero.*

**MSH2:** *Muts Homolog 2 gene.*

**MSH6:** *Muts Homolog 6 gene.*

**MYC:** *MYC proto-oncogene, bHLH transcription factor.*

**ncRNA:** *non-coding RNA.*

**NBN:** *Nibrin Gene.*

**NCBI:** National Center for Biotechnology Information.

**NF1:** *Neurofibromin 1 gene.*

**NGS:** *Next Generation Sequencing.*

**NIH:** National Institutes of Health.

**NLM:** National Library of Medicine.

**NMD:** *Nonsense Mediated Decay.*

**NFR2:** Nuclear Factor Erythroid 2-Related factor 2.

**nt:** nucleótido.

**OC:** *Ovarian Cancer.*

**OMS:** Organización Mundial de la Salud.

**PALB2:** *Partner and Localizer of BRCA2 gene.*

**PARP:** *Poly (Adenosine diphosphate-ribose) polymerase.*

**pb:** pares de bases.

**PCR:** *Polymerase Chain Reaction.*

**pMAD:** *Mis- Splicing And Disease plasmid*

**PR:** *Progesterone Receptor.*

**PRPF3:** Pre-mRNA Processing Factor 3.

**PRPF8:** Pre-mRNA Processing Factor 8.

**PRPF31:** Pre-mRNA Processing Factor 31.

**pSAD:** *Splicing And Disease plasmid.*

**PTC:** *Premature Termination Codon.*

**PTEN:** *Phosphatase and Tensin Homolog gene.*

**qRRM:** quasi RNA recognition motif.

**RAD51:** *RAD51 Recombinase.*

**RAD51C:** RAD51 Paralog C.

**RAD51D:** RAD51 Paralog D.

**RBD/ RRM:** *RNA Binding domain.*

**RBP:** *RNA Binding Proteins.*

**REDECAN:** RED Española de Registros de Cáncer.

**RMN:** Resonancia Magnética Nuclear.

**RNA:** Acido Ribonucleico.

**NFR2:** *Nuclear factor erythroid 2-related factor 2.*

**RNApol:** RNA polimerasa.

**RRM:** *RNA Recognition Motif.*

**RRMH:** *RNA Recognition Motiv Homologous.*

**RS:** *Serine/ Arginine Rich Domain.*

**RT-PCR:** *Reverse Transcription Polymerase Chain Reaction.*

**Rv:** Reverse.

**SEOM:** Sociedad Española de Oncología Médica.

**siRNA:** *small interfering RNA.*

**SMNI:** *Survival Motor Neuron gene 1.*

**snRNP:** *small nuclear Ribonucleoproteins.*

**SREs:** *Splicing Regulatory Elemens.*

**SRp:** *Serine/Arginine rich protein.*

**SRSF:** *Serite/Arginine Splicing Factor.*

**ss:** sitios de *splicing.*

**STK11:** *Serine/Threonine kinase 11 gene.*

**TCGA:** The Cancer Genome Atlas.

**TNBC:** *Triple Negative Breast Cancer.*

**TP53:** *Transfromation-related Protein 53 gene.*

**UMD:** The Universal Mutation Database.

**UTR:** *Untraslated Region.*

**VUS:** *Variant of Unknown Significance.*

**Wt:** *wild type.*

**ZNF9:** *zinc finger protein 9 gene.*

# **RESUMEN**

El cáncer de mama continúa siendo un problema de salud pública a escala mundial, situándose como el cáncer más diagnosticado en mujeres y ocupando el segundo lugar entre las causas de muerte por cáncer en este grupo. La capacidad para identificar y diagnosticar el cáncer de mama ha mejorado notablemente en los últimos años y recientemente estudios de secuenciación a gran escala han demostrado como variantes de pérdida de función en al menos 8 genes (*BRCA1*, *BRCA2*, *PALB2*, *BARD1*, *RAD51C*, *RAD51D*, *ATM* y *CHEK2*) confieren un alto riesgo de desarrollar esta enfermedad (Dorling et al., 2021). En esta tesis doctoral, presentamos los resultados del análisis funcional de las variantes de *splicing* del gen *PALB2* detectadas en más de 113.000 mujeres del proyecto de secuenciación a gran escala BRIDGES (Breast Cancer After Diagnostic Gene Sequencing; <https://bridges-research.eu/>). En este proyecto se analizaron 82 variantes de *PALB2* en los límites intrón-exón con MaxEntScan y a continuación, se seleccionaron 44 variantes para su posterior ensayo funcional de *splicing*. Para este propósito, se construyeron tres minigenes de *splicing* que comprenden los exones 1-12. Las 44 variantes potencialmente *esplíceogénicas* se introdujeron en los minigenes mediante mutagénesis dirigida y se ensayaron en células MCF-7 / MDA-MB-231. Treinta y cinco de las 44 variantes alteraron el *splicing*, de las cuales 25 no mostraron rastros o mostraron cantidades mínimas de transcrito *full-length*. Se caracterizaron 33 transcritos aberrantes producidos por diferentes variantes, 23 de las cuales predecían el truncamiento de la proteína *PALB2*. Asimismo, se analizaron con MES 151 variantes reportadas en ClinVar de las fronteras intrón-exón de *PALB2*, de las que 67 fueron seleccionadas para estudios funcionales, y 11 fueron ensayadas en el mgPALB2\_ex1-3. Todas las variantes ensayadas alteraron el *splicing*, de las cuales 10 no revelaron la presencia del transcrito canónico. Por otro lado, se realizó un mapeo funcional mediante microdeleciones donde se identificaron 3 zonas ricas en ESEs de los exones 11 y 12 de *PALB2*: c.3143-c.3171, c.3229-c.3258 y c.3293-c.3322 y se ensayaron funcionalmente 9 variantes reportadas por ClinVar y localizadas en estos intervalos. Ocho de ellas alteraron el *splicing*. Por último queremos recalcar la sensibilidad y precisión de los ensayos funcionales de *splicing* con minigenes como estrategia para la caracterización inicial de las variantes de *splicing* y la posterior interpretación clínica de las variantes de cualquier enfermedad

genética. La desregulación del *splicing* es uno de los principales mecanismos de inactivación génica en enfermedades hereditarias. Además, la gran mayoría de las variantes estudiadas alteraron el *splicing* (56/64, 87,5%), destacando que 31 fueron clasificadas como patogénicas reduciendo el porcentaje de variantes BRIDGES clasificadas como VUS reportadas por ClinVar, del 57% de VUS-ClinVar al 23% incluyendo nuestros resultados.

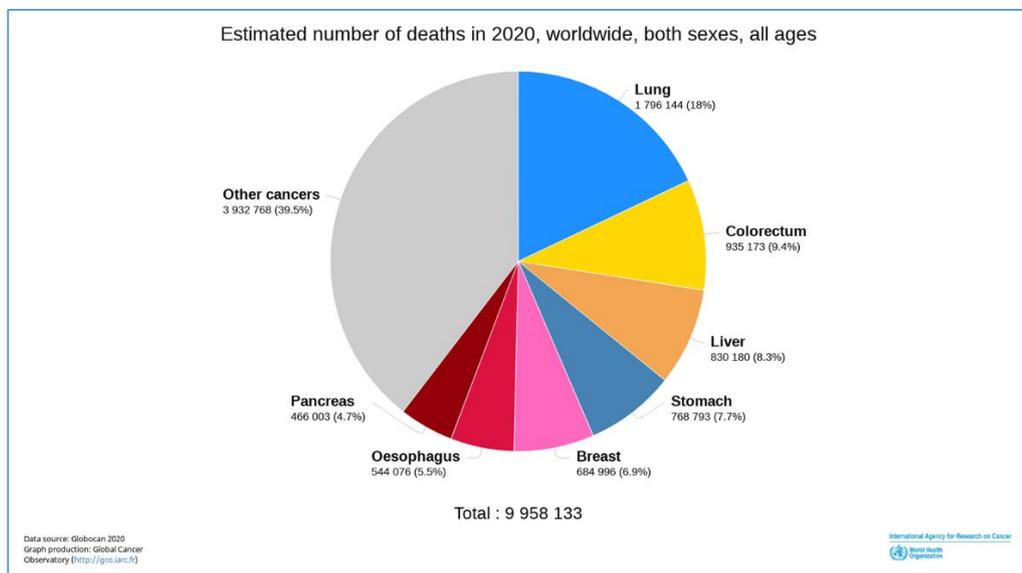
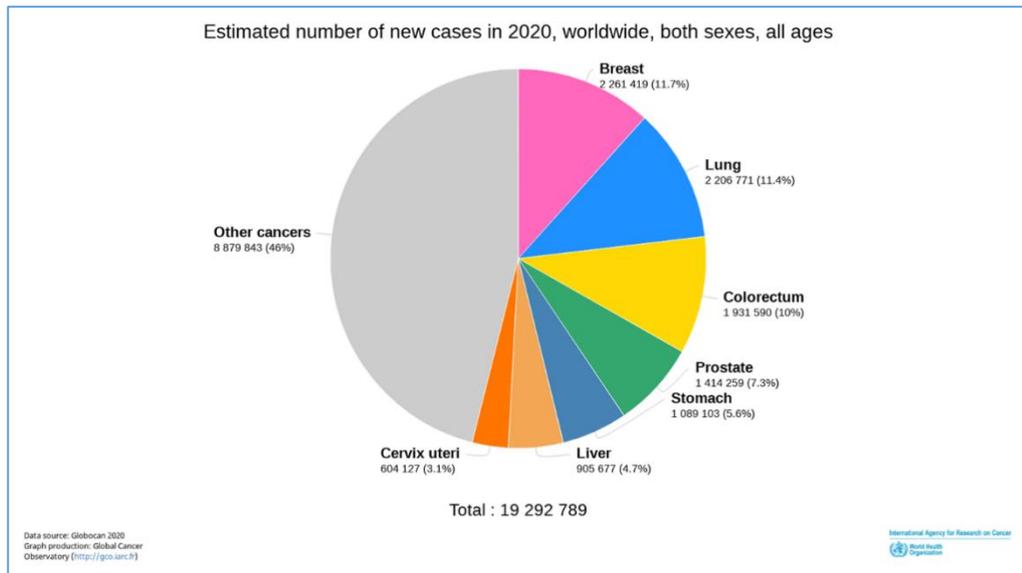


# **INTRODUCCIÓN**

## 1. Cáncer de mama.

El cáncer es una de las principales causas de muerte a nivel mundial que, de acuerdo a la Organización Mundial de la Salud (OMS; <https://www.who.int/es/news-room/fact-sheets/detail/cancer>), provocó cerca de 10 millones de fallecimientos en 2020. Engloba un gran número de enfermedades que se caracterizan por el desarrollo de células anormales, con la capacidad de dividirse, migrar, invadir y dañar tejidos y órganos. El cáncer se origina como resultado de un daño directo en el DNA o/y en los mecanismos de regulación del mismo, debido a la acumulación de mutaciones en genes que controlan procesos celulares esenciales. El cáncer de mama es una de las principales causas de muerte en el mundo desarrollado, siendo el cáncer más diagnosticado a nivel mundial en 2020 (2,26 millones de casos, 11,7%) seguido del cáncer de pulmón (2,21 millones de casos, 11,4%), y el quinto con más mortalidad asociada (685.000 fallecimientos, 6,9%), solo por detrás del cáncer de pulmón (1,8 millones de muertes, 18%), el cáncer colorrectal (935.173 fallecimientos, 9,4%), el cáncer de hígado (830.180, 8,3%) y el cáncer de estómago (7,7%) (Fuente: International Agency for Cancer Research; <https://gco.iarc.fr/today>, acceso 9 de agosto de 2021; Figura 1).

A fecha de hoy, junio de 2021, se estima que en España se alcanzarán los 276.239 casos, según los cálculos de la RED Española de Registros de Cáncer (REDECAN) [www.Redecan.org/ES](http://www.Redecan.org/ES); aunque esta estimación no incluye el posible efecto de la pandemia ocasionada por el COVID-19. Sin embargo, existen grandes variaciones en las tasas de incidencia estimada dentro de los diferentes países de la Unión Europea, siendo elevadas en los países de Europa occidental, especialmente en Bélgica (147 casos por cada 100.000 habitantes), Francia (137/100.000), los Países Bajos (131/100.000), y en el norte de Europa, en el Reino Unido (129/100.000) y en los países nórdicos, Dinamarca (143/100.000), Islandia (131/100.000) y Finlandia (121/100.000). En comparación, las tasas fueron considerablemente más bajas en países de Europa del este como Ucrania



**Figura 1. Incidencia (arriba) y número de muertes (abajo) estimadas por tipo de cáncer a nivel mundial (<https://gco.iarc.fr/today>)**

(54/100.000) y Moldavia (53/100.000). A su vez, las tasas de mortalidad más bajas en la Unión Europea (15-18/100.000) se calcularon en (España, Estonia, Portugal y Bosnia) y las tasas más altas en países de la península balcánica, como Serbia (31/100.000) y Macedonia (36/100.000) (Ferlay et al., 2013).

Los tumores de mama se clasifican basándose en sus perfiles de expresión génica como: luminal, ERBB2 o HER2 positivo, mama normal y *basal-like* o triple negativo (Perou et al., 2000).

La combinación de las características ER y PR (Receptores de estrógenos y progesterona), HER2 (Receptor 2 del Factor de Crecimiento Epidérmico Humano) y Ki67 (Marcador tumoral indicativo del grado de proliferación de las células tumorales) ayuda a clasificar los distintos tipos de tumor, que tienen las siguientes características (Irigoyen et al., 2011):

- Luminal A: Receptores hormonales positivos y HER2 negativo con un Ki67 > 14%.
- Luminal B: Mismos requisitos que el luminal A, pero con un Ki67 superior al 14%.
- Tipo normal: Receptores hormonales positivos y HER2 positivo.
- Her2 positivo: Receptores hormonales negativos y HER2 positivo (Sobreexpresado).
- Triple negativo o *basal like*: Receptores hormonales negativos y HER2 negativo.

Para tratar a los pacientes con enfermedad avanzada y tumores luminales A, suele utilizarse hormonoterapia junto con inhibidores de ciclinas (palbociclib, ribociclib y abemaciclib) que permiten retrasar el uso de la quimioterapia y los pacientes con tumores luminales B se tratan con quimioterapia, tratamientos hormonales y terapias biológicas frente a HER2 (Fuente: Sociedad Española de Senología y Patología Mamaria, <https://www.sespm.es> acceso 21 de agosto de 2021).

El cáncer de mama triple negativo (TNBC) representa aproximadamente el 15-20% de todos los cánceres de mama, tiene unas opciones de tratamiento más limitadas y se asocia con un peor pronóstico en comparación con otros subtipos de cáncer de mama (Foulkes et al., 2010). Los tratamientos utilizados en los pacientes con este tipo de cáncer se basan en la quimioterapia, aunque existen numerosos ensayos clínicos en marcha, como los estudios con inhibidores de PARP en pacientes con mutaciones en *BRCA1* y *BRCA2* (Andreopoulou et al., 2017; Geenen et al., 2018)

### **1.1. Cáncer de Mama Hereditario.**

El cáncer es una enfermedad genética que se genera por la acumulación de mutaciones somáticas en genes esenciales. El cáncer se considera una enfermedad genética esporádica, aunque excepcionalmente aparece asociada a alteraciones genéticas en la línea germinal, provocando formas hereditarias de cáncer, como el síndrome del Cáncer de Mama y Ovario Hereditario

(CMOH). Además, existen otros factores, genéticos y ambientales, que pueden modificar la penetrancia y expresión fenotípica de estas mutaciones (Allender and Arias, 2008).

Durante el siglo XIX se observó que mujeres de determinadas familias, tenían más predisposición a desarrollar cáncer de mama. La primera descripción de predisposición familiar a cáncer de mama fue realizada por el médico francés Pierre Paul Broca, cuya mujer había padecido cáncer de mama a una edad temprana y pertenecía a una familia con casos de cáncer de mama en 4 generaciones. Broca en 1866 (Broca, 1866; Krush, 1979) centró su estudio en cinco generaciones de la familia de su esposa y observó que 9 de 24 mujeres, habían sufrido la enfermedad.

En 1990, Hall y colaboradores encontraron alteraciones en el cromosoma 17q21 asociadas al cáncer de mama temprano, hecho que, además, fue corroborado en 1991 por Narod y colaboradores cuando encontraron el mismo locus 17q12-q23 ligado al cáncer de mama y ovario hereditario (Lenoir, 1991). Este gen, conocido como *BRCA1* (Breast Cancer type 1, OMIM #113705) y que se localiza en el cromosoma 17q21, se clonó en 1994 por primera vez (Miki et al., 1994). Un año más tarde, Wooster y colaboradores descubrieron el gen *BRCA2* (Breast Cancer type 2, OMIM #600185), que se encuentra en el cromosoma 13 banda q12 (Wooster et al., 1995). Estos genes producen proteínas supresoras de tumores y están implicados en el mantenimiento y la estabilización del genoma mediante procesos de reparación del DNA. Además, este hallazgo facilitó el diagnóstico molecular de esta enfermedad.

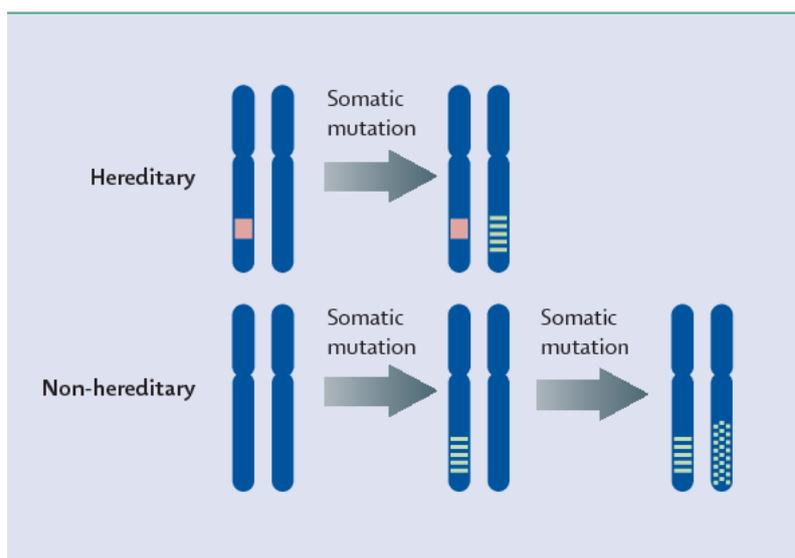
El cáncer de mama es una enfermedad asociada a factores de riesgo ambientales, hormonales y genéticos, siendo el historial familiar de cáncer de mama uno de los factores más contrastados (Beral et al., 2002). No obstante, los antecedentes familiares de cáncer de mama no siempre indican una predisposición hereditaria a la misma. Solo un pequeño porcentaje de todos los casos (5-10%) serán debidos a la presencia de alteraciones en genes de alta penetrancia (genes BC/OC), con un patrón de herencia autosómico dominante, y conferirán riesgo a CMOH (Foulkes, 2008). Sin embargo, a veces, el cáncer hereditario puede identificarse erróneamente como cáncer esporádico debido a fallos en la identificación de una mutación como causa hereditaria del cáncer.

En 1971 Knudson propuso el modelo genético *two hit* en los genes supresores de tumores, el cual defendía la necesidad de dos eventos genéticos independientes para iniciar la transformación neoplásica (Knudson, 1971). Según este modelo, la predisposición hereditaria del cáncer, implica una mutación en la línea germinal, mientras que la tumorigénesis requiere de una segunda mutación somática (Figura 2). Las formas no hereditarias requieren también 2 alteraciones genéticas, pero ambas de carácter somático. Siguiendo el anterior razonamiento y salvo en algunos casos muy específicos, parece evidente pensar que los cánceres no se producen por una única alteración genética sino por la acumulación de alteraciones que fomentan el crecimiento incontrolado de las células.

Un claro ejemplo del modelo *two hit* de Knudson se ve representado en el gen supresor de tumores *BRCA2* implicado en CMOH. *BRCA2* es un gen de la ruta de reparación del DNA de doble hebra mediante recombinación homóloga, que es un mecanismo crítico para la reparación de las roturas de DNA de doble cadena. La inactivación de sólo uno de los alelos de *BRCA2* no produce la enfermedad. Sin embargo, cuando se produce el segundo *hit*, o inactivación del otro alelo, se pierde la funcionalidad de la proteína. En 1998, Chen y colaboradores demostraron que *BRCA1* y *BRCA2* coexistían en el mismo complejo proteico, pero no tenían claro si tenían una interacción directa entre ellos (Chen et al., 1998). *BRCA1* y *BRCA2* fueron hasta 2006 los únicos genes de predisposición a cáncer de mama y ovario hereditario. En 2006, Xia y colaboradores describieron un nuevo compañero de interacción de los genes *BRCA1/2*, *PALB2*, fundamental en la estabilización y localización nuclear de la proteína *BRCA2*, y necesario para el reclutamiento de *BRCA1* en respuesta al daño en el DNA (Xia et al., 2006). A pesar de estas evidencias, hasta un año después *PALB2* no fue reconocido como un importante gen de susceptibilidad de cáncer de mama y, poco después, de cáncer de páncreas (Erkko et al., 2007; Stadler et al., 2011).

A día de hoy, *PALB2* es uno de los ocho genes (*BRCA1*, *BRCA2*, *PALB2*, *BARD1*, *RAD51C*, *RAD51D*, *ATM* y *CHEK2*) que muestran una asociación significativa con cáncer de mama, determinado por dos estudios a gran escala que han involucrado a 177.000 mujeres (Dorling et al., 2021; Hu et al., 2021), y se ha establecido firmemente como un gen de susceptibilidad, ya que las

variantes de pérdida de función de *PALB2* confieren un alto riesgo de desarrollar esta enfermedad. Ambas publicaciones, llegan a un consenso sobre estos 8 genes, sin embargo, otros genes, como *MSH6* y *CDH1*, también estarían asociados a la enfermedad dependiendo de la población de estudio. Así *MSH6*, ha sido asociado significativamente con la aparición del cáncer de mama en pacientes europeos (Dorling et al., 2021), y *CDH1* en pacientes estadounidenses (Hu et al., 2021). Las variantes patogénicas en *PALB2* son relativamente prevalentes y representan ~10% de todas las variantes patogénicas detectadas en los 8 genes centrales del cáncer de mama.



**Figura 2. Modelo de oncogénesis en 2 pasos de Knudson.**  
(Siegfried and Karni, 2018; Desterro et al., 2020)

Es importante tener en cuenta que el CMOH no es el único síndrome hereditario asociado a cáncer de mama, ya que existen otros síndromes con herencia autosómica dominante que predisponen a la aparición de este cáncer, siendo estos: el síndrome de Cowden, asociado a variantes del gen *PTEN* (Liaw et al., 1997; Lloyd and Hospital, 2017), el síndrome de Li-Fraumeni, asociado a mutaciones en el gen *TP53* (Li et al., 1988; Malkin et al., 1990) y el síndrome de Peutz-Jeghers, asociado al gen *LKB1-STK11* (BRUWER et al., 1954; Hemminki et al., 1998).

## 1.2. Diagnóstico genético en cáncer de mama/ovario hereditario.

La selección de pacientes CMOH es un proceso esencial de cara a asegurar la detección de una variante patogénica en un gen de susceptibilidad. En general, los criterios suelen ser bastante similares y pueden asegurar una probabilidad de detección de una variante patogénica mayor del 10%. En España, la selección se basa en criterios definidos por la Sociedad Española de Oncología Médica y los podemos observar en la Tabla 1 (SEOM; Cáncer Hereditario 3ª edición).

**Tabla 1. Criterios de selección de pacientes en el plan de prevención para CMOH. (Fuente: Sociedad Española de Oncología Médica (SEOM)).**

<b>Independientemente del historial familiar:</b>
Mujer con Cáncer de Mama (CM) y Cáncer de Ovario (CO) (Metacrónico o sincrónico)
CM < 40 Años
CM Bilateral (El primer CM < 40 años)
CM Triple Negativo ≤ 60 años
CO epitelial no mucinoso (o trompa o primario de peritoneo)
CM en varón con cáncer de mama
<b>Dos o más familiares de primer grado con alguna combinación de las siguientes características de alto riesgo:</b>
CM bilateral + otro CM < 50 años
CM + CO
2 casos de CM diagnosticados antes de los 50 años
<b>3 o más familiares directos con CM y/o CO:</b>
3 CM + CO

Esta detección precoz permite a las familias portadoras de mutaciones patogénicas asociadas a CMOH favorecerse del programa de prevención de cáncer de mama que se basa en: autoexámenes mensuales, exámenes clínicos semestrales, mamografías, ecografías y RMN anuales, además de ecografías transvaginales semestrales. Por último, las portadoras enfermas podrían participar en programas de quimioprevención con tamoxifeno para prevenir la posible aparición de

un tumor contralateral (Fisher et al., 2005). El plan de prevención también incluye un control del estilo de vida (peso, actividad física y alimentación) y cirugías preventivas como, la mastectomía bilateral profiláctica y la salpingo-ooforectomía bilateral profiláctica (Kurian et al., 2010). Estas cirugías preventivas han demostrado ser muy efectivas en la prevención, provocando una disminución del riesgo de padecer cáncer de mama y ovario hasta en un 90-95% (Mau and Untch, 2017).

## **2. El gen *PALB2* en Cáncer de Mama Hereditario.**

### **2.1. Estructura y función de *PALB2*.**

El gen *Partner And Localizer of BRCA2*, más comúnmente denominado *PALB2* (OMIM #610355), se distribuye a lo largo de ~40 Kb, está compuesto por 13 exones y se sitúa en el brazo corto del cromosoma 16 (16p12.2). La estructura proteica de *PALB2* se compone por 1186 aminoácidos que incluyen un dominio superenrollado muy conservado en su extremo amino terminal. Los exones 1, 2 y parte del 3 codifican este dominio. El extremo 3' del exón 4 genera el motivo asociado a la cromatina (ChAM), indispensable para su función biológica (Ducy et al., 2019), mientras que el dominio WD40 es producido por la región comprendida entre el extremo 3' del exón 7 y el exón 13 (Nepomuceno et al., 2017). Además, MRG15 es otro factor de interacción con *PALB2* involucrado en la asociación a la cromatina. En 2009, Sy y colaboradores determinaron que MRG15 y MRG eran cooperadores de *PALB2* mediante técnicas de purificación por afinidad y análisis de espectrometría de masas (Sy et al., 2009a). *PALB2* existe como oligómeros que pueden formar un complejo con *BRCA1/2* y con la recombinasa *RAD51* (Zhang et al., 2009a, 2009b), jugando un papel fundamental en la recombinación homóloga.

La proteína *PALB2* contiene un dominio superenrollado en la región amino terminal y una serie de repetición WD en su extremo C-terminal. Las repeticiones WD forman juntas una estructura  $\beta$ -propulsora que interacciona directamente con *BRCA2* (Oliver et al., 2017) y están involucradas en la interacción con la DNA polimerasa  $\eta$ , *RAD51*, *RAD51C* y la ubiquitina ligasa *RNF168* (Buisson et al., 2014; Park et al., 2014; Luijsterburg et al., 2017). Además, cualquier

alteración genética dentro de la región WD40 es capaz de perturbar la interacción PALB2-BRCA2 y causar deficiencias en la recombinación homóloga (Park et al., 2014). Por último, se ha visto que el motivo superenrollado del dominio N-terminal de PALB2, interacciona con BRCA1 (Sy et al., 2009b; Zhang et al., 2009a, 2009b). La proteína generada por *PALB2* se ha encontrado asociada con el 50% del BRCA2 celular siendo esencial en su localización cromatínica y su reclutamiento hacia los sitios dañados del DNA (Xia et al., 2006). Una vez ha sido detectado el daño en el DNA, las zonas finales de las hebras de DNA, son eliminadas originando un extremo 3' protuberante libre (DNAss), que se une a la proteína de unión al DNA de cadena simple, RPA. A continuación, PALB2 recluta hacia la zona de rotura a BRCA1, para facilitar el ensamblaje de BRCA2 y RAD51 a la zona final de rotura. RAD51 cataliza en diferentes etapas la invasión generalmente producida por una cromátida hermana, la transferencia de DNA en la fase S/G2 del ciclo celular y finalmente, lidera la reparación del error en el DNA de doble cadena (Sy et al., 2009b; Zhang et al., 2009a).

En 2012, Ma y colaboradores, revelaron la implicación de *PALB2* en los mecanismos de regulación del estrés oxidativo celular (Ma et al., 2012). Este grupo descubrió que, *KEAP1*, un mediador de estrés oxidativo que regula negativamente la función del factor de transcripción antioxidante NRF2, se une a PALB2. Además, determinaron que un motivo de 7 aminoácidos altamente conservado (LDEETGE) situado dentro del dominio de unión de PALB2 a KEAP1 (residuos 79-100), era idéntico al motivo ETGE de NRF2 mediante el que se une a KEAP1; poniendo de manifiesto la implicación de PALB2 en la actividad de NRF2 mediante su unión competitiva a KEAP1.

Además, PALB2 promueve la acumulación y función de NRF2 en el núcleo y reduce el nivel de especies reactivas de oxígeno celular. Asimismo, PALB2 también regula la tasa de exportación de NRF2 desde el núcleo después de la inducción. Estos hallazgos identifican a *PALB2* como un regulador de la homeostasis redox celular y proporcionan un nuevo vínculo entre el estrés oxidativo, el desarrollo de cáncer y la Anemia de Fanconi (Ma et al., 2012).

## **2.2. *PALB2* y predisposición al cáncer.**

Durante los últimos 20 años, la identificación de variantes patogénicas como causa hereditaria de susceptibilidad a cáncer de mama sucedía en una proporción muy pequeña de las mujeres que buscaban asesoramiento genético. Sin embargo, el desarrollo de la tecnología de secuenciación masiva (NGS) ha creado una revolución sin precedentes en el diagnóstico molecular de las enfermedades hereditarias, permitiendo mejorar significativamente el conocimiento existente. El rápido descenso en los costes de secuenciación y la posibilidad de producir grandes volúmenes de datos, ofrecen la posibilidad de estudiar una gran cantidad de genes de forma simultánea, convirtiéndola en una poderosa herramienta para la investigación genómica.

Existen un gran abanico de técnicas de secuenciación (NGS) en el diagnóstico de enfermedades genéticas que pueden ser utilizadas, pero últimamente el uso de paneles de genes ha adquirido una gran relevancia. Esta estrategia permite evaluar de forma simultánea múltiples genes implicados en una enfermedad, siendo especialmente útil cuando una condición clínica puede ser producida por alteraciones en diferentes genes (heterogeneidad genética). Esta técnica permite de una forma económica dirigir la secuenciación hacia un grupo de genes responsables de una patología determinada. Estos paneles pueden ser diseñados “a la carta” y permiten secuenciar genes completos, mutaciones puntuales, deleciones, inserciones y detectar variaciones en el número de copias y translocaciones.

En este contexto, el cáncer de mama y ovario hereditario representan un modelo muy interesante. En un principio, el diagnóstico del cáncer de mama y ovario hereditario estaba basado únicamente en 2 genes, *BRCA1* y *BRCA2* (Miki et al., 1994; Wooster et al., 1995). A día de hoy, varios equipos de investigación han reportado que al menos el 50% de las variantes encontradas en familias con cáncer de mama y ovario hereditario han sido identificadas en diferentes genes que *BRCA1* y *BRCA2* (Castéra et al., 2014). Así, Nielsen y col. propusieron hasta 26 genes BC/OC conocidos o candidatos como genes de susceptibilidad a cáncer de mama/ovario (Nielsen et al., 2016). En 2017 se publicó un estudio de más de 35.000 mujeres con cáncer de mama que fueron

testadas con un panel de 25 genes implicados en cáncer hereditario (Buys et al., 2017). En este estudio identificaron variantes patogénicas en alguno de estos genes y más de la mitad de dichas variantes se encontraron genes distintos a *BRCA1/2*, destacando entre ellos, *CHEK2* (11,7%), *ATM* (9,7%) y *PALB2* (9,3%).

El proyecto europeo BRIDGES (Breast Cancer After Diagnostic Gene Sequencing; <http://bridges-research.eu/>), en el cual nuestro laboratorio ha participado, ha constituido el mayor esfuerzo internacional para la definición de los genes BC/OC y el riesgo asociado a ellos, ya que se han secuenciado un panel de 34 genes confirmados o propuestos BC/OC en más de 113.000 mujeres (Dorling et al., 2021). Estos autores identificaron 5 genes, *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, y *PALB2*, asociados con un riesgo significativo de cáncer de mama en general. Asimismo, detectaron una asociación moderada en otros 7 genes: *BARD1*, *RAD51C*, *RAD51D*, *PTEN*, *NF1*, *TP53* y *MSH6*.

Otros estudios de secuenciación a gran escala han confirmado a *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, y *PALB2*, como el núcleo de susceptibilidad a cáncer de mama y ovario, a los cuales se podría unir *TP53*, aunque la prevalencia de variantes patogénicas en dicho gen es muy baja (Dorling et al., 2021; Foulkes, 2021; Hu et al., 2021; Narod, 2021).

En lo que respecta al gen *PALB2*, se ha visto claramente que existe una asociación entre las familias portadoras de variantes genéticas de pérdida de función (loss of function, LoF) y un aumento del riesgo relativo que varía de moderado a alto, de forma similar a lo que ocurre con las variantes de *BRCA2* (Antoniou et al., 2014; Easton et al., 2015; Couch et al., 2017). Variantes bialélicas LoF en *PALB2* causan Anemia de Fanconi (Tischkowitz and Xia, 2010), mientras que mutaciones germinales monoalélicas LoF en *PALB2* otorgaban un riesgo absoluto de cáncer de mama de 33%-58% o 44%-68% a los 70 años de edad dependiendo del historial familiar de cáncer de mama, de la cohorte y de la edad de aparición (Antoniou et al., 2014; Yang et al., 2020). Además, existe también un riesgo asociado a cáncer de páncreas, ocasionado por alteraciones genéticas en *PALB2*, siendo el segundo gen de susceptibilidad con más mutaciones causantes de este cáncer,

después de *BRCA2*. Asimismo se han visto alteraciones genéticas en *PALB2* asociadas a cáncer de próstata, ovario y mama masculina (Jones et al., 2009; Slater et al., 2010).

A pesar de que generalmente asociamos un gen determinado con un nivel de riesgo específico, la realidad es que el riesgo es otorgado por variantes concretas, y diferentes variantes de un mismo gen pueden aportar niveles de riesgo muy diversos. Así, por ejemplo, Rebbeck y colaboradores demostraron en 2015 como distintas mutaciones en *BRCA1* y *BRCA2* estaban asociadas a niveles de riesgo significativamente distintos en cáncer de mama y ovario (Rebbeck et al., 2015). Por otro lado, Spurdle y colaboradores demostraron que, el riesgo de cáncer de mama o de ovario hasta los 70 años es del 24% (IC del 95% del 10% al 40%) para los portadores de c.5096G>A, p.(Arg1699Gln) en *BRCA1* mientras que era del 58% (IC del 95% del 7% al 72%) para los portadores de c.5095C>T p.(Arg1699Trp) en *BRCA1*, que afecta al mismo codón que la anterior. Estos resultados se pueden comparar con el 4,6% de las mujeres en la población general y el 68% de las portadoras de una mutación patogénica promedio (Spurdle et al., 2012). De la misma manera, es lógico pensar que diferentes variantes genéticas en *PALB2* estén asociadas a niveles de riesgo muy distintos.

Así, en Finlandia la variante fundadora c.1592delT de *PALB2* ha sido identificada en aproximadamente el 1% de las mujeres con cáncer de mama que no fueron seleccionadas en base a un historial familiar positivo y está asociada con un riesgo relativo de 6, equivalente al riesgo en mujeres portadoras de variantes deletéreas de *BRCA2* en el mismo país (Erkko et al., 2008). Por otro lado, en Canadá la variante fundadora c.2323C>T ha sido detectada en el 5% de las mujeres francocanadienses con una aparición temprana de la enfermedad y que no fueron seleccionadas basándose en un historial familiar positivo. En este caso, es difícil estimar la penetrancia del gen *PALB2*. Esto se debe a que las frecuencias alélicas de las variantes de los controles son demasiado bajas para generar una estimación sólida del riesgo relativo (por ejemplo, en más de 6000 controles del estudio, no se encontró la variante c.2323C>T). En la población francocanadiense de Quebec, se han identificado mutaciones fundadoras en *BRCA1*, *BRCA2* y ahora *PALB2*. Sin embargo, la relevancia clínica completa de las mutaciones *PALB2* es aún incierta (Foulkes et al., 2007).

Las variantes de pérdida de función en *PALB2* han sido observadas en poblaciones muy heterogéneas y se ha estimado que, dependiendo de la población, entre un 0,6% y un 3,9% de las familias con historial de cáncer de mama son portadoras de variantes patogénicas en *PALB2*, con riesgos relativos de 3,83 y 5,02 en dos grandes estudios realizados en Estados Unidos y en Europa, respectivamente (Dorling et al., 2021; Hu et al., 2021). En conjunto, se encontraron variantes patogénicas de *PALB2* en el 0,5% de todos los casos de cáncer de mama (87.158 pacientes) en ambos estudios, mientras que el gen más prevalente fue *BRCA2* (1,3%). Estas variantes patogénicas de *PALB2* están fuertemente asociadas a cánceres de mama triple negativo y receptor de estrógenos negativo con un riesgo relativo de 10,36 y 7,35, respectivamente (Dorling et al., 2021). Además, las variantes germinales patogénicas en *BARD1*, *BRCA1*, *BRCA2*, *PALB2* y *RAD51D* han sido asociadas con un alto riesgo (*odds ratio* > 5,0) de TNBC y un riesgo de padecer cáncer de mama general superior al 20% entre las caucásicas. Las variantes patogénicas en *BRIP1*, *RAD51C* y *TP53* se asociaron con un riesgo moderado (*odds ratio* > 2) de TNBC (Shimelis et al., 2018).

### **2.3. *PALB2* y Anemia de Fanconi.**

La anemia de Fanconi (FA), es una enfermedad autosómica recesiva caracterizada por asociarse a características clínicas que a menudo incluyen anomalías del desarrollo que afectan al esqueleto (pulgares y radios ausentes o anormales), riñones, corazón o cualquier otro sistema orgánico importante, además de fragilidad cromosómica, deficiencias en el desarrollo de la médula espinal, anomalías congénitas y predisposición al cáncer, incluyendo leucemias y tumores sólidos (Joenje and Patel, 2001). Esta enfermedad fue reportada por primera vez en 1927 por el pediatra sueco Guido Fanconi. Guido describió una familia con 3 niños afectados por una combinación extraña de anomalías somáticas con anemia aplásica progresiva (Lobitz and Velleuer, 2006). A nivel celular y cromosómico, la característica principal de FA y enfermedades similares es la hipersensibilidad a los entrecruzamientos intercatenarios de DNA que representan la base para el diagnóstico de este trastorno (Gueiderikh et al., 2017).

La anemia de Fanconi afecta a 1 de cada 136.000 neonatos, sin discriminar por grupos étnicos o razas, con una frecuencia estimada de portadores heterocigotos de entre un 0,3 y un 1% (Verlander et al., 1995; Mamrak et al., 2017). La esperanza de vida de los pacientes se reduce a una media de 20 años, principalmente porque estos individuos desarrollan insuficiencias en la médula ósea potencialmente mortales y son susceptibles de desarrollar neoplasias malignas, especialmente leucemia mieloide aguda y en menor medida tumores sólidos (Auerbach and Allen, 1991).

Inmediatamente después de que *PALB2* fuera descubierto, se identificaron variantes patogénicas de este gen en 7 familias con anemia de Fanconi y mutaciones bialélicas de pérdida de función; como c.395delT y c.3113+5G>C. El fenotipo celular de pérdida de función para *PALB2* es similar al producido por *BRCA2* y más grave que otros subtipos de anemia de Fanconi, con aumento de la tasa de rotura cromosómica espontánea, una supervivencia de linfocitos notablemente reducida y un aumento de la rotura de cromosomas por exposición a quimioterapia con mitomicina C (MMC). Además, Reid y colaboradores observaron como los fibroblastos deficientes de *PALB2*, no interaccionaban con RAD51, de la misma forma que ocurre en las células con mutaciones bialélicas de pérdida de función en *BRCA2* y difiriendo de otros tipos de anemia. En el estudio indicado, los siete individuos con anemia de Fanconi desarrollaron cánceres durante su infancia; incluidos 3 tumores de Wilms, 5 meduloblastomas, dos casos de leucemia mieloide aguda y un neuroblastoma (Reid et al., 2007; Xia et al., 2007).

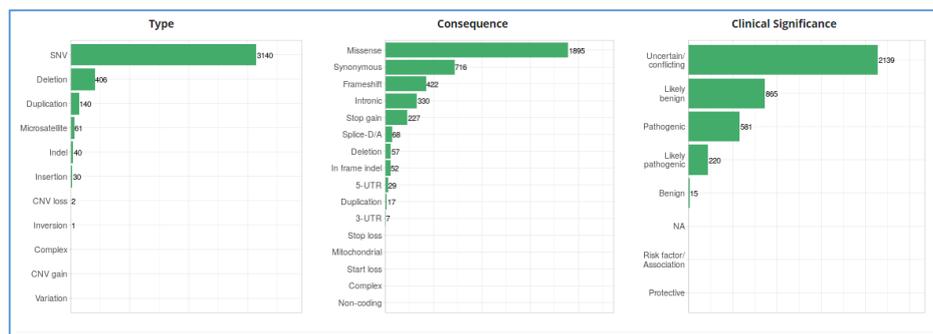
#### **2.4. Variantes de *PALB2*.**

Una variante genética es un cambio que se produce en la secuencia de DNA. De acuerdo al impacto predecible en la traducción a proteína, existen diferentes tipos de variantes:

- *Missense*: variante que provoca un cambio de aminoácido en la proteína.
- *Nonsense*: variante que causa el cambio de un codón de un aminoácido por otro de parada prematura (*Premature Termination Codon –PTC-*).
- *Frameshift*: inserciones o deleciones que alteran la pauta de lectura generando un PTC.

- Sinónima: sustitución de un nucleótido que no provoca un cambio de aminoácido.
- *In-Frame*: inserciones o deleciones que no alteran la pauta de lectura (múltiplos de 3).

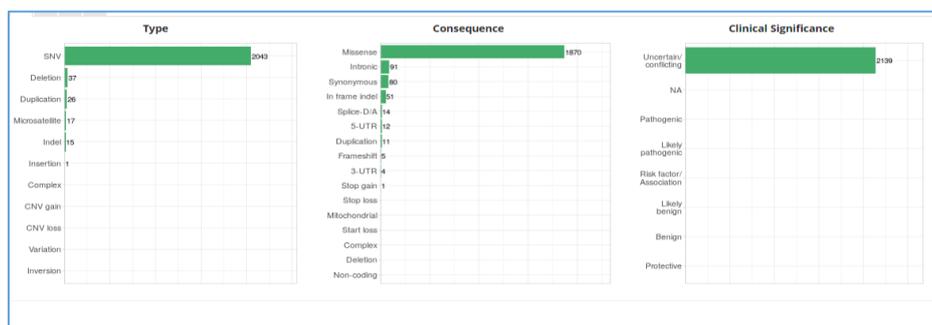
Según la base de datos ClinVar y Simple-ClinVar (<http://simple-clinvar.broadinstitute.org/>, último acceso 3 de agosto de 2021; <https://www.ncbi.nlm.nih.gov/clinvar/?term=palb2%5Bgene%5D>; Figura 3), se han reportado 3820 variantes diferentes en *PALB2*, que se distribuyen del siguiente modo: 3140 sustituciones de nucleótido (82,2%), 406 deleciones (10,6%), 140 duplicaciones (3,7%) y el resto son alteraciones poco frecuentes. De acuerdo al impacto predecible en la proteína, se han registrado 1895 variantes missense (49,6%), 716 sinónimas (18,7%), 422 *frameshift* (11%), 330 intrónicas (8,6%), 227 *nonsense* (5,9%), y sólo 68 de sitios de *splicing* (1,8%) que presumiblemente alterarían el procesamiento de mRNA. A su vez, han sido reportadas variantes en regiones que no se traducen en el 5' y el 3' del transcrito de referencia (*Untranslated Regions*, UTRs).



**Figura 3. Distribución de las variantes de *PALB2* registradas en la base de datos ClinVar.** Captura de pantalla de Simple ClinVar del gen *PALB2* (<http://simple-clinvar.broadinstitute.org/>).

Centrándonos en su significado clínico, 880 (23%) aparecen clasificadas como benignas o probablemente benignas, 801 (21%) como patológicas o probablemente patológicas, mientras que 2139 variantes (56%) se catalogaron como variantes de significado clínico incierto (*variant of uncertain significance*, **VUS**), debido a que no se dispone de suficientes datos para determinar su relevancia clínica. Hay variantes genéticas que son clasificadas directamente como patológicas o probablemente patológicas basándose únicamente en la naturaleza de su cambio, como son las sustituciones y pequeñas deleciones e inserciones que introducen PTCs. Según ClinVar, la mayoría

de las variantes patogénicas en *PALB2* son variantes *frameshift* (~52%) seguidas de variantes *nonsense* (~28%). Sin embargo, la gran mayoría de VUS son variantes *missense* (1870 de 2139 VUS; Figura 4)



**Figura 4. Distribución de variantes de significado clínico desconocido de *PALB2* en ClinVar.** Captura de pantalla de Simple ClinVar del gen *PALB2* (<http://simple-clinvar.broadinstitute.org/>).

Las VUS complican el consejo genético y el manejo clínico de las pacientes ya que no se les puede asignar un riesgo asociado a la enfermedad, de modo que la toma de decisiones (medidas preventivas y/o terapéuticas) se basará únicamente en su historia familiar (Antoniou et al., 2014).

La clasificación clínica de VUS constituye un reto para el consejo genético de las enfermedades hereditarias y específicamente en CMOH, habiendo recibido, en los últimos años, especial atención por parte de un gran número de grupos de investigación. Una de las herramientas utilizadas para estudio de VUS son los ensayos funcionales. Mediante estas técnicas podemos clasificar variantes que provocan una pérdida de función en los genes de susceptibilidad, ya que una disminución en la actividad de estos genes podría estar asociada a un mayor riesgo de cáncer. Por lo tanto, los ensayos funcionales, que pueden establecer y cuantificar las alteraciones en la actividad de una variante, aportan una información esencial para evaluar su patogenicidad (Couch et al., 2008; Fraile-Bethencourt et al., 2019).

Actualmente, no existe un sistema universal de clasificación de variantes de *BRCA1/2* o de *PALB2*, aunque existen aproximaciones utilizadas para clasificar las variantes encontradas en los *BRCA1/2*. Estas aproximaciones se basan en análisis multifactoriales que incluyen estudios de cosegregación, análisis poblacionales, historial familiar, análisis del tumor (pérdida de

heterocigosidad), ensayos funcionales y predicciones bioinformáticas de cambio de aminoácido o impacto en el *splicing* (Goldgar et al., 2004, 2008). La Agencia Internacional para la Investigación del Cáncer (IARC) presentó en 2008 un sistema de clasificación de variantes en genes BC/OC de alto riesgo basado en dicho abordaje de probabilidad multifactorial. Este método, propuesto por el panel de expertos de la IARC, integra datos independientes, para obtener una probabilidad final de patogenicidad. Recomienda clasificar clínicamente las variantes genéticas mediante un sistema de 5 categorías (Rilem, 2020):

- **Clase 1:** la probabilidad de patogenicidad es  $<0,001$ . Sin relevancia clínica.
- **Clase 2:** la probabilidad de patogenicidad es 0,001-0,049. El manejo clínico es similar a una variante de clase 1.
- **Clase 3:** la probabilidad de patogenicidad es 0,05-0,949. El manejo clínico depende del historial familiar y otros factores de riesgo. Estrictamente hablando equivalen a las VUS.
- **Clase 4:** la probabilidad de patogenicidad es 0,95-0,99.
- **Clase 5:** la probabilidad de patogenicidad es  $>0,99$ . Las clases 4 y 5 permiten la puesta en marcha de los protocolos de prevención.

En el contexto de la necesidad de crear unas reglas de clasificación, surgió la iniciativa paralela del consorcio ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles; <https://enigmaconsortium.org/>) que están basados en los criterios del IARC, definiendo las cinco clases indicadas arriba. Últimamente, han adquirido especial relevancia los criterios propuestos por el “American College of Medical Genetics and Genomics and the Association for Molecular Pathology” (ACMG/AMP). Como en los anteriores casos, define 5 categorías clasificatorias: Patogénica, Probablemente Patogénica, VUS, Probablemente Benigna y Benigna (Richards et al., 2015). La guía ACMG permite la clasificación de variantes identificadas en cualquier gen, incluyendo *BRCA1/2* y *PALB2*. Aún así, estos criterios son objeto de una permanente discusión y revisión, así como de modificaciones en los estándares de clasificación en

función de las características específicas del gen en estudio y la enfermedad (Abou Tayoun et al., 2018a; Lee et al., 2018).

### 3. Dogma central de la Biología Molecular.

El dogma central de la biología molecular se basa en el concepto enunciado por Francis Crick, para explicar el flujo de información genética en las células (Crick, 1970).

El DNA está formado por genes que contienen las instrucciones necesarias para la síntesis proteica. La información contenida en el DNA se utiliza como molde, transcribiéndose para formar el RNA en el núcleo, que a su vez se traduce en el ribosoma dando lugar a las proteínas (Figura 5). Las proteínas determinan la estructura y función de las células (unidad funcional más pequeña de un ser vivo).

No obstante, actualmente, el dogma central ha sido modificado después de haberse observado el mecanismo de transcripción inversa, que ocurre en algunos virus de RNA, como el virus de inmunodeficiencia humana (VIH), que para infectar las células necesitan transcribir el RNA en una molécula de DNA. Asimismo, se ha visto que es posible traducir moléculas de DNA y obtener proteínas (Figura 5B).

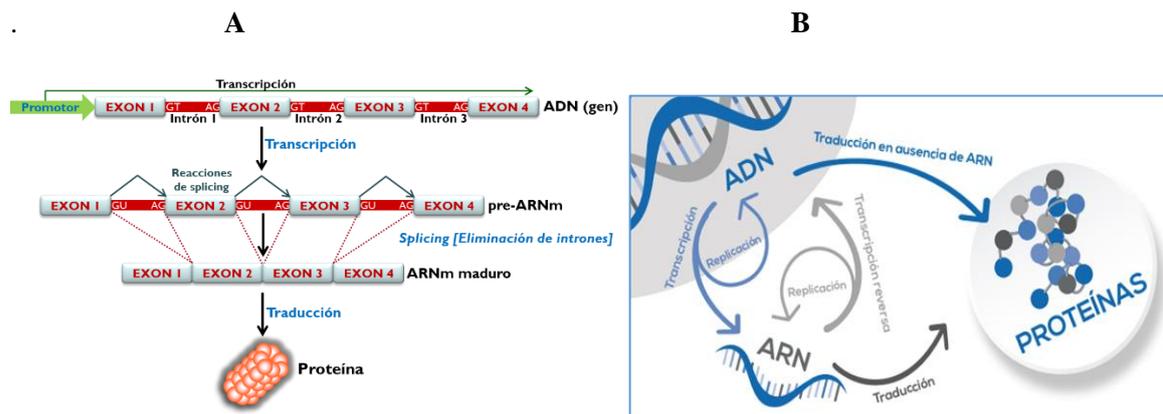


Figura 5 A y B. Esquema representativo del dogma central de la biología molecular.

Sin embargo, la expresión génica es un proceso mucho más complejo, en el cual se pueden distinguir otros mecanismos esenciales. Así, el *splicing* es un proceso que se produce en el núcleo de cualquier célula eucariota y es llevado a cabo por un complejo de ribonucleoproteínas denominado *esplíceosoma* (Wahl et al., 2009). Se basa en el corte y la eliminación de las secuencias no codificantes (intrones) y la unión secuencial de las secuencias codificantes para proteína (exones; Figura 5A). El mecanismo de *splicing* fue descubierto en 1977 (Berget et al., 1977; Chow et al., 1977), hallazgo que les permitió a Phillip A. Sharp y Richard J. Roberts ganar el Premio Nobel de Fisiología o Medicina en 1993. Tres años después Roger D. Kornberg ganó el Premio Nobel de Química por sus estudios sobre la estructura tridimensional de la RNAPol II, enzima encargada de producir el RNA mensajero (mRNA) a partir de una de las cadenas de DNA uniéndose a su secuencia promotora. Además de la RNAPol II, existen otras enzimas de la misma familia con diferentes funciones, como la RNA polimerasa I, que transcribe genes que van a producir los RNA ribosómicos 18S y 28S (Merkl et al., 2020) y la RNA polimerasa III que tiene un importante papel en eucariotas transcribiendo los RNAs más abundantes como los RNAs de transferencia y RNAs pequeños no codificantes, muchos de los cuales no tienen funciones relacionadas con los ribosomas ni con la síntesis de proteínas (Willis and Moir, 2018).

### **3.1. Maduración del mRNA o *splicing*.**

La mayor parte de los genes eucariotas contienen intrones (secuencias no codificantes), que se transcriben, pero no se traducen, y exones (secuencias codificantes) distribuidos de forma alterna. El RNA producido en el proceso de transcripción por la RNAPol II, se conoce como RNA precursor (pre-RNA) que dará lugar a un RNA maduro capaz de ser traducido después de las modificaciones postranscripcionales y el proceso de *splicing* (Gilbert, 1978). El *splicing* es un proceso altamente regulado en el que los exones e intrones deben ser reconocidos de una manera precisa. Los pre-RNA de los organismos vertebrados facilitan la información necesaria para el reconocimiento de un exón por parte de la maquinaria de *splicing* a través de pequeñas secuencias intrónicas altamente conservadas en la evolución, como se muestra en la Figura 6.

### 3.1.1. El *spliceosoma*.

En 1985 Beyer y Osheim observaron por primera vez la estructura del *spliceosoma* mediante microfotografías y vieron cómo determinadas partículas proteicas se unían al mRNA durante el proceso de transcripción. Estas partículas parecían fusionarse entre sí y al hacerlo generaban un lazo que provocaba el corte y la liberación de los intrones. En una gran proporción de los organismos eucariotas coexisten dos tipos diferentes de *spliceosomas*, siendo mucho más frecuente el reconocimiento de los exones por el *spliceosoma* mayor (> 99 veces) que por el menor (Beyer and Osheim, 1988).

El *spliceosoma* no es una enzima preensamblada, sino que está formada por 5 pequeños RNA nucleares (snRNAs) y numerosas proteínas asociadas. Además, es un gran complejo molecular ribonucleoproteico que se somete a numerosos reordenamientos estructurales que conducen a la formación de su centro activo (Kastner et al., 2019). Los snRNAs U1, U2, U4 y U5 son sintetizados por la RNA polimerasa II y sufren una modificación post transcripcional a la que se le añade un casquete de tri-metil-guanosina, mientras que U6 es transcrito por la RNA polimerasa III y adquiere un casquete de  $\gamma$ -monometil guanosina (Reddy and Busch, 1988). El *spliceosoma* formado por los snRNAs U1, U2, U4, U5 y U6 constituiría el denominado *spliceosoma* mayor.

La gran mayoría de los intrones tienen sitios canónicos 5' de *splicing* que comienzan con GT y sitios de *splicing* 3' que terminan en AG (Figura 6). Sin embargo, también existen una pequeña proporción de intrones con sitios de *splicing* no canónicos AT 5' y AC 3' en genes eucariotas (Jackson, 1991; Hall and Padgett, 1994). Estos intrones son denominados intrones U12, son catalizados por el llamado *spliceosoma* menor y representan menos del 0,5% de todos los intrones de cualquier genoma (Wu and Krainer, 1999) (Figura 7).

Existe una alta conservación evolutiva en las secuencias de sitio donador y del punto de ramificación, y aunque la mayoría de las ribonucleoproteínas implicadas son específicas de este complejo (U11, U1, U4atac y U6tac), comparte la snRNP U5 con el *spliceosoma* mayor. Por otro lado, algunos autores defienden que el *spliceosoma* mayor puede actuar también fuera del núcleo

(König et al., 2007). Sin embargo, no se han encontrado evidencias sobre la posibilidad de que los eventos de *splicing* puedan producirse total o parcialmente en el citoplasma (Steitz et al., 2008).

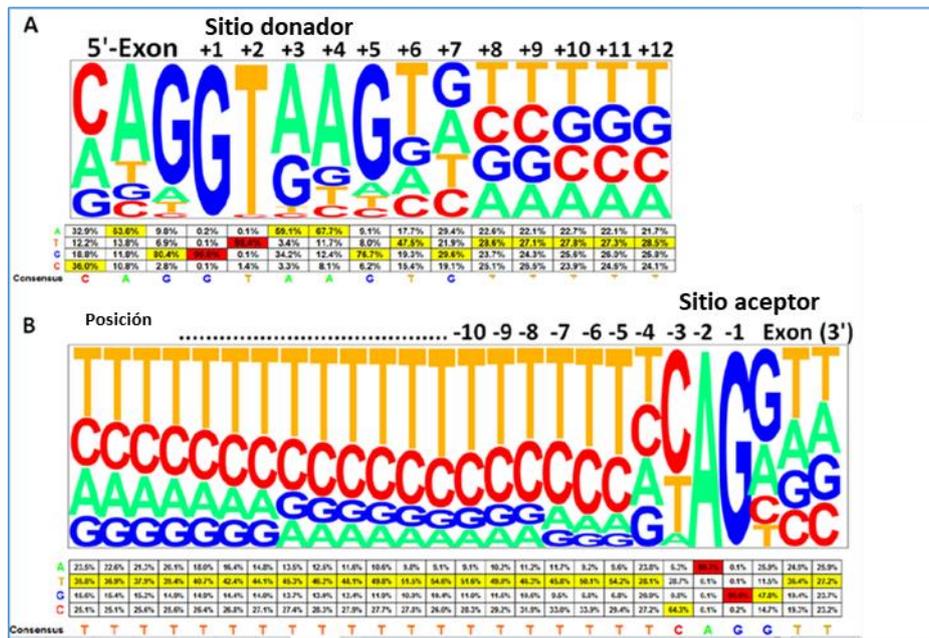


Figura 6. Secuencias básicas de reconocimiento de intrones por el *spliceosoma*. (Ma et al., 2015).

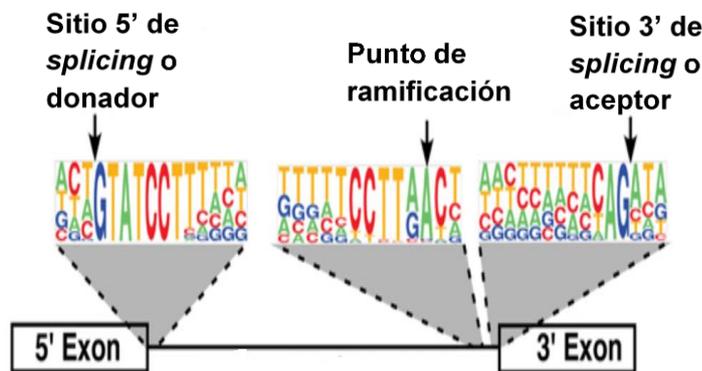


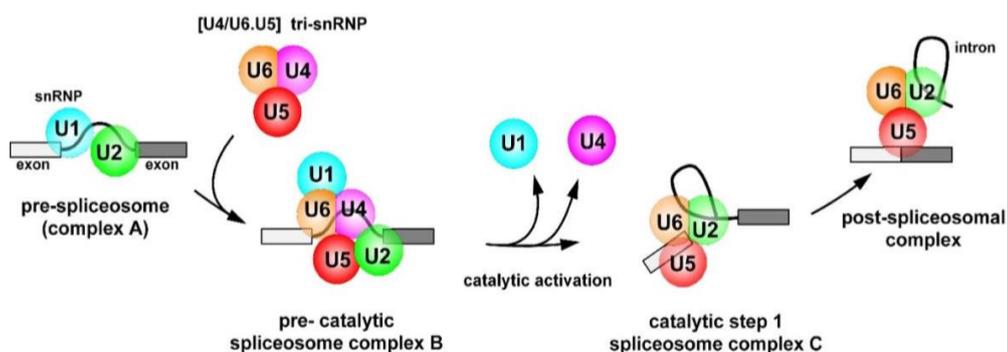
Figura 7. Secuencias básicas de reconocimiento de intrones por el *spliceosoma* menor. (Ma et al., 2015).

### 3.1.2. Conformación dinámica del *spliceosoma*: La reacción de *splicing*.

Para llevar a cabo el proceso de *splicing*, el *spliceosoma* debe ensamblarse y formar el complejo catalítico C (Figura 8). Los snRNPs U1 y U2 reconocen el sitio 5'ss y la secuencia de

pares de bases respectivamente, formando el pre-*espliceosoma* o complejo A. El pre-*espliceosoma* se asocia después con el tri-snRNP preensamblado U4/U6/U5 (complejo B) para formar el *espliceosoma* completo. U6 se pliega en última instancia para formar el centro activo del *espliceosoma* y se encuentra emparejado por sus bases con U4 dentro del tri-snRNP (Wilkinson et al., 2020).

El complejo B necesita un aporte energético para ser catalíticamente activo y va a conseguirlo a partir del ATP, cuando esto sucede, comienza a catalizar las reacciones de transesterificación y pasa a llamarse complejo C, después de que U1 y U4 sean liberados. Una vez se ha formado correctamente el complejo C, se producen dos reacciones de transesterificación. En la primera de ellas, se va a producir un corte en la secuencia de RNA a causa del ataque nucleofílico de la adenina del punto de ramificación al 5' ss. La segunda reacción de transesterificación propicia la unión de los extremos exónicos entre el grupo OH libre del 5' y el 3' ss. Por último, se disocia el complejo y se liberan los snRNPs (Will and Luhrmann, 2011).



**Figura 8. Ensamblaje del *espliceosoma* y reacción de *splicing* (Adaptación de Will and Luhrmann, 2011).**

### 3.1.3. Los sitios de *splicing*.

Las secuencias básicas de *splicing* que van a permitir el correcto funcionamiento del proceso son: el sitio 5' de *splicing* (5' ss) o donador, el sitio 3' de *splicing* (3' ss) o aceptor, el tracto de polipirimidina y el punto de ramificación situado a unos 40 nucleótidos del extremo 3' del intrón

(Figura 6) (Wu and Krainer, 1997). En el sitio donador o 5' ss, el intrón comienza con el dinucleótido GT (GU en el RNA) que está acompañado de nucleótidos conservados tanto en el extremo 3' del exón como en los nucleótidos +3 a +6 intrónicos (Exón/Intrón: MAG/GTRAGT; M=C/A; R=A/G).

En el sitio aceptor o 3' ss del intrón encontramos siempre el dinucleótido AG precedido por una pirimidina esencial en posición -3 (fundamentalmente una C), un nucleótido no conservado en -4, y una cantidad variable de pirimidinas (~11) en las posiciones intrónicas de -5 a -15, que se denomina tracto de polipirimidinas. Asimismo, los dos primeros nucleótidos exónicos están moderadamente conservados, formando la siguiente secuencia consenso del sitio aceptor: Y<sub>n</sub>NCAG|GT (Intrón|Exón). Por tanto, además de las posiciones +1, +2 (GT) y -1,-2 (AG), existen otras altamente conservadas, de modo que cualquier variante en esos nucleótidos puede provocar anomalías en los patrones de *splicing*. Las alteraciones producidas en el proceso de *splicing* están implicadas de forma directa en la gran mayoría de las enfermedades de etiología genética y el conocimiento de las anomalías que evitan el correcto procesamiento del mRNA resulta de gran importancia en la práctica clínica (Baralle et al., 2009).

El dinucleótido GT situado en las posiciones intrónicas +1 y +2 es el elemento más conservado del sitio 5' de *splicing* o donador y se encuentra representado en el 99% de los donadores en los mamíferos. Sin embargo, en una pequeña proporción de casos (<1%), el dinucleótido GT es reemplazado por GC. En los donadores GC la conservación de las bases es mucho más estricta, posiblemente para compensar este cambio en una posición crítica, que sería desfavorable para su reconocimiento y procesamiento (Burset et al., 2000). Los sitios de *splicing* GT-AG y GC-AG son procesados por el *spliceosoma* de tipo U2 estándar (Aebi et al., 1987; Burge et al., 1999; Mount, 2000), existiendo un par de bases de desajuste en la interacción entre donador GC y el snRNA de U1. Por tanto, se ha propuesto que para compensar este desajuste, la secuencia consenso alrededor del donador GC ha evolucionado para maximizar la interacción con otras posiciones en el snRNA de U1 (Jackson, 1991; Burge et al., 1999; Burset et al., 2000; Mount, 2000), aunque los factores externos que facilitan la selección precisa de un donador GC en relación a un donador GT más fuerte están aún por descubrir.

Existe un enriquecimiento sustancial de GC en las regiones donadoras de genes con *splicing* alternativo en humanos, nemátodos y en arábidopsis, sugiriendo que son muy importantes para la regulación del *splicing* alternativo (Churbanov et al., 2008). De hecho, estos donadores son utilizados más frecuentemente por exones con *splicing* alternativo (1 de cada 20 exones humanos) a diferencia de aquellos con *splicing* constitutivo (1 de cada 200 exones humanos) (Thanaraj and Clark, 2001). Tanto el exón 17 de *BRCA2* como el exón 12 de *PALB2* utilizan donadores GC. En *BRCA2*, se han encontrado isoformas fisiológicas alternativas como  $\Delta(E18)$  (liberación del exón 18) y  $\Delta(E17\_E18)$  (liberación de los exones 17 + 18) (de Garibay et al., 2014; Fackenthal et al., 2016). En *PALB2* se han encontrado 44 eventos de *splicing* alternativo fisiológico, de los cuales 6 afectan directamente al exón 12:  $(\Delta(E11)+\blacktriangledown(E12p446))$ ,  $(\Delta(E11)+\blacktriangledown(E12p65))$ ,  $(\blacktriangledown(E12p65))$ ,  $(\Delta(E12p136))$ ,  $(\Delta(E12))$  y  $(\Delta(E11\_E12))$  (Lopez-Perolio et al., 2019). La mayoría de estas isoformas contribuyen de una manera menor a la expresión (<1%), sin embargo, la expresión combinada de  $(\Delta(E11\_E12))$ ,  $(\Delta(E12))$  y  $(\Delta(E11))$  representa entre el 8% y el 34% de la expresión génica general de *PALB2*, aunque las proteínas codificadas por estos dos transcritos no serían funcionales, porque carecen de parte del dominio C-terminal WD40 (que media la interacción de PALB2 con otras proteínas clave en la recombinación homóloga, como BRCA2 o RAD51). A pesar de este dato, no podemos descartar la posibilidad de pensar que las variantes *frameshift* en los exones 11 y/o 12 estuvieran asociadas un menor riesgo de cáncer que este mismo tipo de variantes en otros exones.

### **3.1.4. *Splicing* alternativo.**

Antes de la finalización del proyecto genoma humano, se creía que las 900.000 tipos de proteínas diferentes presentes en el ser humano eran producidas por el mismo número de genes (Valdivia, 2007). Era lógico pensar que como el genoma humano es estructural y funcionalmente más complejo que el de otros organismos más simples, poseería una cantidad muy superior de genes. Sin embargo, se ha visto que el ser humano posee menos de 25.000 genes mientras que el maíz, por poner un ejemplo, cerca de 40.000 (Valdivia, 2007), y produce 81.635 proteínas, según la base de datos Uniprot (<https://www.uniprot.org/proteomes/UP000007305>; último acceso 25 de

agosto de 2021). Así que el paradigma inicial de que un gen producía una proteína cambió por el concepto del *splicing* alternativo.

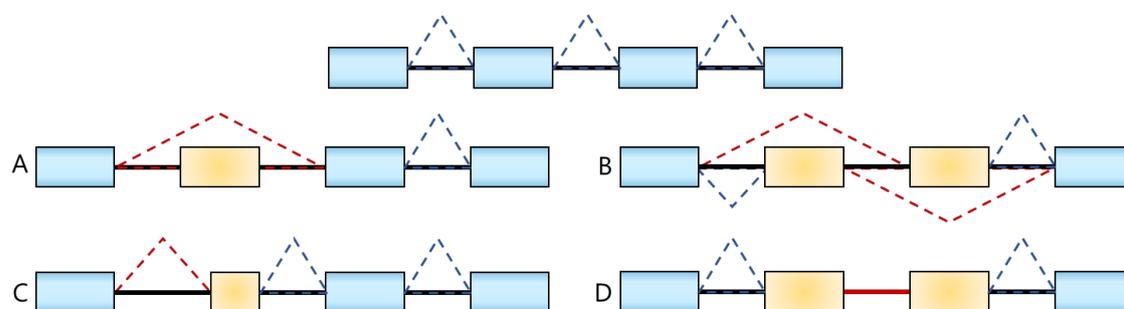
Para explicar la mayor complejidad genómica de los seres humanos en relación con otros organismos multicelulares se plantearon tres hipótesis:

1. La primera defendía la existencia de un alto nivel de regulación de los genes y de los mecanismos involucrados (Ewing and Green, 2000; Venter et al., 2001).
2. La segunda hipótesis sugería que existían más de 200 modificaciones postraduccionales y que eran las responsables de la producción de las diferentes proteínas.
3. La tercera hipótesis estaba relacionada con el *splicing* alternativo, que explicaba de manera eficaz como un único gen era capaz de codificar para múltiples proteínas.

En el proceso de *splicing* alternativo, se generan distintas moléculas de mRNA a partir del mismo gen, conociéndose estas moléculas como isoformas. El 92-94% de los genes en humanos, presentan *splicing* alternativo, explicando la gran riqueza proteómica (Wang et al., 2008). Además, el *splicing* alternativo cumple una función esencial en los procesos de desarrollo y diferenciación celular, y es regulado en muchos casos por factores de *splicing* específicos de tejido (Lee and Rio, 2015).

Existen diferentes patrones de *splicing* alternativo. El evento conocido como exón *skipping* (Figura 9A), o pérdida completa de un exón, que tiene lugar debido a la ausencia de reconocimiento de los sitios canónicos por el *spliceosoma* y como consecuencia la eliminación de uno o varios exones. También se originan diferentes eventos de *splicing* por el uso de sitios alternativos de *splicing* (Figura 9C), que se producen cuando el *spliceosoma* reconoce preferentemente sitios diferentes de los sitios canónicos de *splicing*. La retención intrónica (Figura 9D), aunque es difícil observar en condiciones fisiológicas, es otro evento de *splicing* alternativo y ha sido observado como mecanismo de inactivación de genes supresores de tumores en células tumorales (Jung et al., 2015). Por último, existe el evento de *splicing* alternativo denominado exones mutuamente

excluyentes (Figura 9B), que ocurre cuando se incluye únicamente un exón u otro, pero nunca los dos en el mismo transcrito. Este evento generalmente se produce con el primer y el último exón, por lo que encontramos isoformas con sitios de inicio de la transcripción alternativos o con sitios alternativos de poliadenilación.



**Figura 9. Patrones de *splicing* alternativo. A) Pérdida del exón o *exon skipping*. B) Exones mutuamente excluyentes. C) Uso de sitio de *splicing* alternativo. D) Retención intrónica.**

*PALB2*, como la mayoría de los genes humanos, sufre *splicing* alternativo. En 2019, López-Perolio y colaboradores publicaron un catálogo de isoformas alternativas de *PALB2* encontradas en pacientes sanos, caracterizando 88 isoformas naturales de *PALB2* en diferentes tipos celulares; siendo éstas las más representativas:  $\nabla$ (E1q9),  $\Delta$ (E1q17),  $\Delta$ (E2p6),  $\Delta$ (E2),  $\Delta$ (E4),  $\Delta$ (E5p139),  $\Delta$ (E7p10),  $\Delta$ (E7),  $\Delta$ (E8),  $\Delta$ (E9),  $\Delta$ (E10),  $\Delta$ (E9\_10),  $\Delta$ (E10p2),  $\Delta$ (E10q31),  $\Delta$ (E11),  $\Delta$ (E12) y  $\Delta$ (E11-E12) (Lopez-Perolio et al., 2019). Curiosamente, todos los exones de *PALB2* se ven afectados por uno o más eventos de *splicing* alternativo, lo que sugiere que ningún exón de *PALB2* debe anotarse como constitutivo. La producción y detección de transcritos *in-frame* es crítica para la interpretación clínica de las variantes en este gen, ya que no se puede asumir directamente que los transcritos *in-frame* inactiven *PALB2*. Un claro ejemplo de esto ocurre en el gen supresor de tumores *BRCA1*, donde la variante c.594-2A>C produce el transcrito alternativo *in-frame*  $\Delta$ (E9\_E10), que es funcional y suficiente para una correcta función celular de *BRCA1*, explicando así el bajo riesgo de cáncer de mama observado en las portadoras de esta variante (de la Hoya et al., 2006). Las isoformas alternativas  $\Delta$ (E2p6),  $\Delta$ (E6),  $\Delta$ (E5p24) y  $\Delta$ (E10p3) de *PALB2* podrían codificar proteínas funcionales o parcialmente funcionales, pero su contribución respecto a la

expresión general de *PALB2* (<1%) es demasiado baja para considerarlas “isoformas de rescate”, llamadas así porque retienen total o parcialmente la funcionalidad del transcrito canónico.

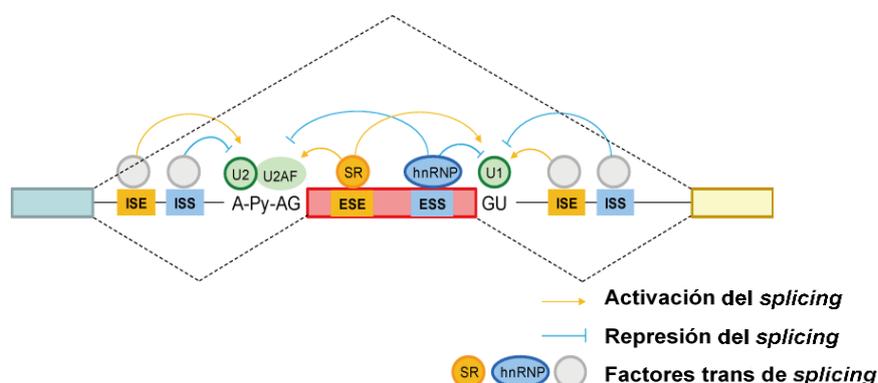
Como acabamos de ver, no todas las isoformas de *splicing* son funcionales, estimándose que un 2% de los eventos de *splicing* alternativo son productos de errores en la maquinaria de *splicing* (Pickrell et al., 2010). De esta manera, el *splicing* alternativo, además de aportar diversidad proteica, regula la expresión de los niveles de proteína, ya que los transcritos con codones stop prematuros (PTCs), serán degradados a través de la ruta del *Nonsense Mediated Decay* (NMD) (Kervestin and Jacobson, 2012).

### 3.1.5. Secuencias reguladoras de *splicing*.

El *splicing* es un proceso esencial que ocurre en el núcleo de las células eucariotas. Este mecanismo, elimina los intrones y une los exones para crear mRNA maduro, proceso que es catalizado por la maquinaria de *splicing* o *spliceosoma*, un complejo enzimático dinámico de RNA-proteína (Wahl et al., 2009; Papasaikas and Valcárcel, 2016). El *splicing* alternativo es un proceso altamente regulado que permite la producción un gran número de isoformas de mRNA a partir de un único pre-mRNA. De hecho, el análisis del transcriptoma a gran escala reveló que hasta el 90% de los genes humanos presentan *splicing* alternativo (Pan et al., 2008). Por lo tanto, el *splicing* alternativo explica en gran medida la diversidad del transcriptoma y del proteoma generados a partir de un número limitado de genes en un genoma (Maniatis and Tasic, 2002; Weatheritt et al., 2016).

Además de los sitios naturales o canónicos de *splicing*, existen otros elementos auxiliares en *cis* que actúan como reguladores del *splicing* (*Splicing Regulatory Elements*, SRE). Son sitios de unión degenerados donde diferentes proteínas pueden unirse y regular el proceso de *splicing* mediante su interacción con el *spliceosoma* (Figura 10) (Smith and Valcárcel, 2000; Wang and Burge, 2008). Las proteínas de unión a RNA (RNA Binding Proteins, RBP) reguladoras de *splicing*, reconocen distintas secuencias de RNA (SRE) y regulan el *splicing* mediante un "código" (Fu, 2004; Wang and Burge, 2008). Dependiendo, de si su localización es exónica o intrónica y del

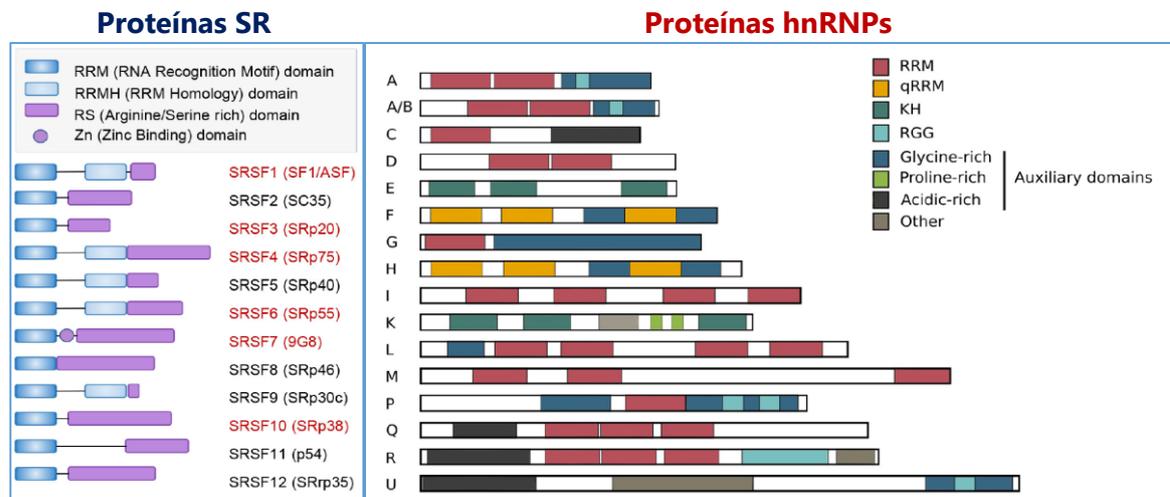
efecto positivo o negativo que ocasionan sobre el reconocimiento de los sitios de *splicing*, los podemos clasificar como *enhancers*, o silenciadores, respectivamente. Así, se pueden distinguir activadores y represores de *splicing* exónicos (ESE y ESS, respectivamente), activadores y represores de *splicing* intrónicos (ISE e ISS, respectivamente). En general, las proteínas SR se unen a ESEs y promueven la inclusión de un exón mediante el reclutamiento del *spliceosoma*. Por el contrario, las proteínas heterogéneas de la familia de las ribonucleoproteínas nucleares (hnRNP) pueden antagonizar el efecto positivo de las proteínas SR al unirse a ESS (silenciadores exónicos de *splicing*) o ISS (silenciadores intrónicos de *splicing*) y reprimir el *splicing* (Geuens et al., 2016).



**Figura 10. Regulación del proceso de *splicing* por elementos en *cis* y factores trans de *splicing* (Proteínas SR, hnRNP, etc).** Human-transcriptome database for alternative *splicing*, [http://www.h-invitational.jp/h-dbas/as\\_mechanism.jsp](http://www.h-invitational.jp/h-dbas/as_mechanism.jsp).

Las proteínas SR constituyen una familia de proteínas de unión a RNA (RNA Binding Proteins) o factores de *splicing* que están presentes en todos los metazoos y en plantas. Específicamente, contienen 1-2 dominios N-terminales de reconocimiento del RNA (RNA Recognition Motif, RRM) y una región C-terminal enriquecida en el dipéptido Arg-Ser (Figura 11) (Shepard and Hertel, 2009; Manley and Krainer, 2010). En general, los dominios RRM reconocen RNA, mientras que los dominios RS participan en diversas interacciones proteína-proteína y proteína-RNA. La mayoría de las proteínas SR están ubicadas exclusivamente en el núcleo, pero

algunas proteínas SR (principalmente SRSF1, SRSF3 y SRSF7) pueden desplazarse entre el núcleo y el citoplasma.



**Figura 11. Estructura de las proteínas SR y hnRNPs.** Lista y dominios de proteínas SR. Las estructuras de los dominios (RRM, RRMH, RS y Zn) se indican como se muestra en el cuadro superior. Los nombres actuales de las proteínas SR son SRSF, pero también se indican entre paréntesis. Entre las 12 proteínas SR centrales, 6 (en letras rojas) se desplazan entre el núcleo y el citoplasma (proteínas SR que se desplazan), mientras que las otras (letras negras) no han demostrado tener actividad de transporte (proteínas SR que no se desplazan) (Jozwiak et al., 2008). La familia hnRNP. Los hnRNP tienen diferentes pesos moleculares que van de 34 a 120 kilodaltons (kDa) y se nombran alfabéticamente desde hnRNP A1 a hnRNP U. Muchos hnRNP están presentes en los mismos complejos, lo que sugiere que múltiples hnRNP comparten estructura y función. Varios dominios estructurales se comparten entre diferentes miembros de la familia. Los miembros de la familia hnRNP están formados por cuatro dominios de unión a RNA únicos (RBDs). Diferentes miembros de la familia portan los mismos RBD, lo que explica en parte sus propiedades compartidas de unión de RNA: motivo de reconocimiento de RNA (RRM), motivo de reconocimiento de quasi-RNA (qRRM), dominio de homología K (KH), y dominio de unión de RNA (RGG) que consta de repeticiones de Arg-Gly-Gly. Los tamaños de estas 16 hnRNP comunes están dibujados a escala entre ellos en esta representación (Jeong, 2017).

Las proteínas SR también participan en un conjunto diverso de procesos celulares adicionales, incluida la exportación nuclear de mRNA, la estabilidad y el control de calidad del mRNA, la traducción, el mantenimiento de la estabilidad genómica, la transformación oncogénica y, además aumentan la inclusión de exones al servir como sitios para el reclutamiento de la maquinaria de *splicing*, el *espliceosoma*. Los potenciadores de *splicing* generalmente se encuentran dentro del exón regulado y suelen ser reconocidos por al menos un miembro de la familia de proteínas SR que recluta la maquinaria *splicing* al intrón adyacente (Blencowe, 2000; Black, 2003).

Entre las principales proteínas SR se encuentran SF2/ASF (SRSF1), SC35 (SRSF2), SRp40 (SRSF5), SRp55 (SRSF6) o 9G8 (SRSF7), entre otras (Busch and Hertel, 2012; Jeong, 2017). Así, por ejemplo, para facilitar el reconocimiento de los sitios donadores atípicos se requiere de la participación de *enhancers* de *splicing* y de las correspondientes proteínas SR (Kralovicova et al., 2011; Fraile-Bethencourt et al., 2017). El exón 17 del gen BC/OC *BRCA2* posee un donador atípico GC que para su reconocimiento necesita de la participación de los factores de *splicing* SC35, Tra2 $\beta$ , SF2/ASF y SRp20 (Fraile-Bethencourt et al., 2017). Por tanto, es factible que el donador GC del exón 12 de *PALB2* esté regulado de una forma similar, y será objeto de estudio en parte de esta tesis doctoral por sus posibles implicaciones en la predisposición hereditaria a cáncer de mama/ovario. Además, las proteínas SR pueden actuar como oncoproteínas cuando están sobre-expresadas. Así, por ejemplo, la proteína SF2/ASF está sobre-expresada en varios tipos de tumores como el de pulmón, colon o mama (Karni et al., 2007; Anczuków et al., 2012).

Por su parte, las hnRNPs representan una gran familia de proteínas de unión a RNA y contribuyen en múltiples aspectos del metabolismo de los ácidos nucleicos ayudando a la maduración de los pre-mRNAs recién sintetizados, estabilizando el transcrito durante su transporte y controlando su traducción. Muchas hnRNPs comparten características generales, pero difieren en la composición de su dominio y en sus propiedades funcionales (Figura 11). Estructuralmente, las hnRNPs poseen 4 dominios diferentes de unión a RNA: RRM (el más abundante), quasi-RRM, un dominio rico en Gly formando una caja RGG y un dominio KH. Se sabe que muchas hnRNPs regulan el *splicing* alternativo, reprimiendo la inclusión de exones o causando la retención de intrones. El papel de las hnRNP en la regulación de la expresión génica ha generado un gran interés en la investigación de enfermedades. Los niveles de expresión de hnRNPs se altera en muchos tipos de cáncer, sugiriendo un papel relevante en carcinogénesis, pudiendo actuar como oncogenes y genes supresores de tumores (Geuens et al., 2016). Además del cáncer, muchas hnRNPs también se han relacionado con diversas enfermedades neurodegenerativas, como la atrofia muscular espinal, la esclerosis lateral amiotrófica, la enfermedad de Alzheimer y la demencia fronto-temporal. Como las neuronas son células que no se dividen, por lo que necesitan una regulación

estricta de la homeostasis del mRNA, siendo muy vulnerables a pérdidas de función en proteínas de unión al RNA, incluidas las hnRNPs (Geuens et al., 2016).

Debido a la gran importancia que tienen los hnRNPs en la regulación de la expresión génica, no es sorprendente que muchas de ellas sean esenciales en el desarrollo de algunos tumores. La hnRNPA1 está asociada con diferentes tipos de cáncer y metástasis. Sus niveles de expresión aumentan drásticamente en muestras de cáncer de pulmón y está asociada con la proliferación tumoral (Liu et al., 2016). Además, la eliminación de hnRNPA2/B1 en células tumorales MDA-MB-231 disminuye la invasión celular y la proliferación (Loh et al., 2015).

En 2013, Anantha R. y colaboradores determinaron que hnRNPC forma parte del complejo de nucleoproteínas que contiene las proteínas supresoras de cáncer de mama PALB2, BRCA2 y BRCA1. hnRNPC modifica su localización nuclear después de la radiación ionizante (IR) y se localiza parcialmente en los sitios de daño en el DNA. El agotamiento de hnRNPC altera sustancialmente el equilibrio normal de los mecanismos de reparación después de la inducción de roturas de DNA de cadena doble, reduciendo el uso de la recombinación homóloga y alterando la progresión de la fase S después de la IR. Además, la pérdida de hnRNPC reduce fuertemente la abundancia de proteínas clave para la recombinación homóloga, como: BRCA1, BRCA2, RAD51 y BRIP1, lo que puede atribuirse, al menos en parte, a una baja regulación de sus mRNA debido a un *splicing* aberrante. La falta de otras proteínas hnRNP produjo poco o ningún efecto sobre la recombinación homóloga. Por lo tanto, dada la inexistencia de otros hnRNP en nuestro complejo BRCA1-PALB2-BRCA2, hnRNPC claramente tiene un papel principal en regulación de la expresión de varios genes clave en la ruta de reparación del DNA de doble cadena (Anantha et al., 2013).

### **3.1.6. *Splicing* y enfermedad.**

Un análisis reciente del proyecto Encyclopedia of DNA Elements (ENCODE) (Dunham et al., 2012) (GRCh38, Ensembl79) indica que la mayor parte del genoma humano se transcribe y consta de aproximadamente 60.000 genes (~20.000 genes que codifican proteínas, ~ 16.000 RNA

largos no codificantes (lncRNAs), ~ 10.000 RNA no codificantes pequeños y 14.000 pseudogenes). Aunque es probable que este inventario de genes pueda cambiar, el número de genes que codifican proteínas es sorprendentemente bajo dada la complejidad proteómica de muchos tejidos (Scotti and Swanson, 2016). Los genes humanos expresan pre-mRNA notablemente complejos, que contienen una media de ocho exones, con intrones que constituyen el 90% de la unidad de transcripción. Los exones de este pre-mRNA están separados por intrones de hasta cientos de miles de nucleótidos de longitud y deben unirse con precisión para mantener el marco de lectura.

A veces el proceso de *splicing* se ve truncado por alteraciones (mutaciones) en los genes, que generan transcritos aberrantes. La correlación entre *splicing* aberrante y enfermedad, tiene gran relevancia en Biomedicina, como consecuencia de la cantidad de variantes deletéreas que afectan al *splicing* de genes de susceptibilidad a enfermedades hereditarias (Scotti and Swanson, 2016). Estas mutaciones pueden alterar los sitios 3' y 5' de *splicing*, el punto de ramificación, el tracto de polipirimidinas y secuencias en *cis* que actúan como activadores o represores. Dada la complejidad de los elementos de la secuencia de los pre-mRNA y los factores de *splicing* que actúan en *trans*, no es sorprendente que este paso de procesamiento del RNA sea particularmente susceptible a variantes germinales y somáticas que están implicadas en la enfermedad (Wang and Cooper, 2007).

La eficiencia del *splicing* puede variar entre individuos debido a variantes en motivos de reconocimiento o reguladores que actúan en *cis* o en los genes que codifican los factores de *splicing* que actúan en *trans*, asimismo, cabe destacar que algunos RNA no codificantes (ncRNA), incluidos los lncRNA y los RNA circulares (circRNA), están implicados en la regulación del *splicing* (Ashwal-Fluss et al., 2014; Barry et al., 2014).

Las alteraciones en el *splicing* pueden causar la enfermedad directamente, modificar la gravedad del fenotipo de la enfermedad o estar relacionadas con la susceptibilidad a ésta. En cada caso, los mecanismos que causan *splicing* aberrante, implican alteraciones de los elementos que actúan en *cis*, dentro del gen afectado o de los factores que actúan en *trans*, requeridos para que se produzca correctamente el proceso *splicing*. Los efectos en *cis* tienen un impacto directo en la

expresión de un solo gen, mientras que los efectos en trans tienen el potencial de afectar la expresión de múltiples genes.

### 3.1.6.1. Mutaciones de *splicing* en *cis*.

En 1992 se creía que el 15% de las mutaciones causantes de enfermedad afectaban al *splicing*, aunque solo se tuvieron en cuenta aquellas mutaciones que alteraban los sitios naturales de *splicing* (Krawczak et al., 1992). Ahora está claro que una gran fracción, pero aún desconocida, de mutaciones afecta al *splicing*. Incluyendo otros componentes del código de *splicing*, como *enhancers* y silenciadores, la proporción de variantes con impacto en el *splicing* (o variantes *esplíceogénicas*) supera incluso el 50% (Cartegni et al., 2002; López-Bigas et al., 2005; Pagenstecher et al., 2006), convirtiéndolas en la principal causa de enfermedades hereditarias. La patogenicidad de las variantes exónicas, suele estar determinada por el efecto predicho sobre el marco de lectura y la funcionalidad de la proteína. Las variantes que habitualmente se catalogan de acuerdo a su impacto en la secuencia proteica como *nonsense*, *missense* o sinónimas, en realidad podrían estar afectando el patrón de *splicing* de los genes. No reconocer este hecho, podría confundir la comprensión de la base molecular de la enfermedad y conducir a una evaluación errónea de las variantes patogénicas.

En 2005, Pagani y colaboradores, analizaron de forma minuciosa los exones 9 y 12 del gen *CFTR*, responsable de la fibrosis quística, y vieron que aproximadamente una cuarta parte de las mutaciones sinónimas daban como resultado un *splicing* aberrante (Pagani et al., 2005). En 2020, Cerasuolo y colaboradores, analizaron la secuencia de *STK11* y de otros 8 genes en dos hermanas jóvenes sospechosas de padecer el síndrome de Peutz-Jeghers (PJS), que mostraban hiperpigmentación mucocutánea típica del PJS. Las dos hermanas portaban una nueva variante germinal en uno de sus alelos producida por un cambio de G a A (c.597G>A), que causaba la eliminación del sitio de *splicing* 5' de exón 4 de *STK11*. Esta variante ocasionaba la formación de una isoforma de mRNA de *STK11* aberrante sin exón 4, que genera un codón de parada prematuro y por consiguiente una proteína truncada. Sin embargo, esta variante también fue detectada en el

padre de las dos hermanas, que, no reflejaba ningún síntoma de PJS. El padre, además presentaba una variante heterocigótica producida por un cambio de T a C (c.465-51T>C) en el intrón 4 de *STK11*, que no portaba ninguna de sus hijas. El análisis con el programa ESEfinder mostró que la variante c.465-51T>C aumentaba la afinidad de unión del factor de *splicing* *SRSF1* por el motivo ESE ubicado en el intrón 4. A través de este mecanismo, este polimorfismo, puede ser responsable de un incremento de la expresión del mRNA de *STK11* en el padre en comparación con sus hijas, atenuando el efecto secundario de la variante patogénica c.597G>A y contribuyendo a la variabilidad fenotípica del PJS en la familia (Cerasuolo et al., 2020). Este ejemplo ilustra la enorme influencia de la variabilidad genética sobre el *splicing*. Aun así, la mayoría de las mutaciones *splicing* que actúan en *cis* permanecen sin ser detectadas (Wang and Cooper, 2007).

### **3.1.6.2. Mutaciones de *splicing* que actúan en *trans*.**

La complejidad estructural y funcional del *esplíceosoma* rivaliza con la del ribosoma. El *esplíceosoma* está compuesto por 5 snRNP y un número variable de proteínas asociadas, incluidas quinasas, fosfatasas y helicasas, muchas de las cuales son necesarias para la correcta función de la maquinaria de *splicing*, así como proteínas asociadas; como factores para exportar el mRNA fuera del núcleo y factores de transcripción (Jurica and Moore, 2003).

Dos enfermedades, la atrofia muscular espinal (AME) y la retinitis pigmentaria (Briese et al., 2005; Mordes et al., 2006), son causadas por mutaciones en genes implicados en el ensamblaje y en la función de snRNP, respectivamente. A pesar de la función de los genes responsables, sorprendentemente, estas enfermedades afectan a distintos subconjuntos de neuronas. La AME es un trastorno autosómico recesivo que afecta principalmente a las  $\alpha$ -motoneuronas de la médula espinal y está causado por la pérdida de función del gen *SMN1*, requerido para el ensamblaje de los snRNP en el citoplasma antes de la maduración final y su importación al núcleo. La pérdida en la producción de snRNP se ha relacionado directamente con la enfermedad (Winkler et al., 2005).

Por otro lado, la retinitis pigmentaria es una de las formas más comunes de ceguera y afecta a 1 de cada 4000 personas en todo el mundo. La enfermedad se produce como resultado de la

degeneración de la retina, debido principalmente a la pérdida progresiva de células fotorreceptoras. La mayoría de los casos son esporádicos; sin embargo, variantes en más de 30 genes causan formas familiares de la enfermedad, pudiendo tener una herencia autosómica recesiva, autosómica dominante o ligada al cromosoma X. Sorprendentemente, tres genes de la forma dominante de retinosis pigmentaria (homólogos del gen del factor de procesamiento de mRNA: *PRPF31*, *PRPF8* y *PPRP3*) codifican proteínas necesarias para el ensamblaje y funcionamiento adecuado del tri-snRNP, U4-U5-U6, un componente esencial del *spliceosoma* (McKie et al., 2001; Chakarova et al., 2002). Probablemente, la retinitis pigmentaria se produzca debido a alteraciones en la función del *spliceosoma*. Es previsible que la especificidad del tipo celular resulte de la sensibilidad de uno o más pre-mRNA específicos de fotorreceptores a la pérdida de la función de U4-U5-U6 tri-snRNP.

### **3.1.6.3. Efectos *trans*-dominantes sobre la regulación del *splicing*: ganancia de función del RNA.**

Las expansiones de tres nucleótidos dentro de las regiones codificantes causan enfermedades como la enfermedad de Huntington, ataxias espinocerebelosas y distrofia muscular oculofaríngea debido a la pérdida y/o ganancia de función de las proteínas codificadas, que contienen trectos anormalmente expandidos de poliglutamina o polialanina (Lalioi et al., 1997; Gatchel and Zoghbi, 2005). La enfermedad mejor caracterizada de todas ellas es la distrofia miotónica, un trastorno hereditario dominante y conocido por ser la segunda causa más común de distrofia muscular. La distrofia miotónica es causada por una expansión de tres nucleótidos CTG en la región 3' no traducida (3' UTR) de *DMPK*, o por una expansión de CCTG en el intrón 1 de *ZNF9* (también conocido como *CNBP*).

La característica molecular clave de la distrofia miotónica, es la mala regulación del *splicing* alternativo fisiológico en un subconjunto de genes, de modo que los patrones de *splicing* embrionario y neonatal se conservan en los tejidos adultos de los pacientes con esta patología. Se puede apreciar que la miotonía y la resistencia a la insulina son el resultado de la expresión

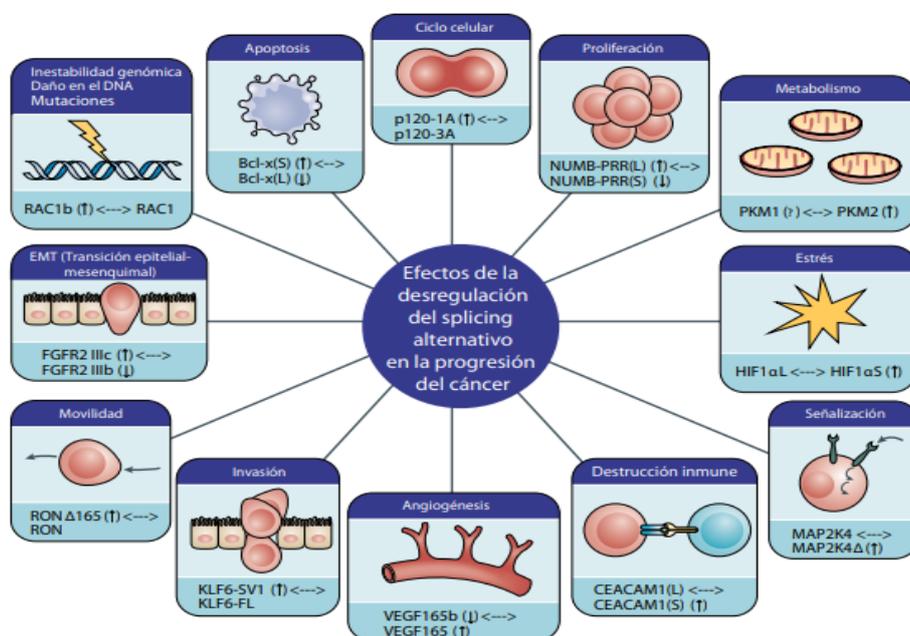
inapropiada de diferentes patrones de *splicing* embrionario en tejidos adultos (Osborne and Thornton, 2006; Ranum and Cooper, 2006; Cho and Tapscott, 2007).

#### **3.1.6.4. *Splicing* y Cáncer.**

The Cancer Genome Atlas (TCGA) fue un proyecto de genómica del cáncer que caracterizó molecularmente más de 20.000 cánceres primarios y emparejó muestras que abarcaban 33 tipos de cáncer. En 2018 Seiler y colaboradores, utilizando los datos de este proyecto, determinaron que las mutaciones somáticas que supuestamente impulsaban la aparición del cáncer, se originaban en 119 genes que codificaban proteínas reguladoras y factores de *splicing* (Seiler et al., 2018). Además de las mutaciones, los cambios en la expresión de los factores de *splicing* a menudo están asociados con procesos de oncogénesis o la pérdida de la supresión tumoral. El 84% de las proteínas de unión al RNA y más del 70% de los factores de *splicing*, presentan fallos en la regulación de sus niveles de expresión (Dvinge et al., 2016; Sebestyén et al., 2016; Sveen et al., 2016). Fallos en la maquinaria de *splicing* de las células cancerosas puede contribuir a prácticamente todos los aspectos de la progresión tumoral (Figura 12) (Kozlovski et al., 2017; Di et al., 2019; Bonnal et al., 2020). Por ejemplo, la retención de varios de los llamados “intrones detenidos”, conduce a la acumulación de pre-mRNA en el núcleo hasta que un estímulo desencadena el *splicing* y, en consecuencia, se produce la expresión rápida de productos génicos, asociados con proliferación, senescencia y apoptosis que contribuyen a la neurogénesis. En el glioblastoma, se interrumpe la retención de intrones programada y la inhibición de PRMT5, que es una arginina N-metiltransferasa importante para la biogénesis de snRNP, y el *splicing* de los intrones detenidos, inhibe de forma potente la progresión del glioblastoma en modelos murinos (Bonnal et al., 2020).

El factor de *splicing* *SRSF1* es un protooncogén poderoso cuya regulación positiva da como resultado la modulación simultánea del *splicing* alternativo en genes objetivo clave. Estos cambios postranscripcionales en la expresión génica resultan, por un lado, en la inactivación de supuestos supresores de tumores como *BIN1* y, por otro lado, en la generación de isoformas oncogénicas. Además, los niveles de expresión de la proteína *SRSF2* aumentan con frecuencia en varios tumores

sólidos, incluidos los cánceres de mama y pulmón. SRSF1 es un objetivo directo de *MYC*, otro protooncogén, y su expresión elevada se correlaciona con aumento del grado tumoral, disminución de la supervivencia y resistencia a la quimioterapia (Karni et al., 2007). Generalmente, estas alteraciones en el *splicing*, se producen en genes que controlan procesos esenciales, como el crecimiento celular, la apoptosis o la motilidad (Urbanski et al., 2018). En particular, la titulación proteica de la actividad de SRSF1 utilizando oligonucleótidos señuelo que contienen sitios de unión a SRSF1 da como resultado la inhibición del crecimiento de células cancerosas y la apoptosis *in vitro* y en modelos murinos (Denichenko et al., 2019).



**Figura 12. Efecto de la desregulación del *splicing* alternativo sobre la progresión del cáncer.** El diagrama ilustra diferentes distintivos del cáncer junto con ejemplos de eventos de *splicing* alternativo que contribuyen a la regulación de los diferentes procesos de tumorigénesis. Las flechas hacia arriba y hacia abajo indican las isoformas que contribuyen más y menos, respectivamente, a cada proceso (Bonnal et al., 2020)

Los reordenamientos cromosómicos de los factores de *splicing* *EWS* y *RBN15* son característicos de los sarcomas de Ewing y de la leucemia megacariocítica aguda pediátrica, respectivamente (Janknecht, 2005; Seiler et al., 2018). Debido a la función de unión al RNA de estas proteínas, estas lesiones genéticas también pueden alterar el proceso de *splicing* del RNA

inmaduro en genes relevantes para el cáncer, incluidos los que codifican las helicasas *RNA-DNA*, *DHX9* y *ARID1A*, con efectos oncogénicos. Además, se ha comprobado como alteraciones de los niveles de expresión de snRNAs en muestras de cáncer de pacientes con carcinoma ductal de mama, inducía errores en el *splicing*, lo que sugiere que los cambios en los niveles de snRNA también pueden contribuir a la progresión del cáncer.

Otro ejemplo más de alteración de los patrones de *splicing* en el cáncer se puede observar en el gen *PKM* de la piruvato quinasa, donde se produce una reversión de los patrones de *splicing* de adulto a embrionario que afecta a dos exones mutuamente excluyentes de este gen, lo cual confiere una ventaja de crecimiento a las células cancerosas al permitir la generación rápida de energía a través de la glucólisis aeróbica (efecto Warburg), desviando la glucosa hacia otros procesos de biosíntesis (Christofk et al., 2008).

Todos los ejemplos citados muestran la gran relevancia que tiene la alteración del *splicing* en la carcinogénesis y en casi todas las etapas distintivas de la progresión tumoral, incluida la invasión celular, la angiogénesis y el metabolismo celular (David and Manley, 2010; Lodomery, 2013; Oltean and Bates, 2014; Biamonti et al., 2018; Urbanski et al., 2018). Por tanto, errores en el proceso de *splicing* y la generación de isoformas específicas del cáncer, relacionadas con alteraciones de la actividad de factores y reguladores de *splicing*, contribuyen a cambios relevantes en el transcriptoma, que fomenta el desarrollo tumoral. Todos estos procesos tienen una gran relevancia para la Oncología Clínica, las respuestas a la terapia contra el cáncer, la aparición de la resistencia a los medicamentos, así como para la generación de nuevos biomarcadores y dianas terapéuticas específicas para el desarrollo de nuevos fármacos antitumorales (Gilbert, 1978; Bonnal et al., 2020). Desde el punto de vista terapéutico, es especialmente llamativa la vulnerabilidad de las células tumorales a la inhibición farmacológica del *splicing*.



# **OBJETIVOS**

El objetivo principal de esta tesis es determinar el papel de la desregulación del *splicing* en el gen de susceptibilidad *PALB2* en cáncer de mama hereditario, y contribuir a la dilucidación de su espectro de predisposición genética.

Objetivos específicos:

1. Construcción y validación de minigenes estables con los exones 1 a 12 de *PALB2*.
2. Análisis bioinformático y selección de variantes BRIDGES candidatas de alterar el *splicing* en los exones 1-12 de *PALB2*.
3. Estudio funcional y clasificación de variantes BRIDGES.
4. Análisis bioinformático de variantes ClinVar candidatas de alterar el *splicing* en los exones 1-12 de *PALB2*.
5. Estudio funcional y clasificación de variantes ClinVar situadas en los exones 1-3.
6. Mapeo funcional de elementos reguladores del *splicing* de los exones 11 y 12 de *PALB2*.

# **MATERIALES Y MÉTODOS**

## **1. Bioinformática.**

### **1.1. Bases de datos y programas:**

La información necesaria para desarrollar el estudio bioinformático se ha obtenido de diferentes bases de datos de carácter internacional. En primer lugar, ha sido necesario consultar la secuencia consenso para el gen *PALB2* y, por otra parte, listar las variantes genéticas detectadas en controles y casos del proyecto BRIDGES y las registradas en la base de datos ClinVar.

#### **1.1.1. Bases de datos de secuencias. Ensembl.**

La base de datos Ensembl ([www.ensembl.org](http://www.ensembl.org)), ha sido utilizada para consultar la secuencia consenso del gen *PALB2* en el genoma humano. El proyecto de base de datos del genoma Ensembl es un proyecto científico del Instituto Europeo de Bioinformática publicado en 1999 en respuesta a la inmediata finalización del Proyecto Genoma Humano (Hubbard et al., 2002).

Esta base de datos de dominio público y acceso gratuito, permite acceder de forma centralizada a todo el genoma humano y a una gran colección de genes de distintas especies. Además, aporta descripciones de las características identificadas en el DNA, denominadas anotaciones y expone otras características importantes como variantes, repeticiones y homologías.

La Tabla 2 muestra el número de acceso a la secuencia del gen *PALB2* y sigue el siguiente formato: ENS+letra+11 dígitos, utilizando la letra P, para proteína, T para transcrito, G para gen, etc.

**Tabla 2. Numero de acceso de la secuencia del gen *PALB2* de Homo Sapiens.**

Especie	Nº de acceso <i>PALB2</i>
Homo Sapiens	<b>ENSG00000083093</b>

### **1.1.2. Lasergene Molecular Biology. Seqbuilder Pro.**

Lasergene Molecular Biology es un software de análisis de secuencias propiedad de la empresa DNASTAR (DNASTAR Inc, Madison, USA).

**Seqbuilder Pro** (<https://www.dnastar.com/software/lasergene/seqbuilder-pro>), es una aplicación de Lasergene para la edición de secuencias, el diseño de plásmidos, el clonaje virtual automatizado y el diseño de oligonucleótidos.

### **1.1.3. Variantes y anotación de transcritos.**

Las variantes, los transcritos y los productos proteicos se nombrarán de acuerdo con las pautas propuestas por The Human Variation Society (HGVS) (<https://vRNAomen.hgvs.org/>), usando la secuencia de referencia Ensembl del transcrito ENSG00000083093 (Genbank NM\_024675.4).

### **1.1.4. ClinVar.**

ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) es un archivo público y de libre acceso de variantes genéticas humanas. Además, incluye interpretaciones de las variantes y su relación con enfermedades y otras afecciones. Se mantiene en el National Center for Biotechnology Information (NCBI), dentro de la National Library of Medicine (NLM) de los National Institutes of Health (NIH). Más de 1300 grupos de investigación y organizaciones han presentado interpretaciones de variantes en ClinVar, incluidos laboratorios de pruebas clínicas, laboratorios de investigación, médicos, registros de pacientes, paneles de expertos, etc (Landrum et al., 2020). Las clasificaciones

incluidas en ClinVar presentan varios niveles de confianza representados por estrellas, de la manera detallada en la Tabla 3.

**Tabla 3. Clasificación por estrellas en la base de datos ClinVar** (Fuente: Fundación Instituto Roche; <https://www.institutoroche.es/recursos/glosario/clinvar>, acceso 10 de junio de 2021).

<b>Número de Estrellas</b>	<b>Descripción y estado de la revisión</b>
<b>0</b>	No existe interpretación sobre la variante o no está apoyada por ninguna evidencia
<b>1</b>	La interpretación de la variante proviene de una sola fuente o por varias, presentando discrepancias.
<b>2</b>	Dos o más fuentes han proporcionado la misma interpretación para una misma variante
<b>3</b>	La interpretación ha sido revisada y validada por expertos.
<b>4</b>	Existen guías de práctica clínica asociadas a la variante

El significado clínico se expresa según los términos estandarizados de ClinVar y del Genetic Testing Registry (GTR), excepto cuando ClinVar procesa datos de OMIM, que establece valores de acuerdo con sus propias pautas. Existen cinco valoraciones diferentes respecto al significado clínico de una variante: Benigna, Posiblemente benigna, Significado clínico desconocido (VUS), Posiblemente patogénica y Patogénica. Estas categorías siguen la nomenclatura sugerida la ACMG y AMP para catalogar variantes en enfermedades mendelianas.

#### **1.1.5. Diseño de oligonucleótidos.**

Los oligonucleótidos empleados para PCR han sido diseñados con la herramienta informática Primer3 input (v3.0.0) (<http://primer3plus.com/web3.0.0/primer3webinput.htm>) o con el programa de diseño de oligonucleótidos Oligo v7.0 (Molecular Biology Insights, Colorado Springs, Colorado, USA). Algunos de estos oligonucleótidos fueron marcados con el fluorocromo

6-FAM (6-Carboxyfluorescein), con el fin de poder utilizarlos en el análisis de fragmentos fluorescente mediante electroforesis capilar.

## **1.2. Selección de variantes candidatas de *splicing*.**

Los datos de las variantes candidatas de alterar el *splicing* en el gen de susceptibilidad a cáncer de mama *PALB2*, han sido seleccionadas del proyecto europeo BRIDGES, en el cual se han secuenciado más de 113.000 mujeres de 25 países de la Unión Europea. Además, hemos seleccionado y analizado funcionalmente variantes de la base de datos ClinVar, que incluye 3820 variantes de este gen (acceso agosto 2021).

### **1.2.1. Human *Splicing* Finder.**

Existen diferentes herramientas informáticas que permiten evaluar la probabilidad que tiene un determinado cambio en el DNA de afectar a secuencias básicas de los sitios canónicos de *splicing* y otros elementos reguladores implicados en el proceso de *splicing*. En nuestro caso, la selección de variantes se ha llevado a cabo con el programa Human Splicing Finder versión 3.0 (HSF; <http://www.umd.be/HSF3/>; en la actualidad este sitio web ha sido cerrado y HSF está incluido dentro de Genomnis, <https://www.genomnis.com/>). HSF incluye varios algoritmos que detectan eliminación o creación de sitios de *splicing* (MaxEntScan -MES- y uno propio de HSF) y activadores o silenciadores. Los criterios de selección de variantes de DNA fueron los siguientes:

- 1) Eliminación de los sitios naturales aceptor y donador de *splicing*: variantes situadas en las posiciones intrónicas -3, -2 y -1 (aceptor), +1, +2, +3, +5, +6 (donador), y en los 2 primeros nucleótidos (aceptor) y los 3 últimos (donador) de cada exón
- 2) Generación de nuevos sitios fuertes alternativos de *splicing*.
- 3) Alteraciones del tracto de pirimidinas: cambios Pirimidina→Purina o disminución de su longitud (delección).

Un total de 82 variantes BRIDGES y 151 de la base de datos ClinVar de las fronteras intrón-exón fueron preseleccionadas para la realización de los análisis *in silico*:

**Exón 1:**

**BRIDGES:** c.47A>G y c.48G>A.

**ClinVar:** c.46A>G.

**Intrón 1:**

**BRIDGES:** c.48+1G>A, c.48+2T>C, c.48+3G>A, c.48+4C>T, c.48+7G>C y c.49-2A>T.

**ClinVar:** c.48+1G>C, c.48+1G>T, c.48+2T>G, c.48+1del, c.48+5C>T, c.48+6G>C, c.49-1del, c.49-2del y c.49-1G>A.

**Exón 2:**

**BRIDGES:** c.50T>G y c.106C>T.

**ClinVar:** c.50dup, c.50T>G, c.106C>T y c.108G>A.

**Intrón 2:**

**BRIDGES:** c.108+1G>A, c.108+2T>C, c.108+7G>A, c.108+8A>G, c.108+9del, c.108+9A>G, c.109-7C>T, c.109-7C>A, c.109-5T>C, c.109-6\_109-4del y c.109-2A>G.

**ClinVar:** c.108+1\_108+2insC, c.108+4A>G, c.108+5G>A, c.108+6T>C y c.109-2A>C.

**Exón 3:**

**BRIDGES:** c.109C>A y c.109C>T.

**ClinVar:** c.109C>A, c.109C>T, c.110G>A, c.11T>C, c.210A>C y c.210A>G.

**Intrón 3:**

**BRIDGES:** c.211+1G>A y c.211+5del.

**ClinVar:** c.211+1G>T, c.211+2T>C, c.211+4A>G, c.211+6T>A, c.212-1G>A, c.212-1G>T, c.212-2A>C, c.212-2A>T, c.212-2A>G y c.212-3C>T.

**Exón 4:**

**BRIDGES:** c.1682A>G.

**ClinVar:** c.212A>G, c.213A>G, c.213A>T, c.213A>C y c.1684G>A.

**Intrón 4:**

**BRIDGES:** c.1684+4A>G, c.1684+7C>T, c.1684+8A>G, c.1684+9A>C, c.1685-2A>C y c.1685-2A>C.

**ClinVar:** c.1684+1G>T, c.1684+1G>A, c.1684+3A>C, c.1684+6del, c.1684+6\_1584+9del, c.1685-3T>C, c.1685-3T>G, c.1685-1G>A y c.1685-1G>C.

**Exón 5:**

**BRIDGES:** c.2513A>C.

**ClinVar:** c.1686G>A, c.2512C>T, c.2512C>G, c.2513A>G y c.2514G>A.

**Intrón 5:**

**BRIDGES:** c.1514+5A>G, c.2514+6A>G, c.2514+8C>T, c.2515-9T>C, c.1515-9T>G y c.2515-2A>G.

**ClinVar:** c.2514+1G>A, c.2514+1G>C, c.2514+3A>G, c.2514+3dup, c.2514+4C>T, c.2514+5A>G, c.2515-3C>T, c.2515-3C>G, c.2515-1G>T y c.2515-1G>A.

**Exón 6:**

**BRIDGES:** no se detectaron variantes candidatas.

**ClinVar:** c.2517T>G, c.2517\_2518del y c.2586G>C.

**Intrón 6:**

**BRIDGES:** c.2586+3C>G, c.2586+4A>T, c.2586+8G>C y c.2587-2A>G.

**ClinVar:** c.2586+1G>T, c.2586+3C>T, c.2586+3C>G, c.2587-2A>G y c.2587-1G>C.

**Exón 7:**

**BRIDGES:** no se detectaron variantes candidatas.

**ClinVar:** c.2587A>G.

**Intrón 7:**

**BRIDGES:** c.2748+1G>T, c.2748+1G>A, c.2748+2dup, c.2748+4A>T, c.2748+8G>A, c.2748+9del y c.2749-1G>T.

**ClinVar:** c.2748+1G>T, c.2748+1G>C, c.2748+1G>A, c.2748+1\_2748+2del, c.2748+2T>C, c.2748+3A>C, c.2747\_2748+4del, c.2749-2A>C, c.2749-1G>C y c.2749-1G>A.

**Exón 8:**

**BRIDGES:** c.2750T>C.

**ClinVar:** c.2749G>A, c.2832del y c.2832C>G.

**Intrón 8:**

**BRIDGES:** c.2834G>C, c.2834+1G>A, c.2834+2T>C, c.2834+3A>G, c.2834+5G>A, c.2834+6T>C, c.2834+8A>G, c.2835-9T>C, c.2835-7G>T y c.2835-3T>C.

**ClinVar:** c.2834+1G>C, c.2834+1G>T, c.2834+2T>G, c.2834+2T>C, c.2834+4T>C, c.2834+5G>A, c.2835-3T>C, c.2835-2A>C, c.2835-1G>A y c.2835-1G>C.

**Exón 9:**

**BRIDGES:** c.2836G>A.

**ClinVar:** c.2835G>A, c.2836G>T, c.2836G>A, c.2837C>G, c.2837del, c.2996G>A, c.2996G>A y c.2996G>C.

**Intrón 9:**

**BRIDGES:** c.2996+4A>G, c.2997-2del y c.2997-1G>A.

**ClinVar:** c.2996+1G>T, c.2996+2T>C, c.2996+3A>C, c.2996+3A>G, c.1996+5G>T, c.1997-3C>G y c.2997-2A>T.

**Exón 10:**

**BRIDGES:** c.3113G>A.

**ClinVar:** c.2997A>T, c.2998G>A, c.2999G>A, c.2999G>T, c.2999del, c.2999\_3001del y c.311T>C.

**Intrón 10:****BRIDGES:** c.3113+3A>G, c.3113+5G>C, c.3113+8T>C y c.3114-5C>T.**ClinVar:** c.3113+1G>A, c.3113+5G>C, c.3114-3A>G y c.3114-1G>A.**Exón 11:****BRIDGES:** no se detectaron variantes candidatas.**ClinVar:** c.3114G>A, c.3114del, c.3116del, c.3116A>C y c.3201G>A.**Intrón 11:****BRIDGES:** c.3201+1G>A, c.3201+3\_3201+4insTG, c.3202-9C>T, c.3202-8G>T y c.3202-8G>A.**ClinVar:** c.3201+1G>T, c.3201+1G>C, c.3201+2dup, c.3201+3A>C, c.3201+4del, c.3201+4A>G, c.3201+5G>T, c.3201+6T>A, c.3202-2A>G, c.3202-2A>C, c.3202-1G>A y c.3202-1G>C.**Exón 12:****BRIDGES:** c.3203G>C y c.3348C>T.**ClinVar:** c.3202G>T, c.3203G>A, c.3204G>A, c.3204G>T y c.3350G>A.**Intrón 12:****BRIDGES:** c.3350+3A>G, c.3350+4A>G y c.3350+5G>A.**ClinVar:** c.3350+1G>A, c.3350+2C>G, c.3350+4A>C y c.3350+6T>C.

### 1.2.2. Herramienta HEXplorer.

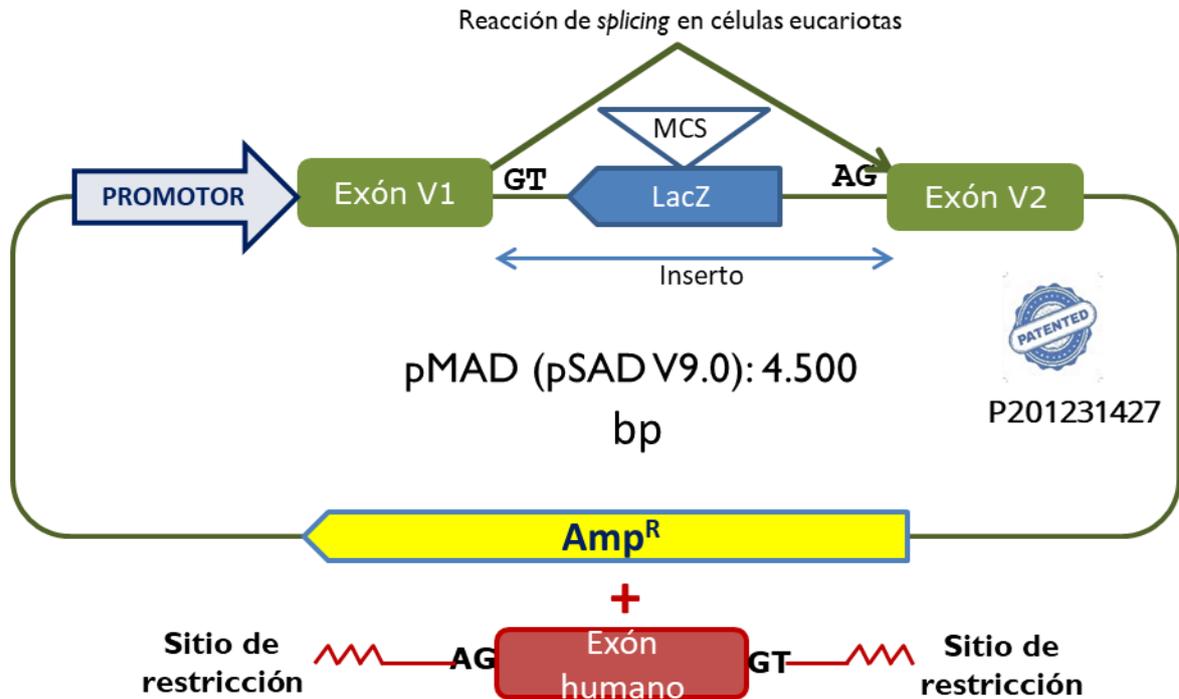
La selección efectiva de los sitios de *splicing* está controlada críticamente por elementos reguladores del *splicing* (SRE, *enhancers*/silenciadores) que pueden activar o reprimir el reconocimiento de ciertos exones. Aunque existen varios algoritmos computacionales que identifican una gran variedad de motivos SRE, su precisión es limitada. HEXplorer ([https://www2.hhu.de/rna/html/hexplorer\\_score.php](https://www2.hhu.de/rna/html/hexplorer_score.php)) realiza un cálculo del balance entre *enhancers* y silenciadores de *splicing*. De este modo, permite, gracias a las diferencias en su sistema de puntuación entre las secuencias mutantes y de referencia, estimar los efectos de una mutación exónica en el reconocimiento de un exón (Erkelenz et al., 2014). El punto de corte habitualmente utilizado para la selección de variantes es menor de -5 (Bueno-Martínez et al., 2021). En esta tesis, se aplicó un criterio aún más restrictivo (< -40) para la selección de variantes de los exones 11 y 12.

## **2. Ensayos funcionales de *splicing* mediante minigenes híbridos.**

El *splicing* es un proceso esencial en la expresión génica mediante el cual los intrones son eliminados y los exones son unidos secuencialmente. Los minigenes híbridos permiten recrear un gen humano implicado en una enfermedad y estudiar los efectos de las mutaciones sobre el *splicing* sin necesidad de muestras de paciente. Los minigenes son vectores plasmídicos de expresión en eucariotas que contienen parte de la secuencia de un gen. Los llamados plásmidos reporteros de *splicing* tienen un diseño sencillo, con un promotor fuerte (usualmente SV40), un exón inicial del vector o V1 con un sitio donador que lanza las reacciones del *splicing*, y un exón final del vector o V2 con un sitio aceptor que finaliza las reacciones (Figura 13).

### **2.1. Minigenes en pMAD.**

Todas las construcciones de este trabajo se realizaron en el vector pMAD (*Mis-Splicing And Disease*) o pSAD (*Splicing And Disease*) v9.0 (Patente P201231427-CSIC), después de haber utilizado el vector de *splicing* pSPL3 (Invitrogen, descatalogado) y comprobar que éste producía algunos errores en el proceso de *splicing* (Sanz et al., 2010). Algunos habían sido descritos por Burn y colaboradores, como la existencia de sitios críticos en la secuencia intrónica del vector y que eran reconocidos por la maquinaria de *splicing* (Burn et al., 1995). Además, se observaron otras desventajas en nuestro laboratorio, como la dificultad para clonar grandes fragmentos o la falta de una selección adecuada de las colonias recombinantes (Acedo et al., 2015).



**Figura 13. Esquema del vector pMAD.**

Por todo ello y con el objetivo de mejorar las reacciones de *splicing* producidas por el vector, se desarrolló el vector de *splicing* pSAD, con el que se consiguió solucionar los problemas generados por pSPL3 (Acedo et al., 2015; Fraile-Bethencourt et al., 2017) (Patente P201231427-CSIC). Posteriormente, se introdujeron nuevas mejoras en el vector, obteniéndose el plásmido pSAD v9.0 o pMAD. Las principales ventajas del plásmido pMAD son: i) Reducción del tamaño del vector (pSPL3, 6,03 kb vs. pMAD 4,5 kb); ii) Incorporación de un sitio de clonación múltiple con más de 20 sitios diferentes de restricción; iii) Inclusión de un segundo mecanismo de selección ( $\beta$ -galactosidasa, LacZ); iv) Fortalecimiento de los sitios donador de V1 y aceptor de V2; v) Eliminación de sitios crípticos del vector que puedan interferir en las reacciones de *splicing*. En conjunto, todas estas mejoras permiten clonar más fácilmente fragmentos de mayor tamaño y mejorar las reacciones de *splicing* que tengan lugar dentro del nuevo constructo/minigén.

## 2.2. Minigenes de *PALB2*.

Para este estudio, se construyeron tres minigenes a partir del vector pMAD, que contienen en conjunto los exones 1 a 12 de *PALB2*: mgPALB2\_ex1-3, mgPALB2\_ex4-6 y mgPALB2\_ex5-12. También se clonó el mgPALB2\_ex2-6, pero fue descartado porque no producía el transcrito canónico esperado.

El minigén mgPALB2\_ex1-3 tiene un tamaño de 5068 pares de bases (pb) con un inserto de 974 pb (Figuras 14 y 15), contiene los exones 1, 2 y 3 de *PALB2* y se generó en dos pasos independientes. El inserto con los exones 2 y 3 y las correspondientes secuencias intrónicas flanqueantes (680 pb) fue diseñado en nuestro laboratorio y se sintetizó por la empresa Genewiz (Genewiz, South Plainfield, NJ). Este fragmento se clonó en el vector pSAD v9.0. El exón 1 se amplificó con Phusion High Fidelity Polymerase (Fisher Scientific, Waltham, MA, USA) y los oligonucleótidos indicados en la Tabla 4. Este producto PCR fue clonado en el minigén mgPALB2\_ex2-3 por *Overlapping Extension PCR* (Bryksin and Matsumura, 2010), originando una fusión *in-frame* del exón V1 del vector pMAD y el exón 1 de *PALB2* (78 pb).

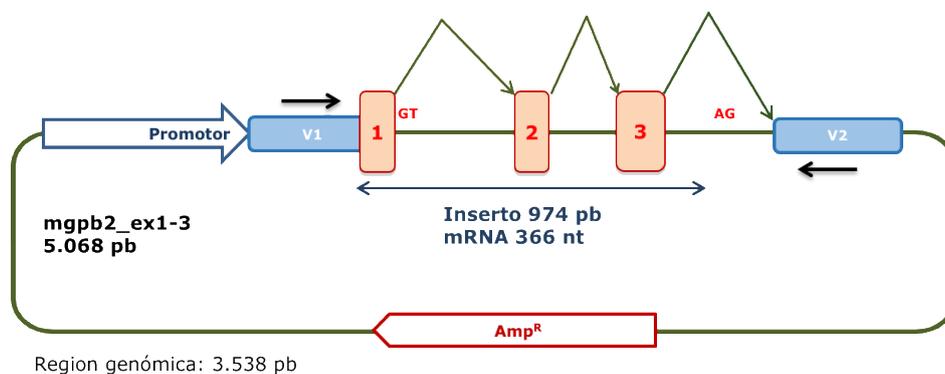


Figura 14. Estructura del minigén de los exones 1 a 3 de *PALB2*.

EXON V1 (pSAD)-  
GGCTGCTCTTTTGGTCTGTGCGCTGCCGATGGACGAGCCTCCCGGAAGCCCTCAGCTGTGAGGAGAAGGAAAAGgtgccgg Ex1  
gggtgcgggaagggcgacgcaggactctgaccccccttccagggttttaggcctggctttgtcctcggcagtcgagggc  
agcagtatcatctgaccacccccctctcttagggctgagggggcacagaaaatgaaaatgtggaatagcgtgggtctcactgg  
tggcgtttaagaacaaatttcattacagtagcagcgcaatgttccagatgataggaacaggttatgcaagatcctgaaatggt  
cagagcttgggtctttttgagaacaaaagtagattgttatggaccagtgtactcctcctgctcttggccaagggaccgcccaag  
cactgcatccccctcctctgactccacctttccacttgcccagattgttgggtgttttctctccagTTAAAGGAGAAATTAG Ex2  
CATTCTTGAAAAGGGAATACAGCAAGCACTAGCCCGCTTCAgtaagtgaatcgtattctcaattaaggtgttatagtacaa  
acaatttaaaaacagttcttgactctataaaactttaagaaaacgtatttctggggctgttttgcctcctctagCGTGCCCAA  
AGAGCTGAAAAGATTAAGCATTCTATTAAAGAAAACAGTAGAAGAACAGATTGTTTGTCTCAGCAGGATCTCTCACCGCAGCTAA Ex3  
AACACTCAGgtaaatctagaccattcactttatgctgtttattattcatttccaggatattttggctattgtctttttcc  
acagtgtgaagataatgactagcaatagacgcttaatttttaatttttaattttttttttttgagacagagtcactct  
ttgttaccaggctggagtgagtgccacaatcttggctCGGCCG  
EagI

**Exones 1-3.** Estructura del Inserto: ex1 (78 pb) - ivs1 (216 pb) // ivs1 (200 pb)- ex2 (60 pb) - ivs2 (117 pb) - ex3 (103 pb) - ivs3-1 (200 pb)

**Figura 15. Secuencia del inserto del minigén mgPALB2\_ex1-3.**

El minigén **mg\_PALB2\_ex4-6** tiene un tamaño de 8856 pb con un inserto de 4375 pb (Figuras 16 y 17) también fue clonado en 2 pasos independientes. Primero los exones 5 y 6 y sus correspondientes secuencias intrónicas adyacentes se amplificaron con la enzima polimerasa Phusion-High Fidelity (Fisher Scientific) y los oligonucleótidos indicados en la Tabla 4. Este inserto fue clonado en el vector pMAD mediante *Overlapping Extension PCR* (Bryksin and Matsumura, 2010), para generar el minigén mgPALB2\_ex5-6. A continuación, el inserto que contenía el exón 4 y parte de sus secuencias intrónicas contiguas se introdujo en el constructo anterior por clonaje clásico, mediante digestión con las enzimas de restricción *XbaI/BamHI* (New England Biolabs, Ipswich, MA, Estados Unidos) y su posterior ligación con Rapid DNA Ligation Kit (Thermo Scientific, Waltham, Massachusetts, Estados Unidos), para producir el minigén final mgPALB2\_ex4-6.

El minigén **mgPALB2\_ex5-12** tiene un tamaño de 9392 pb con un inserto de 4947 pb (Figuras 18 y 19) que se clonó en 3 etapas distintas. Primero, el exón 7 se amplificó con la enzima polimerasa Phusion-High Fidelity (Fisher Scientific) y los oligonucleótidos indicados en la Tabla 4. Posteriormente, este inserto se clonó en el constructo previo mgPALB2\_ex5-6 por *Overlapping Extension PCR* (Bryksin and Matsumura, 2010), para producir el minigén mgPALB2\_ex5-7. Por otro lado, los exones 8, 9, 10, 11 y 12 y sus secuencias intrónicas adyacentes, se sintetizaron por la



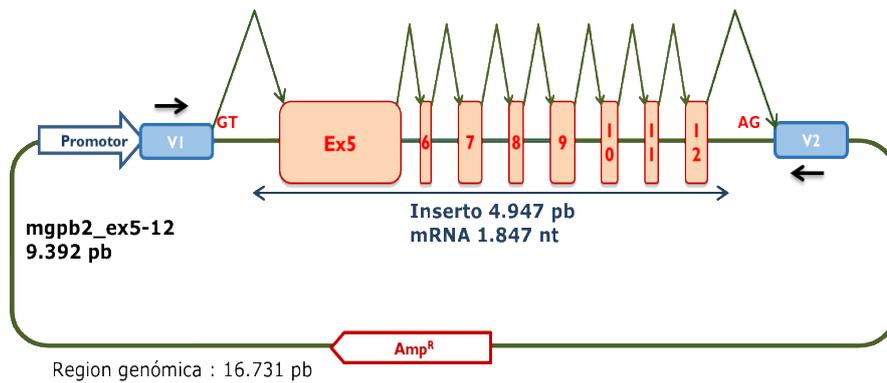


Figura 18. Estructura del minigén de los exones 5 a 12 de *PALB2*.

```

agtcatggatgggaaaagttaaacatttttagtataatgagtttgaagctcacctttgtgggtattacatttaagaa
tggttaaacatggtcttcttgataggactcattgttaaacattaaagttcattctgggaaaataaggttcattaaaatggtctt
ttaaacttaggagatcctattctctttgtcatcagtgaaacagattgtctgtttgtgggtttgttactattttgtgactta
ttttctcttttagGGAAGAAAAGCGTCATCAAAAAGAGGATTCCTCTTCTGGAGTAATAGTGCTTATTTA TCCTTGGATGA
TGATGCTTTCACGGCTCATTTCATAGGGATGGAATGCTGAGTTAAAGCAACTACTGTCTTCTCAGTATCACAGACTTTCAG
GTTCCTGATGAGACTTTGACCTCTTAAAGCTTGAAAAGTGAAGTGAAGTCTGCTCAGAAAACCACTGGAGCCCTTTGAGTCAAA
AATGTTTGAGAGAGACTCTTAAAGAGGGAAGCTGTATTTTTCCAGAGAACTGAGTCTTAAACGCA TGGATACAGAAATGGA
GGACTTAGAAGAGGACTTATTTGCTACCGAAAAATCACATCCAAAAGGCCAACTCGCAAAAGCCAGCATACAAGACGGG
CCTTTCTCATCATATTACTTTATACTCTTAAATACGGTTCGCCCTGATGATAATGACAGGCCTACCAAGACATGTGTC
ACCTGCTTTCCTCATCTAGTACTACTCCAGCTTTGGCCCTCAAGGCTCTATGAAAAAGCATCTACAGAAATGTGTCGAGC
AACTTGTGCACACCCCAACTTGTCTATTTGAAAGACTCAGTCTGTCTTGCCAGTGATCTAAACAATTCGACAGTTCAGGAC
CCACGAAAACCCATACCCACTGCAAGTGTGAGGAGGCAAGGCAACCTACTGCTGACTGTGACTCTGTCCGCCAGGAAAC
ACCTCCACCCTTGTGACTTCACTTTTAAAGAAAATCAGCTCTGTAGAAAACATGCCAGGAGCTGCATAAAACATTCCGCGA
ACAGTgacacattcctctgtgaaatctctctgaaagaaatgaaatgccttagtgaatgtaaacagcatgactgctctgagc
cattgggcttcacagcttaagaatgggttgaagctgtctttgatagattcattgtaacattaaagttcattctggggaaa
ttaaggttcattaaaatgttcttttaaatatgggaggtcctattctctgttatcagtgaaacagttgctattggagctt
gctgctgtataagaggaataaagacatacgaagtagacattttgatgagtggttaatgcaagcagacattatacaaaagt
gtagactaatgtagtgcctttgtttcacagACTGAAACAGCAGAGCTTCCTGCTCTGATAGCATAAACCCAGGCAACCTAC
AATTGGTTTCAGAGTTAAAGgtcagaagaatactctctcagtgctcgtgtctacatagaaaacttaataagtcgaaaa
gaattcagtcataatgctctgtctttaaactataaagataatggtaaacagatcagaaaaacagattggatgggtta
ttttctcaatattacctctgttttattggttgaagctctctcttaaaagtctctgcaagcaggaattcgatatacaagtt
atcgatacagtagtaagtttgataggcattttttggttacataaagacacaaagctctttctttcacctgcataaagact
aaaattacatacctactgtttcattgaataataatcatcacttgcataactatgctttgcataaaacagcactgagtgcc
actttaaagaaactgttccattgtgtcagAATCCTTCAGGTTCTGTTCGCTAGATGTGAGTGCCATGTTTTGGGAAAGGCC
GGTTGTAAAGGCCATGTATCACTAAGTCTGCGAAGATGAGTTTCTCTTTGGAAGCTCTGGATGCTTGGCAGTGGGAAAAA
CTTTATACCTGGCACTTCGCAAGAGtaagtggaatctcagctgaaagagatcttgcagcatttgcctgataatgtagt
ggcagcttaccaaaattggagctatgaccatgcaagcagaaacagagatgagggctgcacatcccaattatccccattgt
tagaattatccctgcataggtaaatgctcagtaagcattatgactattatgcatatagttatttagatatacagtaaat
aaaaagagtttctgagccttcaaatgatgaaaattatcctgtacagtgagaatacaaaaagaatgtagataaatttggaaa
ctggatatacaaaaatgaacaaccaagcataaatttggctgctttgtttatttaggtCCAGATTTACAGATAGTCCAG
TGCCGTGATGTGATAACTCTGCTGTGTAGCTTTGGGAAATTTGGAATCAGAGAGATCAGgtatgtaattccc caaggagtgt
ttgttttccctcatctctgtcctgtcagctgggttttaagtgacaggttaaacctaggcttgagcttgaagaatctgaaaag
atcctaaagagagagagatttgtttaaaaaaaatcaataagaatgacatccctgactgaaagttctatttaaaatggaacctag
gctgggctctcagggctgagggcagagctgtgaaactgggagcagagttgtagtgagcagattgtgccaactgca
ctccagcttaggagcagagcaagactgtctcaaaaaaaaataaaagtgaaactgctttaaattaaaaggt
tactctcacatcacccttttctcttatatttggcttagGGCATTGTTTTGTTCTCTGATGATGAAAGTGAAGCAAGTGA
CTACTAGTCTGGAATAATAAAGCTGTGCTTGCCCTGACAAAAGAGGGGCTAGTGTAGTACAGTGGGACCTTCTGATCAA
CAAGTAGAGTCTAGCAGTTTGCAGAAAGTGGAGGtaagaaaagcattgattgatttttaactattagatgaagaatgattt
atcacaggtttcagagaagaattgggtaactagatctcgttttctgtgctgggggtgtaataaagcattgaccgcatcaaca
ctaggttatgacatagaagcaggttagtgaggtagaacagacatgctcagggatgaagcacaagaatgcttgacgacatact
ggtagatattagaagaataatgctcattgtagaagaatttagtatcacagctttctgggttagattttttctctgata
taggttagttatataatgtagtccaacatgaggagaagggctactagagactgctttagtgcaaaactgactttcata
ctgttttaatacagAGGCAAGAAAACCAATTTTGTAGTCCCTGAGGAGACTACTAATCTTTGCTGAGGTCAGGAGGAT
CGAAGAAGCTCTGCTGGTACTACTATTATGAACAACATTGTTATTTGtaagcttctcctaggtctcagttcccctactc
ttagttaggagataatacctcaatttcaaggttgtgtgagagataaataagagagatgtagtaaacatgattgtgggtttg
tgtgtctgtgtgtgtttttgtgtgtttgagacagagctcctcctatcaccaggtgggagacagtggtatgatgctg
aacatacaaatgaaactattggcaaaatcaaccacaggtctacttttacctaactatgactaaagaaaactaaggagctc
tagtttttcccgtgctcactcaagacatgctatgataaagaaaatatttctgaatactggtttgtggagaagatgtg
atcagcttattttttgttactaaagGAATTTAAAACATGCTCACTCTGAAAAGATGCACATTGATGATTTCTACCAAG
CTTCAGTCTGTCACAAAGCCTATTCTGAAATGtaagtaagtaagctggctgggaccactttgtgttactgttggctcattaa
gtgagagcagtaagtcataagcagtaataagaacacccctcccctgggactctttttcaaatggatagcaattgtttgtcctt
tctgaaaatttaggatgaaaataagttgtgtgactgattgctggtaattttctgctgtgaagcagaagtcagctcaggt
aataactcaacagttctcagacggcagggaaaaaaatacaagcaggtggttaaatcctggatactcagagcttactcaggtcatt
gctttaaattgtttgtttttgtctcctgagacatttttttctctgacatactttagcagcttatttgggataattttttt
ctccgaaattagGGCTCTCTTTATGTCTGAGTATCCCTGTGCCAAAGAGTGTGCTGTGCCAAGCCCTGTGTTTTCAG
CTCATTTGATTAAACCTAAGACGACTCTAGCCTGGGTGTGATGCTGACTGCTTCTCCAGGGCAAGGCTGCGAGGcaagtg
tgcaataactgctactctatgggtgggacattctgaaaagcactgtgcaaacacaaaaactagatattttgtgtcacttagaa
aatggaaaacatgaaactatatacatgataggaatattcagaagtaggtagaagattagggactgaggtcaaacccagga
ttcctcgtacctgactctactagtagtaggggttatataggttttaattaggttttccaagcctgtttttt

```

**Ex5**

**Ex6**

**Ex7**

**Ex8**

**Ex9**

**Ex10**

**Ex11**

**Ex12**

**Exons 5-12. Insert structure:** ivs4 (266 bp) - ex5 (830 bp) - ivs5 (364 bp) - ex6 (72 bp) - ivs6 (209 bp) // ivs6 (191 bp) - ex 7 (162 bp) - ivs7 (114 bp) // ivs7 (250 bp) - ex 8 (86 bp) - ivs8 (200 bp) // ivs8 (200 bp) - ex9 (167 bp) - ivs9 (200 bp) // ivs9 (200 bp) - ex 10 (117 bp) - ivs10 (200 bp) // ivs10 (200 bp) - ex11 (88 bp) - ivs11-1 (200 bp) // ivs11-2 (200 bp) - ex 12 (149 bp) - ivs12 (250 bp)

Figura 19. Secuencia del inserto en el minigén mgPALB2\_ex5-12.

Todos los productos RT-PCR, procedentes de la transfección de las células MCF-7 con los minigenes *wild type* presentados anteriormente, fueron caracterizados mediante electroforesis en geles de agarosa, como se indica en la sección 2.2.4 de Materiales y Métodos.

**Tabla 4. Secuencia de oligonucleótidos utilizados para la construcción de los minigenes: mgPALB2\_ex1-3, mgPALB2\_ex4-6 and mgPALB2\_ex5-12.**

Exones	Generación Inserto	Clonaje <sup>1</sup>	Oligonucleótidos (5' → 3') <sup>2</sup>
Exón1	Hi-Fi PCR	OEP	[AGTCACCTGGACAACCTCAAAGGCACCTTT]GGCTGC TCTTTCGTTCTGTC
			[TGCATAACCTGTTCTATCATCTGGAACAT]TGCGCTG CTACTGTAATGAAA
Exones 2-3	Síntesis Génica	Restricción	-
Exón 4	Hi-Fi PCR	Restricción	[TATATATCTAGA]GTTAAGAGAAGAGATTGTGTGA [TATATAGGATCC]ATACATTTTCCTTTTCAGTGTT
			[GGTGGCGCCGCTCTAGAACTAGTGGATCCCCGG]A GTCATGGATGGGAAAAGTAA
Exones 5-6	Hi-Fi PCR	OEP	[GACGGTATCGATAAGCTTGATATCGAATTCCTGCA]T TGGCATAGAACTTTAAGAGG
			[GCCAATGCAGGAATTCGATATCAAGCTTATCGATA]C ATGATGTAAGTTTGATAGGCA
Exón 7	Hi-Fi PCR	OEP	[AAAGCTGGGTACCGGGCCCCCCTCGAGGTCGACG]C CTCATCTCTGTTCTGCCTT
			-
Exones 8-12	Síntesis Génica	Restricción	-

<sup>1</sup> OEP= *Overlapping Extension PCR*. <sup>2</sup> Las colas de restricción están entre corchetes.

### 2.2.1. Transformación.

Las células ultracompetentes de *Escherichia coli* XL10-Gold (Agilent, Santa Clara, Estados Unidos) tienen el siguiente genotipo: TetrD(*mcrA*)183 D(*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte* [F' *proAB lacIqZDM15 Tn10* (Tetr) Amy Camr] (donde los genes enumerados significan alelos mutantes y los genes con el símbolo F', son de tipo *wild type*). Estas bacterias fueron transformadas siguiendo el protocolo sugerido por la casa comercial y cultivadas en medio Luria Bertani (LB) (Fisher Scientific), incluyendo todos los minigenes finales y los constructos intermedios descritos en apartados anteriores.

### **2.2.2. Selección de colonias recombinantes.**

Las bacterias transformadas se sembraron en placas de medio LB cuya composición es: 1% Triptona, 0,5% extracto de levadura, 1% NaCl premezclado con agar-LB-Miller (Merck/VWR, West Chester, PA, EE.UU.) suplementadas con ampicilina a una concentración final de 100 µg/ml, como método de selección de colonias transformadas. Las bacterias fueron incubadas a 37°C durante 18 horas.

Además, en los primeros pasos de las construcciones de nuestros minigenes y para distinguir las colonias bacterianas que incluían los plásmidos con los insertos de interés, de aquellas que únicamente incluían el vector vacío, se trataron las placas con IPTG (Thermo Scientific) y X-Gal (Thermo Scientific) para poder seleccionar como positivas, las bacterias blancas (con inserto) respecto a las azules, que únicamente albergaban el vector pMAD.

### **2.2.3. Rastreo de colonias por PCR.**

Mediante el rastreo de colonias por PCR, conseguimos de forma rápida y fácil constatar que el proceso de inserción se realizó de forma eficaz y precisa. Este método consiste en la amplificación directa por PCR del extracto resultante del hervido de cada colonia aislada, utilizando el Mix de PCR DreamTaq Green 2x (Thermo Scientific) y 2 oligonucleótidos específicos del vector situados en las zonas flanqueantes al sitio de clonación múltiple del vector pMAD (T7: 5'-TAATACGACTCACTATAGGG-3', y pSAD-SEQ-RV: 5'-GGATAACAATTTACACAGG-3'). De esta manera se pueden distinguir aquellos vectores que incluyen el inserto deseado del resto, por el tamaño del producto amplificado. Además, esta misma técnica ha sido empleada utilizando un oligonucleótido localizado en el interior del inserto, y otro situado en zonas flanqueantes al sitio de clonación múltiple.

Las reacciones de amplificación se realizaron en un termociclador Veriti 96 (Applied Biosystems, Madrid, España) con el programa: 95°C 1-3 min; 35 ciclos x [95°C 30" - 50°C 30" - 72°C 1 min por kb]; 72°C 5 min con la mezcla presentada en la Tabla 5.

**Tabla 5. Mezcla de amplificación en el rastreo de colonias por PCR.**

<b>DreamTaq Mix</b>	<b>12,5µL</b>
<b>Oligonucleótido FW</b>	0,2 µM
<b>Oligonucleótido RV</b>	0,2 µM
<b>Agua</b>	11,5 µL
<b>Volumen final</b>	25 µl

---

#### **2.2.4. Electroforesis en gel de agarosa.**

Todos los productos amplificados fueron verificados mediante electroforesis en gel de agarosa al 1,2% en tampón TBE x1 (Tris 100 milimolar (Mm), Borato 90 mM, EDTA 1mM) para aquellos fragmentos inferiores a 1 kb y al 1% para los fragmentos superiores, usando siempre el estándar de tamaño 1 kb Plus DNA ladder de 1 kb (Thermo Scientific).

#### **2.2.5. Extracción de DNA plasmídico y comprobación de las construcciones.**

Una vez las colonias habían sido seleccionadas mediante el rastreo especificado anteriormente, se procedió a extraer el DNA plasmídico de los clones que contenían el inserto. Para ello se utilizó el kit GeneJET™ Plasmid Miniprep Kit (Thermo Scientific). Esta extracción se realizó a partir de un cultivo bacteriano en saturación tras 18 horas de incubación a 37°C en medio LB (Caldo-LB-Miller, Merck) suplementado con ampicilina a una concentración final de 100 µg/ml. Por último, se prepararon reacciones de secuencia utilizando oligonucleótidos específicos de nuestro minigén, que fueron procesadas por la empresa Macrogen y analizadas posteriormente en nuestro laboratorio con la ayuda del programa SnapGene Viewer v5.3.2 ([www.snapgene.com](http://www.snapgene.com)).

### 2.3. Mutagénesis dirigida.

La técnica de mutagénesis dirigida permite generar cualquier alteración detectada en el DNA de un paciente o mutaciones artificiales en los minigenes construidos. Por lo tanto, no es necesario obtener muestras biológicas de pacientes.

#### 2.3.1. Mutaciones puntuales.

Las mutaciones reproducidas en los minigenes mgPALB2\_ex1-3, mgPALB2\_ex4-6 y mgPALB2\_ex5-12 se generaron siguiendo las recomendaciones del kit de mutagénesis QuikChange Lightning Site-Directed Mutagenesis (Agilent). El protocolo de mutagénesis dirigida consta de una amplificación mediante PCR de alta fidelidad con la enzima QuikChange, de un minigén *wild type* (wt) con oligonucleótidos que introducen la mutación mediante la mezcla representada en la Tabla 6, con el programa: 95°C, 2 min; 18 ciclos x [95°C, 20'' - 55°C, 10'' - 68°C, 30'' por kb]; 68°C 10'.

**Tabla 6. Mezcla de amplificación PCR de mutagénesis, PCR inserción y deleción.**

<b>Buffer QuickChange 10x</b>	<b>2,5µl</b>
<b>Plásmido molde</b>	50 nanogramos(ng)
<b>Oligonucleótido Fw 10mM</b>	1 µl
<b>Oligonucleótido Rv 10mM</b>	1 µl
<b>XL dNTP Mix</b>	0,5 µl
<b>Enzima QuickChange</b>	1 unidad
<b>Quick solution</b>	0,75 µl
<b>Volumen final</b>	25 µl

Una vez finalizada la amplificación, se realiza un tratamiento con DpnI (Thermo Scientific) para digerir el DNA metilado que corresponde al plásmido molde sin mutación.

A continuación, las células ultracompetentes XL10-Gold (Agilent) se transformaron con los minigenes mutantes siguiendo el proceso descrito en el apartado 2.2.1. Por último, estos minigenes portadores de mutaciones fueron extraídos y comprobados mediante secuenciación de la región donde se han producido la mutación de forma artificial de forma similar al apartado 2.2.5.

Un total de 88 oligonucleótidos específicos de mutagénesis dirigida fueron diseñados para el estudio de 44 variantes procedentes del proyecto europeo BRIDGES en los minigenes mgPALB2\_ex1-3, mgPALB2\_ex4-6 y mgPALB2\_ex5-12 (Tabla 7). Además, se emplearon 22 oligonucleótidos para el estudio de 11 variantes que habían sido seleccionadas de la base de datos ClinVar en el minigén mgPALB2\_ex1-3 (Tabla 8) y 18 oligonucleótidos fueron utilizados para estudiar funcionalmente 9 variantes exónicas de los exones 11 y 12, en el minigén mgPALB2\_ex5-12, procedentes también de la base de datos ClinVar (Tabla 9). Todas estas variantes fueron seleccionadas en los estudios bioinformáticos previos siguiendo las directrices definidas en el apartado: 1.2.

**Tabla 7. Mutaciones y secuencia de oligonucleótidos utilizados para la mutagénesis dirigida de las variantes BRIDGES.**

Variantes	Exón/Intrón	Oligonucleótidos (5'→3')
c.47A>G	Ex1	CAGCTGTGAGGAGAAGGAAAGGGTGCCGGGGGTGCGGGAA
		TTCCCGCACCCCGGCACCCTTTTCCTTCTCCTCACAGCTG
c.48G>A	Ex1	GCTGTGAGGAGAAGGAAAAAGTGCCGGGGGTGCGGGAAG
		CTTCCCGCACCCCGGCACTTTTTCCTTCTCCTCACAGC
c.48+1G>A	Ivs1	GTGAGGAGAAGGAAAAGATGCCGGGGGTGCGGGAAGGG
		CCCTTCCCGCACCCCGGCATCTTTTCCTTCTCCTCAC
c.48+2T>C	Ivs1	AGCTGTGAGGAGAAGGAAAAGGCGCCGGGGGTGCGGG
		CCCGCACCCCGGCGCCTTTTTCCTTCTCCTCACAGCT
c.48+4C>T	Ivs1	AGCTGTGAGGAGAAGGAAAAGGTGTCGGGGGTGCGGG
		CCCGCACCCCGACACCTTTTTCCTTCTCCTCACAGCT
c.48+7G>C	Ivs1	TGTGAGGAGAAGGAAAAGGTGCCGGGGGTGCGGGAAGGG
		CCCTTCCCGCACCCCGGCGCACCTTTTTCCTTCTCCTCACA
c.49-2A>T	Ivs1	TTCAGAACACTCAAAGGTATCAGTCAGACTACTGAAATGTA
		TACATTTAGTAGTCTGACTGATACCTTTGAGTGTCTGAA
c.108+1G>A	Ivs2	GACACTAGCCCGCCTTCAGATAAGTGAATCGTATTCTCA
		TGAGAATACGATTCACTTATCTGAAGGCGGGCTAGTGTC
c.108+2T>C	Ivs2	ACTAGCCCGCCTTCAGGCAAGTGAATCGTATTCTC
		GAGAATACGATTCACTTGCCTGAAGGCGGGCTAGT
c.109-6_109-4del	Ivs2	TTCTGGGGCTGTTTTTGTCTCTAGCGTGCCCAAAGAGCTGA
		TCAGCTCTTTGGGCACGCTAGAGACAAAACAGCCCGAGAA
c.109-2A>G	Ivs2	GGGGCTGTTTTTGTCTCCTCTGGCGTGCCCAAAGAGCTG

		CAGCTCTTTGGGCACGCCAGAGGAGACAAAAACAGCCCC
<b>c.211+1G&gt;A</b>	Ivs3	CCGCAGCTAAAACACTCAGATAAATCTAGACCATTCACT AGTGAATGGTCTAGATTTATCTGAGTGTTTTAGCTGCGG
<b>c.211+5del</b>	Ivs3	GCAGCTAAAACACTCAGGTAATCTAGACCATTCACTTATGCC GGCATAAGTGAATGGTCTAGATTACCTGAGTGTTTTAGCTGC
<b>c.1684+4A&gt;G</b>	Ivs4	ATTTATTCAAGTGAAAGGTAGATCAAGATGTGTTTGATGAT ATCATCAAACACATCTTGATCTACCTTTCACTTGAATAAAT
<b>c.1685-2A&gt;G</b>	Ivs4	ACTGTTTCAAATCAAAGTCATTATTATTTGATTAGAGTGGG CCCCTCTAATCAAATAATAATGACTTTGATTTGAAACAGT
<b>c.1685-2A&gt;C</b>	Ivs4	GTGACTTATTTTTCTTCTTTCCGGGAAGAAAAGTCGTCATCA TGATGACGACTTTTCTTCCGAAAGAAGAAAAATAAGTCAC
<b>c.2513A&gt;C</b>	Ex5	CATAAACATTCCGTCGAACCGGTACAATCCATTTCTCT AGAGGAAATGGATTGTACCGGTTGACGGAATGTTTATG
<b>c.2515-2A&gt;G</b>	Ivs5	ATGTGACTTTTGTTTTACCGACTGAAACAGCAGAGCTT AAGCTCTGCTGTTTCAGTCCGTGAAAAAAAAGTCACAT
<b>c.2586+4A&gt;T</b>	Ivs6	CAGAACTGTTGCCATTGTGTCGGAATCCTTCAGGTTCTT AGGAACCTGAAGGATTCGACACAATGGCAACAGTTCTG
<b>c.2587-2A&gt;G</b>	Ivs6	CAGAACTGTTGCCATTGTGTCGGAATCCTTCAGGTTCTT AGGAACCTGAAGGATTCGACACAATGGCAACAGTTCTG
<b>c.2748+1G&gt;T</b>	Ivs7	ATACCTGGCACTTCGCAGAGTTAAGTGGGAATCTCGAGC GCTCGAGATTCCCCTTAACTCTGCGAAGTGCCAGGTAT
<b>c.2748+1G&gt;A</b>	Ivs7	GGAAAACTTTATACCTGGCACTTCGCAGAGATAAGTGGG CCCCTTATCTCTGCGAAGTGCCAGGTATAAAGTTTTTCC
<b>c.2748+2dup</b>	Ivs7	ACCTGGCACTTCGCAGAGGTAAAGTGGGAATCTCGAGCTGA TCAGCTCGAGATTCCCCTTTACCTCTGCGAAGTGCCAGGT
<b>c.2748+4A&gt;T</b>	Ivs7	CTGGCACTTCGCAGAGGTATGTGGGAATCTCGAGCTGAA TTCAGCTCGAGATTCCCACATACCTCTGCGAAGTGCCAG
<b>c.2749-1G&gt;T</b>	Ivs7	GGCTGCTTTGTTTTATTTATGTTCCAGTATTACAGATAGTT AACTATCTGTAATACTGGAACATAAATAAAACAAAGCAGCC
<b>c.2750T&gt;C</b>	Ex8	TTGTTTTATTTAGGCTCCAGTATTACAGATAGTTCCA TGGAACCTATCTGTAATACTGGAGCCTAAATAAAACAA
<b>c.2834G&gt;C</b>	Ex8	TTTGAAATCAGAGAGATCACGTATGTAATCCCAGGAGT ACTCCTTGGAATTACATACGTGATCTCTCTGATTTCCAAA
<b>c.2834+1G&gt;A</b>	Ivs8	TTGAAATCAGAGAGATCAGATATGTAATCCCAGGAGTG CACTCCTTGGAATTACATATCTGATCTCTCTGATTTCCAA
<b>c.2834+2T&gt;C</b>	ivs8	AAATTTGGAATCAGAGAGATCAGGCATGTAATCCCAGG CTTGGAATTACATGCCTGATCTCTCTGATTTCCAAATTT
<b>c.2834+3A&gt;G</b>	ivs8	GGAAATCAGAGAGATCAGGTGTGTAATCCCAGGAGTGAT ATCACTCCTTGGAATTACACACCTGATCTCTCTGATTTCC
<b>c.2834+5G&gt;A</b>	ivs8	ATCAGAGAGATCAGGTATATAATCCCAGGAGTGATTTGT ACAAATCACTCCTTGGAATTATATACCTGATCTCTCTGAT
<b>c.2834+6T&gt;C</b>	ivs8	TCAGAGAGATCAGGTATGCAATCCCAGGAGTGATTTGT ACAAATCACTCCTTGGAATTGCATACCTGATCTCTCTGA
<b>c.2996+4A&gt;G</b>	Ivs9	TTTGCAGAAGATGGAGGGTAGGAAAAGCATTGATTGATT AATCAATCAATGCTTTTCTACCCTCCATCTTCTGCAA
<b>c.2997-2del</b>	Ivs9	TTTCATACTGTTTTAATTACGAGGCAAAGAAAACCAATTTT AAAATTGGTTTTCTTTGCCTCGTAATTAAAACAGTATGAAA
<b>c.2997-1G&gt;A</b>	Ivs9	TTTCATACTGTTTTAATTACAAAGGCAAAGAAAACCAATTTT AAAATTGGTTTTCTTTGCCTTTGTAATTAAAACAGTATGAAA
	Ex10	CAACATTGTTATTTAGTAAGCTTTCCCTCTAGGTCTCTC

<b>c.3113G&gt;A</b> (p.Trp1038Ter)		GAGGACCTAGAGGGAAAGCTTACTAAATAACAATGTTG
<b>c.3113+3A&gt;G</b>	Ivs10	AACATTGTTATTTGGTGAGCTTTCCTCTAGGTCCTC GAGGACCTAGAGGGAAAGCTCACCAAATAACAATGTT
<b>c.3113+5G&gt;C</b>	Ivs10	TGTTATTTGGTAACCTTTCCTCTAGGTCCTCAGTTC GAACTGAGGACCTAGAGGGAAAGGTTACCAAATAACA
<b>c.3201+1G&gt;A</b>	Ivs11	ACAAAGCCTATTCTGAAATGATAAGTAATGACTGGCTGGGA TCCCAGCCAGTCATTACTTATCATTTTCAGAATAGGCTTTGT
<b>c.3201+3_3201+4insTG</b>	Ivs11	ACTGTTTCAAATCAAAGTCATTATATTTGATTAGAGTGGG CCCCTCTAATCAAATAATAATGACTTTGATTTGAAACAGT
<b>c.3201+6T&gt;A</b>	Ivs11	CTATCTGAAATGGTAAGAAATGACTGGCTGGGACCACT AGTGGTCCCAGCCAGTCATTTCTTACCATTTTCAGAATAG
<b>c.3348C&gt;T</b> (p.Gly1161=)	Ex12	CTTCTCCAGGGCAGGCTGGTAGGCAAGTGTGCATAACT AGTTATGCACACTTGCCTACCAGCTGCCCTGGAGGAAG
<b>c.3350+4A&gt;G</b>	Ivs12	AGGCTGGCAGGCAGGTGTGCATAACTGCTACTCTATGG CCATAGAGTAGCAGTTATGCACACCTGCCCTGCCAGCCT
<b>c.3350+5G&gt;A</b>	Ivs12	GGCTGGCAGGCAAATGTGCATAACTGCTACTCTATGG CCATAGAGTAGCAGTTATGCACATTTGCCTGCCAGCC

**Tabla 8. Variantes y secuencia de oligonucleótidos utilizados para la mutagénesis dirigida de las variantes procedentes de ClinVar.**

<b>Variantes</b>	<b>Exón/Intrón</b>	<b>Oligonucleótidos (5'→3')</b>
<b>c.46A&gt;G</b>	Ivs1	CTGTGAGGAGAAGGAAGAGGTGCCGGGGGTGCGGGAAGGG CCCTTCCCGCACCCCGGCACCTCTTCTTCTCCTCACAG
<b>c.48+1G&gt;C</b>	Ivs1	GCTGTGAGGAGAAGGAAAAGCTGCCGGGGGTGCGGGAAGG CCTTCCCGCACCCCGGCAGCTTTTCTTCTCCTCACAGC
<b>c.48+1del</b>	Ivs1	CAGCTGTGAGGAGAAGGAAAAGTGCCGGGGGTGCGGGAAG CTTCCCGCACCCCGGCACCTTTTCTTCTCCTCACAGCTG
<b>c.48+2T&gt;G</b>	Ivs1	AGCTGTGAGGAGAAGGAAAAGGGGCCGGGGGTGCGGGAAG CTTCCCGCACCCCGGCCCTTTTCTTCTCCTCACAGCT
<b>c.49-2del</b>	Ivs1	TGCCAGTATTGTTGGTGTTTTTCTTCTTCCGTTAAAGGA TCCTTTAACGGAAGAAGAAAAACCAACAATACTGGGCA
<b>c.49-1G&gt;A</b>	Ivs1	TTCTTCCAATTAAGGAGAAATTAGCATTCTTGAAAAGGG CCCTTTTCAAGAATGCTAATTTCTCCTTTAATTGGAAGAA
<b>c.49-1del</b>	Ivs1	TGCCAGTATTGTTGGTGTTTTTCTTCTTCCATTAAGGA TCCTTTAATGGAAGAAGAAAAACCAACAATACTGGGCA
<b>c.108+1_108+2insC</b>	Ivs2	CCTTCAGGCTAAGTGAATCGTATTCTCAAATTAAGGTGTT AACACCTTAATTTGAGAATACGATTCACCTTAGCCTGAAGG
<b>c.109-2A&gt;C</b>	Ivs2	TTTGTCTCCTCTCGCGTGCCCAAAGAGCTGAAAAGATTAA TTAATCTTTTCAGCTCTTTGGGCACGCGAGAGGAGACAAA
<b>c.211+1G&gt;T</b>	Ivs3	CCGAGCTAAAACACTCAGTTAAATCTAGACCATTCACTT AAGTGAATGGTCTAGATTTAACTGAGTGTTTTAGCTGCGG
<b>c.211+2T&gt;G</b>	Ivs3	ACCGAGCTAAAACACTCAGGCAAATCTAGACCATTCACT AGTGAATGGTCTAGATTTGCCTGAGTGTTTTAGCTGCGGT

**Tabla 9. Variantes y secuencia de oligonucleótidos utilizados para la mutagénesis dirigida de las variantes exónicas de los exones 11 y 12 de *PALB2*, procedentes de ClinVar.**

Variantes	Exón/Intrón	oligonucleótidos (5'→3')
<b>c.3155A&gt;T</b> <b>p.(Asp1052Val)</b>	Ex11	ATGCACATTGTTGATTCTTACCAAGCTTCAGTCTGTCACA
		TGTGACAGACTGAAGCTTGGTAAGAATCAACAATGTGCAT
<b>c.3157G&gt;T</b> <b>p.(Asp1053Tyr)</b>	Ex11	ACTCCTGAAAAAGATGCACATTGATTATTCTTACCAAGCT
		AGCTTGGTAAGAATAATCAATGTGCATCTTTTTTCAGGAGT
<b>c.3161C&gt;T</b> <b>p.(Ser1054Phe)</b>	Ex11	GATGCACATTGATGATTTTTTACCAAGCTTCAGTCTGTCAC
		GTGACAGACTGAAGCTTGGTAAAAATCATCAATGTGCATC
<b>c.3242A&gt;G</b> <b>p.(Glu1081Phe)</b>	Ex12	CCTGTGCCAAAGGGAGTGAGTCGTTGCGAAGCCCTGTGTT
		AACACAGGGCTTCGCAACGACTCACTCCCTTTGGCACAGG
<b>c.3244A&gt;G</b> <b>p.(Ser1082Gly)</b>	Ex12	GTCATCCCTGTGCCAAAGAGGGTGAGTCGTTGCGAAGCCC
		GGGCTTCGCAACGACTCACCCCTCTTTGGCACAGGGATGAC
<b>c.3251C&gt;T</b> <b>p.(Ser1084Leu)</b>	Ex12	CAAAGAGAGTGAGTTGTTGCGAAGCCCTGTGTTTCAGCTC
		GAGCTGAAACACAGGGCTTCGCAACAACACTCACTCTCTTTG
<b>c.3294_3298del</b> <b>p.(Lys1098fs)</b>	Ex12	TGTTTCAGCTCATTTGTGATTAACCCTAACTCTCAGCGTGG
		CCACGCTGAGAGTTAGGGTTAATCACAATGAGCTGAAACA
<b>c.3306C&gt;G</b> <b>p.(Ser1102Arg)</b>	Ex12	CTAAGACGACTCTCAGGGTGGGTGTGATGCTGTACTGTCT
		AGACAGTACAGCATCACACCCACCCTGAGAGTCGTCTTAG
<b>c.3306C&gt;T</b> <b>p.(Ser1102=)</b>	Ex12	AACCCTAAGACGACTCTCAGTGTGGGTGTGATGCTGTACT
		AGTACAGCATCACACCCACACTGAGAGTCGTCTTAGGGTT

**Tabla 10. Variantes y secuencia de oligonucleótidos utilizados para la mutagénesis dirigida de las microdeleciones de los exones 11 y 12 de *PALB2*, procedentes de ClinVar.**

Variantes	Exón/Intrón	Oligonucleótidos (5'→3')
<b>c.3116del130</b>	Ex11	ATCAGCTTATTTATTTTTTGTATCTAAGGATGCACATTGATGATTCTTACCAAGCTTCAG
		CTGAAGCTTGGTAAGAATCATCAATGTGCATCCTTAGATAACAAAAATAAATAAGCTGAT
<b>c.3143del129</b>	Ex11	GGAATTTAAAACTGGTCAACTCCTGAAAATCAGTCTGTGCACAAAGCCTATTCTGAAATG
		CATTTCAGAATAGGCTTTGTGACAGACTGATTTTCAGGAGTTGACCAGTTTTTAAATCC
<b>c.3169del130</b>	Ex11	AAAAAGATGCACATTGATGATTCTTACCAATGGTAAGTAATGACTGGCTGGGACCACTT
		AAGTGGTCCCAGCCAGTCATTACTTACCATTTGGTAAGAATCATCAATGTGCATCTTTTT
<b>c.3204del130</b>	Ex12	TGGGATATTTATTTTTCTCCGAAATTAGGGTGCCAAAGAGAGTGAGTCGTTGCGAAGCCC
		GGGCTTCGCAACGACTCACTCTCTTTGGCACCctaatctcgagaaaaataatcccc
<b>c.3229del130</b>	Ex12	TAGGGGCTTCTCTTTATTGTCTGAGTCATAGCCCTGTGTTTCAGCTCATTGTGATTAAC
		GTTAATCACAATGAGCTGAAACACAGGGCTATGACTCAGGACAATAAAGAGAAGCCCTTA
<b>c.3254del144</b>	Ex12	GTCATCCCTGTGCCAAAGAGAGTGAGTCGTAATCTCAGCGTGGGTGTGATGCTGTACTGT
		ACAGTACAGCATCACACCACGCTGAGAGTACGACTCACTCTCTTTGGCACAGGGATGAC
<b>c.3293del130</b>	Ex12	CTGTGTTTCAGCTCATTGTGATTAACCCTAACTGTCTTCTCCAGGGCAGGCTGGCAGGC
		GCCTGCCAGCTGCCTGGAGGAAGCAGTTAGGGTTAATCACAATGAGCTGAAACACAG
<b>c.3318del130</b>	Ex12	CCCTAAGACGACTCTCAGCGTGGGTGTGATCAGGCAAGTGTGCATAACTGCTACTCTATG
		CATAGAGTAGAGTTATGCACACTTGCTGTATCACACCCACGCTGAGAGTCGTCTTAGGG

### **2.3.2. Microdeleciones.**

La mutagénesis dirigida también se empleó para introducir microdeleciones exónicas. El estudio de microdeleciones permite detectar mediante ensayos funcionales de *splicing* secuencias exónicas esenciales (*enhancers* o silenciadores) en la regulación del *splicing*. Mediante este procedimiento introdujimos 8 microdeleciones en el minigén mgPALB2\_ex5-12 utilizando 16 oligonucleótidos específicos, cuyas secuencias se muestran en la Tabla 10. Las microdeleciones se generaron en el mgPALB2\_ex5-12 con el kit de mutagénesis QuickChange Lightning Site-Directed Mutagenesis (Agilent). A continuación, los minigenes producto de las deleciones se transformaron en bacterias XL10-Gold y se comprobó la inclusión de la microdelección mediante secuenciación como hemos visto en apartados anteriores.

### **2.4. Transfección.**

La transfección es la técnica utilizada para introducir los minigenes *wild type* y mutantes en células humanas en cultivo. Para este estudio se utilizó la línea celular humana MCF-7, que son células tumorales obtenidas de efusión pleural de una paciente con carcinoma mamario metastásico (Brooks et al., 1973). En el protocolo utilizado, se sembraron aproximadamente  $2 \times 10^5$  células MCF-7 (células de adenocarcinoma mamario humano) hasta alcanzar un 90% de confluencia en 0,5 ml de medio MEME (Minimum Essential Medium Eagle, 10% Fetal Bovine Serum, 1% Non-Essential Amino Acid Solution, 2mM) (Sigma-Aldrich, St. Louis, MO).

A continuación, las células fueron transfectadas con 1 microgramo ( $\mu\text{g}$ ) de cada constructo y 2 microlitros ( $\mu\text{l}$ ) de lipofectamina LTX (Life Technologies, Carlsbad, CA). A las 48 horas, incubamos las células con 300  $\mu\text{g}/\text{mL}$  de cicloheximida (Sigma-Aldrich) durante 4 horas, para inhibir el NMD. Por último, se extrajo y purificó mediante columna el RNA celular con el kit universal de purificación de RNA de Genematrix (EURx, Gdansk, Poland) incluyendo tratamiento con DNAsa I.

## 2.5. Retrotranscripción

Se obtuvo cDNA mediante el kit RevertAid First Strand cDNA Synthesis (Life Technologies, Carlsbad, CA, Estados Unidos) a partir de RNA purificado de células transfectadas con los minigenes *wild type* y mutantes. El proceso de retrotranscripción se llevó a cabo a partir de 400 nanogramos (ng) de RNA mediante la enzima RevertAid First Strand cDNA Synthesis Kit (Life Technologies) usando el oligonucleótido RTPSPL3-RV (5'TGAGGAGTGAATTGGTCGAA-3') específico del exón V2 del vector, con el que conseguimos amplificar únicamente los transcritos producidos por el minigén y no los procedentes de las células transfectadas. Este proceso se efectuó en 2 pasos: 1) incubación de 400 ng de RNA con 2 µl del oligonucleótido RTPSPL3-RV a 65° (volumen final de 12 µl) durante 5 minutos para conseguir la unión de los oligonucleótidos; 2) las muestras fueron incubadas durante 1 hora a 42°C (volumen final de 20 µl) y los componentes que se indican en la Tabla 11. Finalmente, se inactivó la retrotranscriptasa mediante incubación a 70°C durante 5 minutos.

**Tabla 11. RT-PCR MIX.**

<b>400ul RNA+ RTpspl3-RV+Agua</b>	<b>12 µl</b>
<b>Buffer 5x</b>	4 µl
<b>dNTPS</b>	2 µl
<b>Ribolock</b>	1 µl
<b>RevertAid h-minus RT</b>	1 µl
<b>VOLUMEN FINAL</b>	20 µl

---

A continuación, 40 ng de cDNA, fueron amplificados en un volumen final de 50 µl con los oligonucleótidos PSPL3\_RT-FW (5'-TCACCTGGACAACCTCAAAG-3') y RTpSAD-RV (Patent P201231427) utilizando la enzima *Hot start* Platinum-Taq DNA polymerase (Life Technologies). Para realizar este proceso, las muestras fueron desnaturalizadas a 94°C durante 2

minutos, seguido de 35 ciclos x [94°C/30 s, 60°C/30 s, 72°C (1 min/kb)] y un paso de extensión final de 5´ a 72°C. Por último, todos los productos PCR se analizaron en un gel de agarosa al 1% y fueron secuenciados por la empresa Macrogen (Macrogen, Madrid, España).

## **2.6. Análisis de transcritos.**

Los transcritos producidos por cada minigén fueron caracterizados mediante tres metodologías: electroforesis en gel de agarosa (sección 2.2.4), análisis de fragmentos mediante electroforesis capilar fluorescente (área de picos visualizados) y secuenciación Sanger capilar de los transcritos generados (Macrogen, Madrid, España), que en determinadas circunstancias venía precedida por extracción y purificación de bandas en gel de agarosa de bajo punto de fusión (Agarosa Low Melt, LM) al 0,8%, (Condalab, Madrid, España) y el uso del kit Wizard SV Gel and PCR clean-up system (Promega Biotech Ibérica, Madrid, España). Sólo en el caso de la variante c.1684+4A>G se realizó un análisis densitométrico del patrón de bandas visualizado en geles de agarosa.

### **2.6.1. Cuantificación de transcritos. Electroforesis capilar.**

La detección de transcritos mediante electroforesis capilar es una técnica de alta sensibilidad y resolución que permite distinguir diferencias mínimas en el tamaño de los fragmentos analizados. Los amplicones analizados por esta metodología fueron obtenidos mediante reacciones RT-PCR en las que se ha utilizado un oligonucleótido marcado con FAM (RT-pSPL3-FW o RT--pSAD-RV), cuya secuencia se ha detallado anteriormente. Las reacciones de amplificación se realizaron por triplicado a partir de cDNA y utilizando la polimerasa Platinum-Taq en un termociclador Veriti 96 (Applied Biosystems, Madrid, España) con el siguiente programa: 94°C 2 min; 26 ciclos x [94°C/30 s, 60°C/30 s, 72°C (1 min/kb)], y un paso final de 5´ a 72°C.

El análisis de los resultados se realizó con el programa informático Peak Scanner™

Software v1.0 (Thermo Fisher Scientific;

<https://www.thermofisher.com/order/catalog/product/4381867#/4381867>). La cuantificación relativa de cada transcrito se realizó obteniendo los datos del área de cada pico detectado y calculando la media aritmética y la desviación estándar de los tres valores obtenidos.

### **3. Interpretación clínica de variantes.**

Los laboratorios clínicos de genética molecular humana están detectando cada vez más variantes nuevas en pacientes, gracias a la secuenciación masiva y a los paneles de genes. Esto ha permitido asociar un gran número de genes como posibles causantes de enfermedad. Si bien algunas enfermedades están asociadas con un solo gen, otras están asociadas con un gran abanico genético.

En 2013, la ACMG y la AMP establecieron como objetivo el desarrollo de unas recomendaciones para el uso de una terminología única y común para clasificar variantes utilizando todas las evidencias científicas disponibles. Para evaluar las opiniones de la comunidad científica, se enviaron encuestas a más de 100 laboratorios de biología molecular en Estados Unidos y Canadá que se registraron en GeneTests.org, para comentar sus preferencias terminológicas. El resultado de la encuesta y el foro abierto determinó el uso de un sistema de nomenclatura de cinco niveles utilizando los términos: "patogénico", "probablemente patogénico", "significado incierto", "probablemente benigno" y "benigno". Una vez hubieron acuñado estos términos, era necesario crear un sistema de puntuación basado en un conjunto de criterios que permitía situar una variante en uno de los cinco niveles de clasificación.

Este sistema de clasificación está destinado a la interpretación de variantes observadas en pacientes con sospecha de trastornos hereditarios (principalmente mendelianos) en un entorno de laboratorio de diagnóstico clínico. No está destinado a la interpretación de variaciones somáticas, variantes farmacogenómicas o variantes en genes asociados con trastornos multigénicos complejos de herencia no mendeliana.

Este enfoque de clasificación puede ser algo más estricto de lo que los laboratorios de genética clínica han aplicado hasta la fecha. Aunque también puede aumentar la proporción de variantes clasificadas como de significado incierto. Se espera que este enfoque reduzca de forma

sustancial el número de variantes que se informadas como "causales" de la enfermedad sin tener suficiente evidencia de apoyo para esa clasificación. Es importante tener en cuenta que cuando un laboratorio clínico informa una variante como patogénica, esto puede alterar el tratamiento o la vigilancia de un paciente o eliminar a un miembro de la familia con genotipo negativo del plan de prevención, basado en esa determinación.

El sistema de clasificación se basa en dos conjuntos de criterios: uno para la clasificación de variantes patogénicas o probablemente patogénicas (Tabla 12) y una para clasificación de variantes benignas o probables benignas (Tabla 13) (Richards et al., 2015). Los criterios de patogenicidad para variantes patogénicas o posiblemente patogénicas se ponderan como: muy fuerte (PVS1), fuerte (PS1–4); moderado (PM1–6) o de apoyo (PP1–5) y están representados en la Tabla 12.

**Tabla 12. Criterios de clasificación de variantes patogénicas** (Adaptacion de Richards et al., 2015).

EVIDENCIAS DE PATOGENICIDAD	CATEGORÍAS
Muy fuerte	<ul style="list-style-type: none"> <li>• Variantes PVS1 (<i>nonsense, frameshift, variantes de sitio de splicing</i> (<math>\pm 1,2</math>), <i>deleciones de un solo exón o multiexónicas</i>) en genes donde la “pérdida de función” es un mecanismo conocido de la enfermedad.</li> <li>• Hay que ser cautelosos al clasificar:             <ol style="list-style-type: none"> <li>1. Variantes de genes donde “la pérdida de función” (LoF) no es un mecanismo de enfermedad conocido.</li> <li>2. Variantes de pérdida de función (LoF) en el extremo 3’.</li> <li>3. Variantes que causan <i>exon skipping</i>, pero dejan el resto del transcrito intacto.</li> <li>4. Variantes que inducen la producción de múltiples transcritos.</li> </ol> </li> </ul>

<p><b>Fuerte</b></p>	<p>1. <b>PS1:</b> Mismo cambio de aminoácidos que una variante previamente clasificada como patogénica sin importar cambio de nucleótidos. Ejemplo: Val→Leu causado por G&gt;C o G&gt;T en el mismo codón.</p> <p><b>Advertencias:</b> los cambios que afectan al <i>splicing</i> pueden tener una repercusión diferente.</p> <p>2. <b>PS2:</b> Variantes <i>de novo</i> en un paciente con la enfermedad y sin antecedentes familiares.</p> <p>3. <b>PS3:</b> Se han desarrollado ensayos funcionales <i>in vitro</i> o <i>in vivo</i> que apoyan un efecto dañino sobre el gen o producto génico.</p> <p>4. <b>PS4:</b> La prevalencia de la variante en los individuos afectados es significativamente superior comparada con la prevalencia en controles.</p>
<p><b>Moderado</b></p>	<ul style="list-style-type: none"> <li>• <b>PM1:</b> La variante se encuentra localizada en un punto caliente mutacional y/o dominio funcional crítico y bien establecido (por ejemplo, centro activo enzimático).</li> <li>• <b>PM2:</b> Ausencia de la variante en controles (o una frecuencia extremadamente baja si es recesiva) en el Exome Sequencing Project, 1000 Genomes Project, o Exome Aggregation Consortium.</li> <li>• <b>PM3:</b> Para trastornos recesivos, detectados en trans con una variante patológica.</li> <li>• <b>PM4:</b> Cambios en la longitud de la proteína como resultado de deleciones/inserciones en el marco en una región sin repeticiones o variantes de <i>stop-loss</i> (afectan al codón de parada canónico alterando la longitud del transcrito).</li> </ul>

	<ul style="list-style-type: none"> <li>• <b>PM5:</b> Variante <i>missense</i> nueva, donde se produce un cambio de aminoácido diferente a uno previamente clasificado como patogénico. Ejemplo: Arg156His es patogénica; y el cambio que se observa es Arg156Cys.</li> </ul>
<p><b>Criterios Secundarios</b></p>	<ul style="list-style-type: none"> <li>• <b>PP1:</b> Cosegregación con enfermedad en varios miembros de la familia en un gen conocido asociado a dicha enfermedad.</li> </ul>
	<ul style="list-style-type: none"> <li>• <b>PP2:</b> Variantes <i>missense</i> en un gen en el que variantes <i>missense</i> son una causa común de enfermedad.</li> </ul>
	<ul style="list-style-type: none"> <li>• <b>PP3:</b> Múltiples evidencias computacionales apoyan un efecto perjudicial sobre el gen o producto genético (conservación evolutiva, impacto en el <i>splicing</i>, etc.).</li> </ul>
	<ul style="list-style-type: none"> <li>• <b>PP4:</b> El fenotipo o los antecedentes familiares del paciente son muy específicos de una enfermedad con una sola etiología genética.</li> </ul>
	<ul style="list-style-type: none"> <li>• <b>PP5:</b> Una fuente fiable y acreditada informa que la variante es patógena, pero la evidencia funcional no está disponible para realizar una evaluación independiente.</li> </ul>

**Tabla 13. Criterios de clasificación de variantes Benignas** (Adaptacion de Richards et al., 2015).

EVIDENCIAS DE PATOGENICIDAD	CATEGORÍAS
Por si misma	<ul style="list-style-type: none"> <li>• <b>BA1:</b> La frecuencia de los alelos está por encima del 5% en Exome Sequencing Project, 1000 Genomes, o ExAC.</li> </ul>
Fuerte	<ul style="list-style-type: none"> <li>• <b>BS1:</b> La frecuencia alélica es mayor de la esperada para el trastorno.</li> <li>• <b>BS2:</b> La variante ha sido observada en un individuo sano para un trastorno recesivo (homocigótico), dominante (heterocigótico) o ligado al cromosoma X (hemicigótico) con una penetrancia esperada a una edad temprana</li> <li>• <b>BS3:</b> Los estudios funcionales <i>in vitro</i> o <i>in vivo</i> no muestran efectos perjudiciales en la función de las proteínas o el <i>splicing</i></li> <li>• <b>BS4:</b> Falta de segregación familiar.                      *. Advertencia: la presencia de fenocopias para fenotipos comunes (es decir, cáncer, epilepsia) puede imitar la falta de segregación entre los individuos afectados. Además, las familias pueden poseer más de una variante patogénica que contribuye a un trastorno autosómico dominante, lo que confunde aún más una aparente falta de segregación.</li> </ul>
Criterios Secundarios	<ul style="list-style-type: none"> <li>• <b>BP1 :</b> Variantes missense en un gen para el que principalmente han sido descritas como causantes de enfermedad las variantes truncantes.</li> <li>• <b>BP2:</b> observadas con otra variante en trans para un gen/trastorno dominante de alta penetrancia o en cis con cualquier patrón de herencia.</li> <li>• <b>BP3:</b> Delecciones o inserciones <i>in-frame</i> situadas en regiones con secuencias repetidas, en una región del gen de la que se desconoce su función.</li> </ul>

<ul style="list-style-type: none"> <li>• <b>BP4:</b> Varios programas informáticos sugieren que no produce impacto en los genes o sobre el producto genético (conservación, evolución, impacto en el <i>splicing</i>, etc.).</li> </ul> <p><b>*Advertencia:</b> BP4 solo se puede utilizar una vez durante la evaluación de una variante.</p>
<ul style="list-style-type: none"> <li>• <b>BP5:</b> Variantes encontradas en casos con una base molecular de la enfermedad alternativa.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>BP6:</b> Una fuente contrastada ha reportado recientemente la variante como benigna, pero las evidencias científicas no están disponibles para desarrollar una interpretación independiente.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>BP7:</b> Variantes sinónimas para las cuales los algoritmos predictivos de <i>splicing</i> no predicen ningún impacto en la secuencia consenso, ni la creación de nuevos sitios de <i>splicing</i>, y además el nucleótido afectado no está muy conservado evolutivamente.</li> </ul>

### 3.1. Recomendaciones propias en la clasificación de variantes de sitios de *splicing*.

En nuestra opinión, una vez que los datos de los ensayos funcionales de *splicing* estén disponibles (ya sea del análisis de minigén o del RNA derivado del paciente), los códigos predictivos de *splicing* no deberían contribuir a la clasificación de variantes, sino que deberían ser reemplazados por códigos funcionales de *splicing*. De lo contrario, surgirían inconsistencias internas en el Sistema de clasificación ACMG/AMP.

Además, el sistema ACMG/AMP asume que cada pieza de evidencia que contribuye a la clasificación final es independiente (Tavtigian et al., 2018), una suposición que difícilmente

cumplen los códigos predictivos de *splicing* y los códigos funcionales, ya que la mayoría de los análisis de *splicing* se realizan a partir de variantes bioinformáticamente preseleccionadas (Sanoguera-Miralles et al., 2020).

Varios documentos del panel de expertos de ClinGen-SVI y ClinGen reconocen el uso de minigenes para el desarrollo de ensayos funcionales de *splicing* (PS3/PS4) (Richards et al., 2015; Lee et al., 2018; Brnich et al., 2020). Sin embargo, la asignación de un código PS3/BS3 para los ensayos funcionales con minigenes (o de la misma manera, los ensayos de RT-PCR realizados en pacientes) no es obvia, especialmente si la variante de *splicing* produjo dos o más transcritos diferentes. Hasta donde sabemos, ni el ClinGen-SVI ni ningún otro panel de expertos han abordado este tema. El dr. Miguel de la Hoya con el que colaboramos, propone descomponer lecturas complejas en transcripciones individuales, asignando fortalezas de código PS3/BS3 a cada transcrito, individualmente y luego generar una fuerza de código PS3/BS3 combinada, basada en el juicio de expertos.

Siguiendo estas directrices, para clasificar clínicamente las variantes BRIDGES analizadas en esta tesis, se asignaron a los mRNA individuales diferentes intensidades del código PS3 siguiendo el mismo razonamiento propuesto por ClinGen-SVI para las supuestas variantes de pérdida de función (árbol de decisión PVS1) (Abou Tayoun et al., 2018b). Con este objetivo, primero identificamos los transcritos de *PALB2* biológicamente relevantes y las regiones codificantes de *PALB2* críticas para la función de las proteínas. Todos estos criterios de clasificación están descritos en profundidad en (Valenzuela-Palomo et al., 2021).

# **RESULTADOS**

## 1. Análisis de variantes detectadas en el proyecto BRIDGES.

### 1.1 Análisis Bioinformático.

Una gran variedad de herramientas informáticas permite determinar la probabilidad que tiene una determinada variante de DNA de afectar a los sitios canónicos de *splicing* y a otros elementos reguladores implicados en este proceso. En esta tesis doctoral, se ha realizado un análisis bioinformático con el programa Max Ent Scan (MES), siguiendo las directrices del proyecto BRIDGES de las variantes situadas en las fronteras intrón/exón, comparando las predicciones de los elementos en *cis* detectados en la secuencia consenso de referencia, con aquellas secuencias que incluyen las variantes que queremos analizar.

Las variantes candidatas se seleccionaron de acuerdo con estos criterios: a) rotura del sitio de *splicing* en las posiciones  $\pm 1,2$  (AG / GT); b) cambios importantes en la puntuación de MES ( $\geq 15\%$ ) (Houdayer et al., 2012; Moles-Fernández et al., 2018); c) creación de sitios de *splicing* alternativos *de novo* (corte  $\geq 3,0$ ); d) independientemente de las predicciones de MES, analizaremos todas aquellas variantes situadas en otras posiciones conservadas distintas de las secuencias consenso del aceptor y del donador, como cambios de pirimidina a purina, deleciones en el tracto de polipirimidina, sustituciones de nucleótidos conservados en las posiciones intrónicas -3C, + 3R, + 4A, + 5G, + 6T, así como el primero y los tres últimos nucleótidos del exón. Siguiendo este método, se han analizado 82 variantes de DNA del gen *PALB2* procedentes del proyecto BRIDGES con los programas citados anteriormente. Las Tablas 14, 15 y 16 muestran las variantes ordenadas por minigenes (mgPALB2\_ex1-3, mgPALB2\_ex4-6 y mgPALB2\_ex5-12; sección 1.2 de Materiales y Métodos) y exones.

**Tabla 14. Variantes de *PALB2* estudiadas *in silico* del proyecto BRIDGES en el minigén *mgPALB2\_ex1-3*.**

Variantes <sup>1</sup>	Exón/ Intrón	MES <sup>2</sup>					
		wt	mut	Cambio (%) <sup>2</sup>	<i>de novo</i> - wt	<i>de novo</i> -mut	Cambio (%)
<b>c.47A&gt;G</b>	Ex1	5.74	0.01	-99.8%			
<b>c.48G&gt;A</b>	Ex1	5.74	-3.48	-166.9%			
<b>c.48+1G&gt;A</b>	Ivs1	5.74	-2.43	-142.3%			
<b>c.48+2T&gt;C</b>	Ivs1	5.74	-2	-134.8%			
<b>c.48+3G&gt;A</b>	Ivs1	5.74	9.77	+68.5%			
<b>c.48+4C&gt;T</b>	Ivs1	5.74	2.88	-49.8%			
<b>c.48+7G&gt;C</b>	Ivs1	5.74	5.74	0	2.84	6.49	+328.5%
<b>c.49-2A&gt;T</b>	Ivs1	9.28	0.92	-90.1%	1.37	8.49	+719.7%
<b>c.50T&gt;G</b>	Ex2	9.28	9.02	-2.8			
<b>c.106C&gt;T</b>	Ex2	10.86	9.66	-11.1			
<b>c.108+1G&gt;A</b>	Ivs2	10.86	2.68	-75.3%			
<b>c.108+2T&gt;C</b>	Ivs2	10.86	3.1	-71.5%			
<b>c.108+7G&gt;A</b>	Ivs2	10.86	10.86	0			
<b>c.108+8A&gt;G</b>	Ivs2	10.86	10.86	0			
<b>c.108+9del</b>	Ivs2	10.86	10.86	0			
<b>c.108+9A&gt;G</b>	Ivs2	10.86	10.86	0			
<b>c.109-7C&gt;T</b>	Ivs2	10.06	10.13	+0.7%			
<b>c.109-7C&gt;A</b>	Ivs2	10.06	9.17	-8.9%		.	
<b>c.109-5T&gt;C</b>	Ivs2	10.06	9.86	-2.0%			
<b>c.109-6_109-4del</b>	Ivs2	10.06	7.27	-27.7%			
<b>c.109-2A&gt;G</b>	Ivs2	10.06	2.11	-79%	0.33	4.38	+1427.3%
<b>c.109C&gt;T</b>	Ex3	10.06	9.82	-2.4%			
<b>c.109C&gt;A</b>	Ex3	10.06	10.18	+1.2%			
<b>c.211+1G&gt;A</b>	Ivs3	8.76	0.58	-93.4%			
<b>c.211+5del</b>	Ivs3	8.76	6.99	-20.2%			

<sup>1</sup>El símbolo [+] implica ganancia, mientras que [-] pérdida.

<sup>2</sup>Las variantes seleccionadas se encuentran marcadas en **rojo** (13 variantes).

**Tabla 15. Variantes de *PALB2* estudiadas *in silico* del proyecto BRIDGES en el minigén *mgPALB2\_ex4-6*.**

Variantes <sup>1</sup>	Exón/ Intrón	MES					
		wt	mut	Cambio (%) <sup>2</sup>	<i>de novo</i> - wt	<i>de novo</i> - mut	Cambio (%)
<b>c.1682A&gt;G</b>	Ex4	8.88	8.39	-5.5%	-2.42	5.26	+317.4%
<b>c.1684+4A&gt;G</b>	Ivs4	8.88	7.24	-18.5%			
<b>c.1684+7C&gt;T</b>	Ivs4	8.88	8.88	0			
<b>c.1684+8A&gt;G</b>	Ivs4	8.88	8.88	0			
<b>c.1684+9A&gt;C</b>	Ivs4	8.88	8.88	0			
<b>c.1685-3T&gt;C</b>	Ivs4	11.15	10.96	-1.7			
<b>c.1685-2A&gt;C</b>	Ivs4	11.15	3.11	-72.1%	1.5	4.07	+366.0%
<b>c.1685-2A&gt;G</b>	Ivs4	11.15	3.2	-71.3%			
<b>c.2513A&gt;C</b>	Ex5	7.09	4.28	-39.6%			
<b>c.2514+5A&gt;G</b>	Ivs5	7.09	10.03	+41.5%			
<b>c.2514+6A&gt;G</b>	Ivs5	7.09	8.68	+22.4%			
<b>c.2514+8C&gt;T</b>	Ivs5	7.09	7.09	0			
<b>c.2515-9T&gt;C</b>	Ivs5	9.47	9.64	+1.8%			
<b>c.2515-9T&gt;G</b>	Ivs5	9.47	8.57	-9.5%			
<b>c.2515-2A&gt;G</b>	Ivs5	9.47	1.51	-84.1%			

<sup>1</sup>El símbolo [+] implica ganancia, mientras que [-] pérdida.

<sup>2</sup>Las variantes seleccionadas se encuentran marcadas en **rojo** (5 variantes).

**Tabla 16. Variantes de *PALB2* estudiadas *in silico* del proyecto BRIDGES en el minigén *mgPALB2\_ex5-12*.**

Variantes <sup>1</sup>	Exón/ Intrón	MES					
		wt	mut	Cambio (%) <sup>2</sup>	<i>de novo</i> - wt	<i>de novo</i> - mut	Cambio (%)
<b>c.2586+3C&gt;G</b>	Ivs6	6.80	8.68	+27.8%			
<b>c.2586+4A&gt;T</b>	Ivs6	6.80	4.30	-36.8%			
<b>c.2586+8G&gt;C</b>	Ivs6	6.80	6.80	0			
<b>c.2587-2A&gt;G</b>	Ivs6	3.31	-4.63	-239.9%	1.7	5.55	+426.47%
<b>c.2748+1G&gt;T</b>	Ivs7	11.08	2.57	-76.8%			
<b>c.2748+1G&gt;A</b>	Ivs7	11.08	2.90	-73.8%			
<b>c.2748+2dup</b>	Ivs7	11.08	1.20	-89.9%			
<b>c.2748+4A&gt;T</b>	Ivs7	11.08	9.81	-11.5%			
<b>c.2748+8G&gt;A</b>	Ivs7	11.08	11.08	0			
<b>c.2748+9del</b>	Ivs7	11.08	11.08	0			
<b>c.2749-1G&gt;T</b>	Ivs7	9.58	0.98	-89.8%	-0.01	7.77	+77800%
<b>c.2750T&gt;C</b>	Ex8	9.58	8.13	-15.1%			
<b>c.2834G&gt;C</b>	Ex8	9.80	7.26	-25.9%			
<b>c.2834+1G&gt;A</b>	Ivs8	9.80	1.62	-83.5%			
<b>c.2834+2T&gt;C</b>	Ivs8	9.80	2.05	-79.1%			
<b>c.2834+3A&gt;G</b>	Ivs8	9.80	6.99	-28.7%			
<b>c.2834+5G&gt;A</b>	Ivs8	9.80	7.88	-19.6%			
<b>c.2834+6T&gt;C</b>	Ivs8	9.80	9.37	-4.4%			
<b>c.2834+8A&gt;G</b>	Ivs8	9.80	9.80	0			
<b>c.2835-9T&gt;C</b>	Ivs8	8.14	8.67	+6.5%			
<b>c.2835-7G&gt;T</b>	Ivs8	8.14	8.56	+5.2%			
<b>c.2835-3T&gt;C</b>	Ivs8	8.14	7.20	-11.6%			
<b>c.2836G&gt;A</b>	Ex9	8.14	7.99	-1.8%			
<b>c.2996+4A&gt;G</b>	Ivs9	9.21	6.45	-30.0%			
<b>c.2997-2del</b>	Ivs9	6.69	-5.61	-183.9%			
<b>c.2997-1G&gt;A</b>	Ivs9	6.69	-2.05	-130.6%	-7.67	4.22	155.0%
<b>c.3113G&gt;A</b>	Ex10	8.72	3.93	-54.9%			
<b>c.3113+3A&gt;G</b>	Ivs10	8.72	7.43	-14.8%			
<b>c.3113+5G&gt;C</b>	Ivs10	8.72	3.53	-59.5%			
<b>c.3113+8T&gt;C</b>	Ivs10	8.72	8.72	0			
<b>c.3114-5C&gt;T</b>	Ivs10	4.13	6.39	+54.7%			
<b>c.3201+1G&gt;A</b>	Ivs11	11.01	2.83	-74.3%			
<b>c.3201+3-3201+4insTG</b>	Ivs11	11.01	6.56	-40.4%			
<b>c.3201+6T&gt;A</b>	Ivs11	11.01	9.48	-13.9%			
<b>c.3202-9C&gt;T</b>	Ivs11	6.42	6.81	+6.1%			
<b>c.3202-8G&gt;T</b>	Ivs11	6.42	7.48	+16.5%			
<b>c.3202-8G&gt;A</b>	Ivs11	6.42	6.92	+7.8%			
<b>c.3203G&gt;C</b>	Ex12	6.42	5.94	-7.5%			

<b>c.3348C&gt;T</b>	Ex12	3.10	1.91	-67.4%	-0.75	6.99	+1032%
<b>c.3350+3A&gt;G</b>	Ivs12	3.10	2.92	-5.8%			
<b>c.3350+4A&gt;G</b>	Ivs12	3.10	2.53	-18.4%	1.67	6.46	+486.8%
<b>c.3350+5G&gt;A</b>	Ivs12	3.10	1.01	-38.4%	-0.75	6.99	+1032%

<sup>1</sup>El símbolo [+] implica ganancia, mientras que [-] pérdida.

<sup>2</sup>Las variantes seleccionadas se encuentran marcadas en **rojo** (26 variantes).

En total, se seleccionaron 44 variantes (en rojo en las Tablas 14, 15 y 16) para estudios funcionales, debido a que podían estar asociadas a un potencial efecto deletéreo sobre el *splicing*. Un total de 11 y 33 variantes alteraron los sitios 3'ss y 5'ss, respectivamente. MES predijo que siete variantes de sitios de *splicing* (c.49-2A>T, c.109-2A>G, c.2587-2A>G, c.2749-1G>T, c.2997-1G>A, c.3348C>T, c.3350+4A>G) generarían simultáneamente un sitio *de novo*. Las variantes c.2748+4A>T, c.2834+6T>C, c.3113+3A>G y c.3201+6T>A fueron seleccionadas, a pesar del pequeño cambio en el valor de MES, porque afectaban posiciones conservadas del 5'ss (secuencia consenso Exón/Intrón: MAG/gtragt). La prevalencia de cada una de las variantes BRIDGES en casos y controles se indica en la tabla S1 del anexo.

## 1.2 Construcción de los minigenes.

Los minigenes híbridos son construcciones artificiales, fáciles de manipular, que recrean genes humanos implicados en una enfermedad, y permiten estudiar el proceso de *splicing* y el efecto de variantes sobre este proceso, sin necesidad de muestras de pacientes. Se han construido 3 minigenes en el vector de *splicing* pMAD (ver sección 2.2 de Materiales y Métodos). El primero, **mgPALB2\_ex1-3** incluye una región genómica que contiene los exones 1, 2, 3, el intrón 2 completo y parte de los intrones 1 y 3 del gen *PALB2* según la siguiente estructura: **ex1** (78 bp) – ivs1-1 (216 bp) // ivs1-2 (200 bp) - **ex2** (60 bp) – ivs2 (117 bp) – **ex3** (103 bp) – ivs3-1 (200 bp).

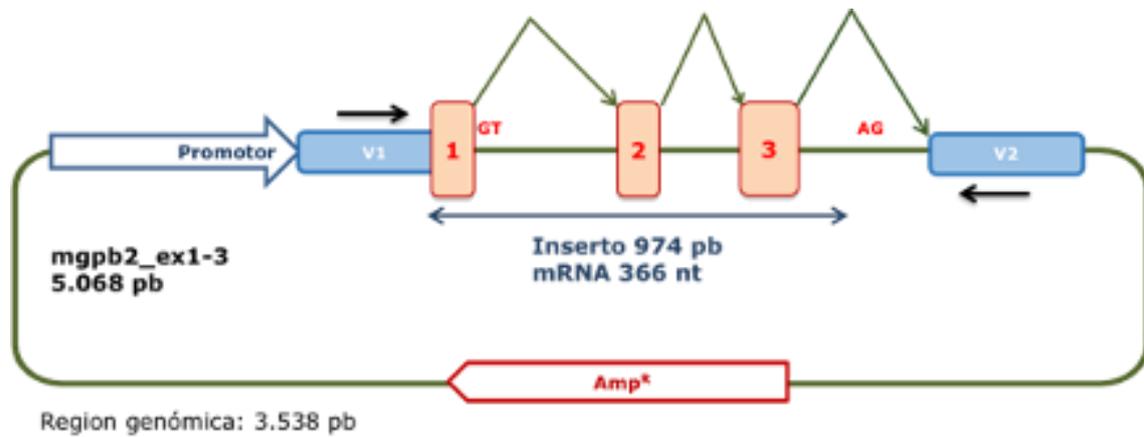
El segundo minigén, **mgPALB2\_ex4-6**, contiene los exones 4, 5 y 6, el intrón 5 completo y parte de los intrones 3, 4, y 6, con la estructura: ivs3 (416 bp) – **ex4** (1473 bp) – ivs4-1 (734 bp) // ivs4-2 (266 bp) – **ex5** (830 bp) – ivs5 (364 bp) – **ex6** (72 bp) – ivs6-1 (209 bp).

Por último, el minigén **mgPALB2\_ex5-12**, alberga los exones 5, 6, 7, 8, 9, 10,11 y 12, el intrón 5 completo y parte de los intrones 4, 6, 7, 8, 9, 10, 11 y 12 con la siguiente estructura: ivs4-2 (266 bp) – **ex5** (830 bp) – ivs5 (364 bp) – **ex6** (72 bp) – ivs6-(209 bp) -ivs6-2 (191 bp) – **ex7** (162 bp) – ivs7-1 (114 bp) – *SalI* – ivs7-2 (250 bp) – **ex8** (86 bp) – ivs8-1 (200 bp) // ivs8-2 (200 bp) – **ex9** (162 bp) – ivs9-1 (200 bp) //ivs9-2 (200 bp) – **ex10** (117 bp) – ivs10-1 (200 bp) // ivs10-2 (200 bp) – **ex11** (88 bp) – ivs11-1 (200 bp) // ivs11-2 (200 bp) – **ex12** (149 bp) – ivs12 (250 bp).

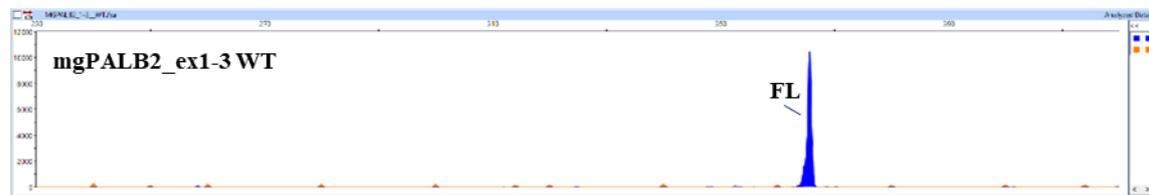
Estos tres minigenes, contienen insertos cuya suma total es de 10.292 pares de bases de *PALB2* cubriendo una región genómica de 33.574 pb. El minigén mgPALB2\_ex1-3 contiene una fusión del exón V1 del vector de *splicing* y el exón 1 de *PALB2*, lo que permite el estudio de variantes en el 5' ss del exón 1 de *PALB2*.

El minigén mgPALB2\_ex1-3 tiene un tamaño de 5068 pares de bases y produce un transcrito canónico (también llamado *wild type* o *full-length* en esta memoria) estable de 366 nucleótidos, como podemos ver en la Figura 20. Por lo tanto, se considera este minigén apto para el estudio funcional de variantes de *splicing*.

A

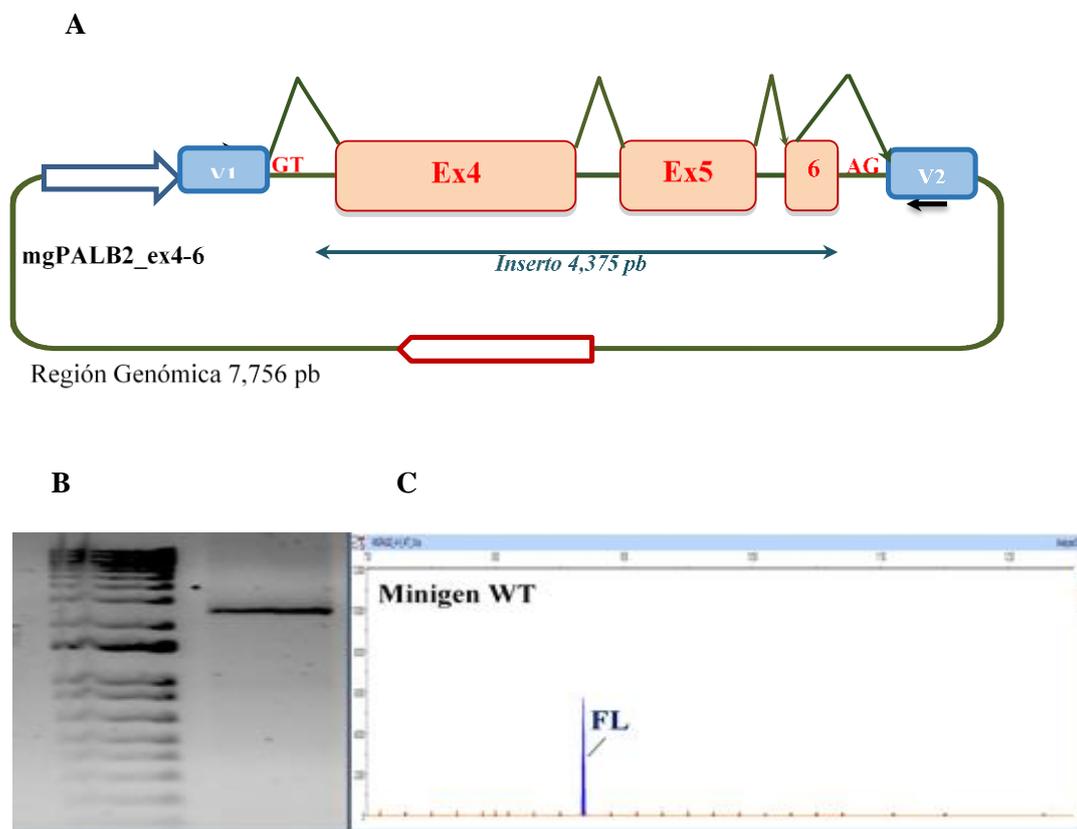


B



**Figura 20.** El minigén mgPALB2\_ex1-3. A) Estructura del minigén de los exones 1 a 3 de *PALB2*. B) Electroforesis capilar del transcrito canónico del minigén *wild type* mgPALB2\_ex1-3 con los oligonucleótidos: RT-PSPL3-Fw y RT-pSAD-rV-FAM (366 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ600).

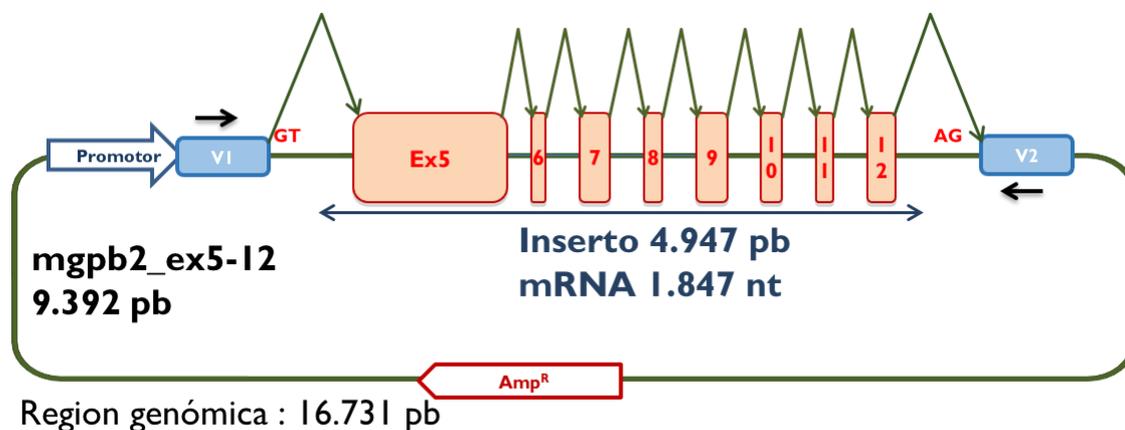
El minigén **mgPALB2\_ex4-6** tiene un tamaño de 7756 pares de bases y produce un transcrito estable de 2556 nucleótidos (Figura 21). Por lo tanto, consideramos este minigén apto para el estudio funcional de variantes de *splicing*.



**Figura 21.** El minigén **mgPALB2\_ex4-6**. **A)** Estructura del minigén de los exones 4 a 6 de *PALB2*. **B)** Gel de agarosa 1% que muestra el transcrito V1-ex4-ex5-ex6-V2 de 2556 nt. **C)** Electroforesis capilar fluorescente del transcrito canónico del minigén *wild type* **mgPALB2\_ex4-6** con los oligonucleótidos: RT-PALB2-ex4-fw y FAM-RT-PALB2-ex6-rv-FAM (918 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

El minigén **mgPALB2\_ex5-12** tiene un tamaño de 9392 pb y produce un transcrito estable de 1.847 nucleótidos, como podemos ver en la Figura 22. Por lo tanto, consideramos este minigén apto para el estudio funcional de variantes de *splicing*.

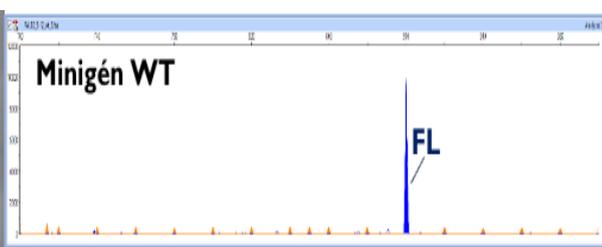
A



B



C



**Figura 22.** El minigén **mgPALB2\_ex5-12**. A) Estructura del minigén de los exones 5 a 12 de *PALB2*. B) Gel de agarosa al 1% que muestra el transcrito V1-ex5-ex6-ex7-ex8-ex9-ex10-ex11-ex12-V2 con un tamaño de 1847 nt. C) Electroforesis capilar del minigén **mgPALB2\_ex5-12** con los oligonucleótidos: RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM, con un tamaño de 906 nt. Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

### 1.3 Ensayos funcionales.

Se transfectaron aproximadamente  $2 \times 10^5$  células MCF-7 (células de adenocarcinoma de mama humano), con 1  $\mu$ g del minigén (apartado 2.4 de Materiales y Métodos). Además, para comprobar la reproducibilidad de los resultados de *splicing*, también se transfectaron células MDA-MB-231 (línea celular de cáncer de mama triple negativo) con los minigenes WT, **mgPALB2\_ex1-**

3 y mgPALB2\_ex5-12 y las siguientes variantes: c.47A>G, c.48+1G>A, c.48+2T>C, c.48+4C>T, c.2749-1G>T, c.2834+3A>G, c.3201+3\_3201+4insTG y c.3350+4A>G.

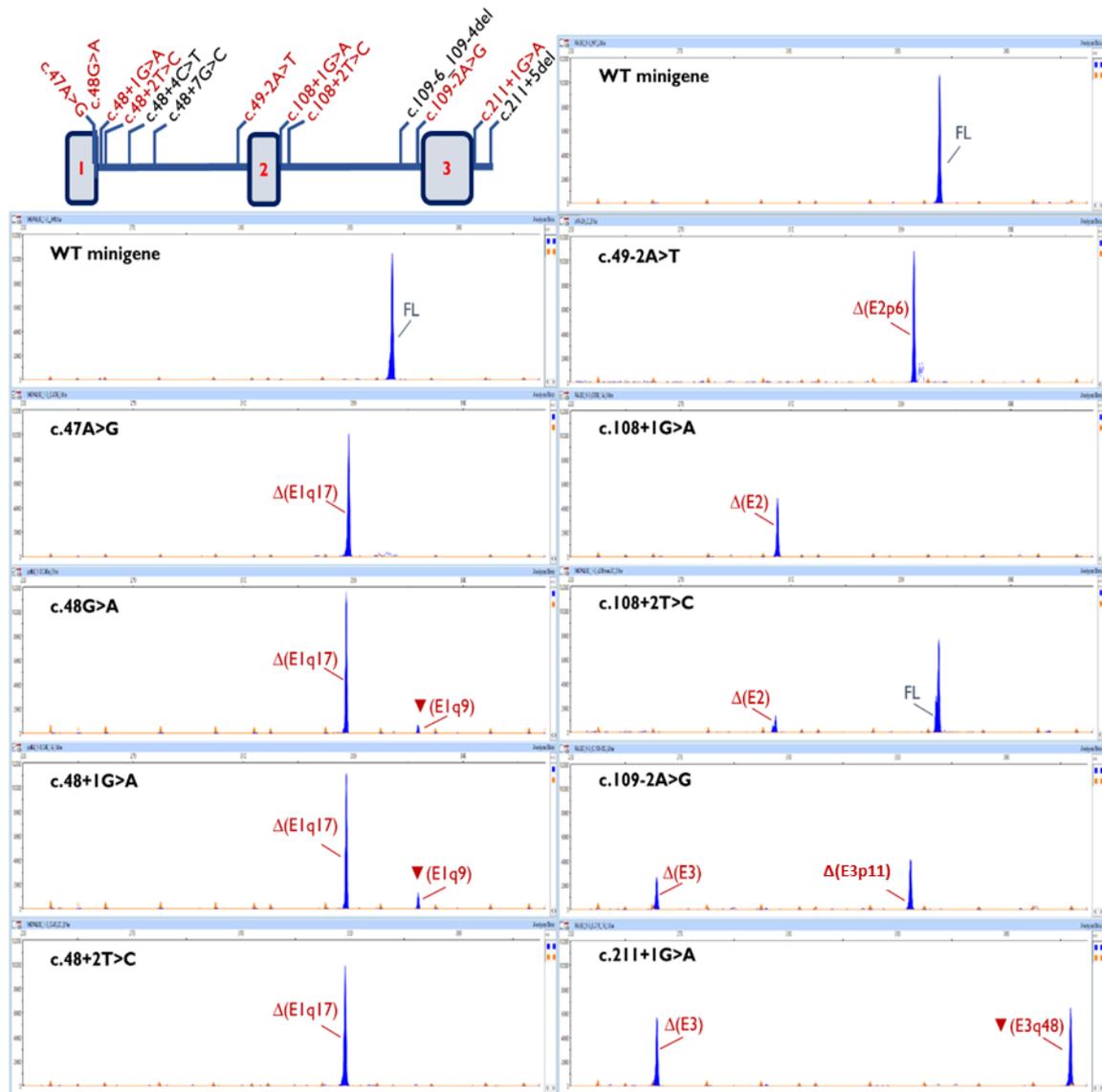
### 1.3.1 Resultados del minigén mgPALB2\_ex1-3.

De las 25 variantes procedentes del proyecto BRIDGES, situadas en los exones 1, 2 y 3 de *PALB2* y secuencias intrónicas adyacentes, 13 fueron seleccionadas para su estudio funcional siguiendo los criterios bioinformáticos mostrados en el apartado 1.2 de Materiales y Métodos (Tabla 14). De las 13 variantes ensayadas, 6 afectan al sitio donador del exón 1, (**c.47A>G**, **c.48G>A**, **c.48+1G>A**, **c.48+2T>C**, **c.48+4C>T** y **c.48+7G>C**), 3 a los sitios de *splicing* del exón 2 (**c.49-2A>T**, **c.108+1G>A** y **c.108+2T>C**) y 4 a los del exón 3 (**c.109-6\_109-4del**, **c.109-2A>G**, **c.211+1G>A** y **c.211+5del**).

Mediante mutagénesis dirigida se han generado los minigenes mutantes a partir de mgPALB2\_ex1-3 *wild type* (Figura 20). Las células MCF-7 fueron transfectadas con los minigenes mutantes, y los productos de la amplificación con los oligonucleótidos (RTPSPL3-FW y RT-PSAD-RV-FAM), tras una RT-PCR del RNA procedente del proceso de transfección, se comprobaron por diferentes técnicas como: electroforesis en gel de agarosa, electroforesis capilar y secuenciación capilar Sanger. A continuación, se muestran los resultados obtenidos.

#### 1.3.1.1 Resultados de las variantes del exón 1 de *PALB2*.

La variante *missense* **c.47A>G** p.(Lys16Arg) está registrada y catalogada en ClinVar como variante de significado clínico incierto (VUS). Según el estudio bioinformático, esta variante elimina el sitio donador del exón 1 (Tabla 14). Los ensayos funcionales desarrollados con esta variante revelaron la presencia única de la isoforma aberrante ( $\Delta$ (E1q17), 100%) (Figura 23; Tablas 17-18).



**Figura 23. Ensayos funcionales de *splicing* de variantes del minigén mgPALB2\_ex1-3.** Mapa de variantes ensayadas y electroforesis capilar de fluorescencia de productos RT-PCR de los minigén *wild type* mgPALB2\_ex1-3 y mutantes. Se amplificaron con los oligonucleótidos FAM-RT-PSPL3-Fw y RT-PSAD-rv (Transcripción completa VI-ex1-ex2-ex3-V2 = 366 nt) Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ600).

La variante sinónima **c.48G>A** p.(Lys16=), y la variante de sitios de *splicing* **c.48+1G>A**, están registradas y catalogadas en ClinVar como patogénica y posiblemente patogénica, respectivamente. De acuerdo a las predicciones bioinformáticas, ambas variantes causan la eliminación del sitio donador del exón 1 (Tabla 14). Sin embargo, los ensayos funcionales revelaron

la producción de  $\Delta(E1q17)$  en un 88,6% y 77%, respectivamente. Asimismo, otras isoformas, aunque de forma minoritaria, también se vieron afectadas por el *splicing*. **c.48G>A** generaba  $\nabla(I1^{ms})$  en un 5% y  $\nabla(E1q9)$  en un 5,4%, además de porcentajes mínimos del transcrito *full-length* (Figura 23; Tablas 17-18). Los mismos transcritos fueron observados en los resultados producidos por **c.48+1G>A**, pero en diferentes porcentajes; un 13,8%, y un 9,2%, respectivamente (Figura 23; Tablas 17-18).

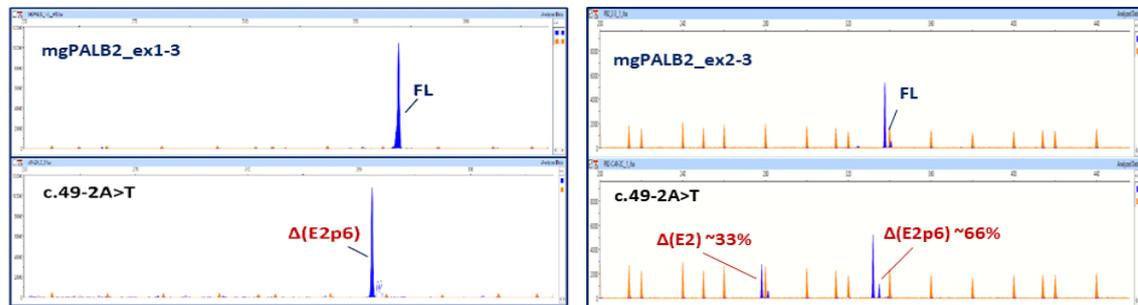
La variante **c.48+2T>C**, está catalogada en ClinVar como posiblemente patogénica. Los estudios bioinformáticos predicen que esta variante elimina el sitio donador del exón 1 (Tabla 14). Esta variante, convierte un donador GT en un donador atípico GC, que no es reconocido por la maquinaria de *splicing*, ya que los resultados del ensayo funcional muestran la presencia única del transcrito aberrante ( $\Delta(E1q17)$ , 100%) (Figura 23; Tablas 17-18).

Las dos últimas variantes ensayadas en este exón, fueron **c.48+4C>T** y **c.48+7G>C**. Ambas variantes se encuentran registradas en ClinVar y catalogadas como VUS y como benigna o posiblemente benigna, respectivamente. Según el estudio bioinformático, **c.48+4C>T**, produce la eliminación del aceptor del exón 1 y el estudio de **c.48+7G>C** predecía la creación de un donador situado 9 nucleótidos *downstream* del natural (Tabla 14). Sin embargo, los ensayos funcionales concluyeron que ninguna de las variantes presentaba transcritos aberrantes, en niveles significativos (Figura 25).

### **1.3.1.2 Resultados de las variantes del exón 2 de PALB2.**

Tres variantes del exón 2 fueron ensayadas funcionalmente en el minigén mgPALB2\_ex1-3. En primer lugar, la variante de sitio de *splicing* **c.49-2A>T**, está registrada e interpretada como benigna y como posiblemente patogénica, creando conflictos de interpretación, ya que no existen evidencias funcionales en ClinVar acerca de esta variante. El análisis bioinformático, predice la eliminación del aceptor natural del exón 2 y crea un sitio críptico *de novo* 6 nucleótidos *downstream* (Tabla 14). Los resultados del ensayo funcional revelaron como único transcrito aberrante la isoforma ( $\Delta(E2p6)$ , 100%). La variante **c.49-2A>T** fue ensayada también en el minigén

mgPALB2\_ex2-3, mostrando diferentes resultados. Así, en mgPALB2\_ex2-3 produjo  $\Delta(E2)$  en un 33,3% y  $\Delta(E2p6)$  en un 66,6% (Figuras 23-24; Tablas 17-18). Estos resultados, demuestran la importancia de mantener el contexto genómico en el desarrollo de minigenes para conseguir la reproducibilidad de las reacciones de *splicing*.



**Figura 24. Electroforesis capilar de c.49-2A>T en los minigenes mgPALB2\_ex1-3 y mgPALB2\_ex2-3.** Se amplificaron con los oligonucleótidos FAM-RT-PSPL3-Fw y RT-PSAD-rv. Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ600).

**Tabla 17. Eventos de *splicing* producidos por variantes encontradas en pacientes BRIDGES del gen *PALB2*.**

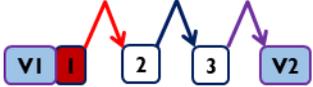
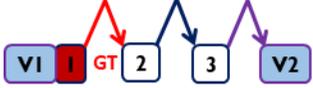
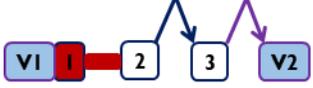
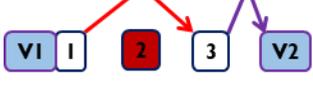
Variantes (HGVS) <sup>1</sup>	Resumen Bioinformático <sup>2</sup>	Transcritos <i>Full-length</i>	Transcritos-PTC <sup>3</sup>	Transcritos <i>In-Frame</i> <sup>3</sup>
<i>Wild type</i> mgPALB2_ex1-3		100%		
<b>c.47A&gt;G</b> p.(Lys16Arg)	[-]5'ss (5.74→0.01)	-	Δ(E1q17) [100%]	
<b>c.48G&gt;A</b> p.(Lys16=)	[↓]5'ss (5.74→3.48)	0.9%±0.1%	Δ(E1q17) [88.6%±0.1%] ▼(I1 <sup>ms</sup> ) [5%]	▼(E1q9) [5.4%±0.1%]
<b>c.48+1G&gt;A</b>	[-]5'ss (5.74→-2.43)	-	Δ(E1q17) [77%] ▼(I1 <sup>ms</sup> ) [13.8%]	▼(E1q9) [9.2%]
<b>c.48+2T&gt;C</b>	[-]5'ss (5.74→-2)	-	Δ(E1q17) [100%]	
c.48+4C>T	[-]5'ss (5.74→2.88)	94.5%±0.5%	Δ(E1q17) [5.5%±0.5%]	
c.48+7G>C	5'ss (5.74→5.74) [+]5'ss (6.49) 9-nt <i>downstream</i>	100%		
<b>c.49-2A&gt;T</b>	[-]3'ss (9.28→0.92) [+]3'ss (8.49) 6-nt <i>downstream</i>	-		Δ(E2p6) [100%]
<b>c.108+1G&gt;A</b>	[-]5'ss(10.86→2.68)	-		Δ(E2) [100%]
c.108+2T>C	[-]5'ss(10.86→3.1)	85.5%±0.3%		Δ(E2) [14.5%±0.3%]
c.109-6_109-4del	[↓]3'ss(10.06→7.27)	100%		
<b>c.109-2A&gt;G</b>	[-]3'ss (10.06→2.11) [+]3'ss (4.38) 11-nt <i>downstream</i>	-	Δ(E3) [41% ±4.0%] Δ(E3p11) [59% ±4.0%]	
<b>c.211+1G&gt;A</b>	[-]5'ss (8.76→0.58)	-	Δ(E3) [45.7%±0.4%]	▼(E3q48) [54.3%±0.4%]
c.211+5del	[↓]5'ss (8.76→6.99)	100%		
<i>Wild type</i> mgPALB2_ex4-6		100%		
c.1684+4A>G	[↓]5'ss (8.88→7.24)	41.7%±1.2 %		Δ(E4) [58.3%±1.2%]
<b>c.1685-2A&gt;C</b>	[↓]3'ss (11.15→3.11)	-	Δ(E5p139) [67%±3.6%] Δ(E5p5) [15.4%±1.5%] ▼(E5p88) [6.4%±1.3%] Δ(E5p10) [5.7%±0.7%] Δ(E5p97) [5.5%±0.2%]	

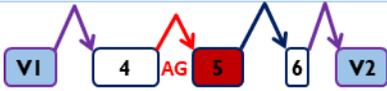
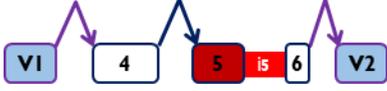
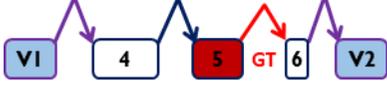
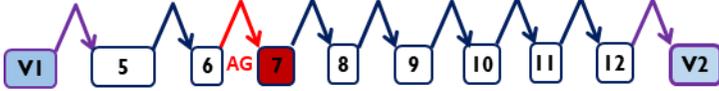
<b>c.1685-2A&gt;G</b>	[↓]3'ss (11.15→3.2) [+]3'ss (4.07) 139-nt <i>downstream</i>	-	Δ(E5p139) [73.3%±0.7%] ▼(E5p88) [8.7%±0.1%] Δ(E5p10) [7.3%±0.7%] Δ(E5p97) [6.7%±0.1%] Δ(E5p5) [4%±0.3%]	
c.2513A>C p.(Gln838Pro)	[↓]5'ss (7.09→4.28)	73.1± 2%	▼(I5) [21.2%± 1.8%] ▼(E5q106)[5.7%± 0.3%]	
<b>c.2515-2A&gt;G</b>	[↓]3'ss (9.47→8.57)	-	▼(I5) [100%]	
<i>Wild type</i> mgPALB2_ex5-12		100%		
c.2586+4A>T	[↓]5'ss (6.8→4.3)	100%		
<b>c.2587-2A&gt;G</b>	[-]3'ss (3.31→2.66) [+]3'ss (5.5) 10-nt <i>downstream</i>	-	Δ(E7p10) [100%]	
<b>c.2748+1G&gt;A</b>	[-]5'ss(11.08→2.90)	.		Δ(E7) [100%]
<b>c.2748+1G&gt;T</b>	[-]5'ss(11.08→2.57)	-		Δ(E7) [100%]
<b>c.2748+2dup<sup>4</sup></b>	[-]5'ss (11.08→1.2)	-		Δ(E7) [83,7%]
c.2748+4A>T	[↓]5'ss(11.08→9.81)	100%		
<b>c.2749-1G&gt;T</b>	[-]3'ss (9.58→0.98) [+]3'ss (7.77) 7-nt <i>downstream</i>	-	Δ(E8p7) [96.9%± 0.26%]	Δ(E8p15) [3.1%± 0.26%]
c.2750T>C	[↓]3'ss (9.58→8.13)	100%		
c.2834G>C p.(Arg945Thr)	[↓]5'ss (9.8→7.26)	90%± 0.4%	Δ(E8) [7.8%± 0.3%] ▼(E8q69) [2.2%±0.1%]	
<b>c.2834+1G&gt;A</b>	[-]5'ss (9.8→1.62)	-	Δ(E8) [82.1%±0.5%] ▼(E8q69) [17.9% ±0.5%]	
<b>c.2834+2T&gt;C</b>	[-]5'ss (9.8→2.05)	-	Δ(E8) [100%]	
c.2834+3A>G	[↓]5'ss (9.8→6.99)	80.2% ±0.3%	Δ(E8) [19.8%±0.3%]	
c.2834+5G>A	[↓]5'ss (9.8→7.88)	40.7% ±0.9%	Δ(E8) [59.3% ±0.9%]	
c.2834+6T>C	[↓]5'ss (9.8→9.37)	100%		
c.2996+4A>G	[↓]5'ss (9.21→6.45)	12.9%±1.1%		Δ(E9) [87.1%±1.1%]
<b>c.2997-2del</b>	[↓]3'ss (6.69→5.61)	-	Δ(E10p2) [66.6% ±0.2%]	Δ(E10) [33.3% ± 0.2%]
<b>c.2997-1G&gt;A<sup>4</sup></b>	[-]3'ss (6.69→-2.05)	-	Δ(E10p2) [95.8±0.9%]	

	[+]3'ss (4.22) 2-nt <i>downstream</i>			
c.3113G>A	[↓]5'ss (8.72→3.93)	10.5%±0.91%	Δ(E10q31) [52.9%±4.3%] Δ(E10q41) [2.2%±0.2%]	Δ(E10) [28.1%±2.5%] Δ(E9_10) [6.3%±0.6%]
c.3113+3A>G	[↓]5'ss (8.72→7.43)	8%±0.95%	Δ(E10q31) [62.7%±0.1%] Δ(E10q65) [4.9%±0.3%]	Δ(E10) [19.9%±0.34%] Δ(E9_10) [4.5%±0.3%]
<b>c.3113+5G&gt;C</b>	[↓]5'ss (8.72→3.53)	4.9%±0.7%	Δ(E10q31)[90%±0.48%] Δ(E10q41)[2.8%±0.25%] Δ(E10q65) [2.3%±0.12%]	
<b>c.3201+1G&gt;A</b>	[-]5'ss(11.01→2.83)	-	Δ(E11) [100%]	
<b>c.3201+3_3201+4insTG</b>	[↓]5'ss(11.01→6.56)	-	Δ(E11) [100%]	
c.3201+6T>A	[↓]5'ss(11.01→9.48)	11.5% ± 0.3%	Δ(E11) [88.5%±0.3%]	
c.3348C>T p.(Gln838Pro)	[-]5'ss (3.1→1.91) [+]5'ss (6.99) 4-nt <i>upstream</i>	68%±0.36%	Δ(E12q4) [26.7%±0.75%] Δ(E12) [5.3% ±0.31%]	
<b>c.3350+4A&gt;G</b>	[-]5'ss (3.1→2.53) [+]5'ss (6.46) 4-nt <i>downstream</i>	-	▼(E12q4) [56.1% ±7.8%] Δ(E12) [43.9% ±7.8%]	
<b>c.3350+5G&gt;A</b>	[-]5'ss (3.1→1.01)	-	Δ(E12) [100%]	

<sup>1</sup> Fuente en negrita: sin rastros o <5% del transcrito *full-length*. <sup>2</sup> [-] eliminación del sitio de *splicing*; [+] Nuevo sitio. [↓] Reducción de la puntuación MES. 3 PTC: codón de terminación prematura; <sup>3</sup> Δ, pérdida de secuencias exónicas; ▼ inclusión de secuencias intrónicas; E (exón), p (cambio de aceptor), q (cambio de donador). Cuando es necesario, se indica el número exacto de nt insertados o eliminados. Por ejemplo, la transcripción ▼ (E1q9) denota el uso de un sitio donador alternativo que se localiza 9 nucleótidos *downstream* del exón 1, provocando la adición de 9-nt al mRNA maduro. <sup>4</sup> Se encontraron 3 transcritos no caracterizados: 888-nt [4.2%±0.9%] (c.2997-1G>A) y, 592-nt [8.9%±0.3%] y 710-nt [7.4%±0.2%] (ambos de c.2748+2dup).

**Tabla 18. Anotación de transcritos en el mgPALB2\_ex1-3, mgPALB2\_ex4-6 y mgPALB2\_ex5-12 de acuerdo a la secuencia de PALB2 NM\_024675.4.**

Transcritos <sup>1</sup>	RNA-HGVS	PROTEIN-HGVS	Diagrama de eventos de <i>splicing</i> <sup>2</sup>
<u>Δ(E1q17)</u>	r.32_48del	p.(Cys11Phefs*26)	 Δ(E1q17)
<u>▼(E1q9)</u>	r.48_49ins48+1_48+9	p.(Lys16_Leu17insVPG)	 ▼(E1q9)
<u>▼(I1<sup>mg</sup>)</u>	r.48_49ins48+1_48+416	p.(Leu17Valfs*20)	 ▼(I1 <sup>mg</sup> )
<u>Δ(E2p6)</u>	r.49_54del	p.(Leu17_Lys18del)	 Δ(E2p6)
<u>Δ(E2)</u>	r.49_108del	p.(Leu17_Gln36del)	 Δ(E2)
<u>Δ(E3p11)</u>	r.109_117del	p.(Arg37Serfs*2)	 Δ(E3p11)
<u>▼(E3q48)</u>	r.211_212ins211+1_211+48	p.(Ser70_Glu71ins16)	 ▼(E3q48)
<u>Δ(E3)</u>	r.109_211del	p.(Arg37Asnfs*106)	 Δ(E3)
<u>Δ(E4)</u>	r.212_1684del	p.(Glu71_Lys561del)	 Δ(E4)

$\Delta$ (E5p5)	r.1685_1689del	p.(Gly562Glufs*14)	 $\Delta$ (E5p5)
<u><math>\Delta</math>(E5p139)</u>	r.1685_1823del	p.(Gly562Valfs*20)	 $\Delta$ (E5p139)
$\nabla$ (E5p88)	r.1685_1686ins1685+1_1685+88	p.(Lys563Aspfs*44)	 $\nabla$ (E5p88)
$\Delta$ (E5p10)	r.1685_1694del	p.(Gly562Valfs*34)	 $\Delta$ (E5p10)
$\Delta$ (E5p97)	r.1685_1781del	p.(Lys563Metfs*4)	 $\Delta$ (E5p97)
$\nabla$ (I5)	r.2514_2515ins2514+1_2514+363	p.(Thr839Valfs*14)	 $\nabla$ (I5)
$\nabla$ (E5q106)	r.2514_2515ins2514+1_2514+106	p.(Thr839Valfs*14)	 $\nabla$ (E5q106)
<u><math>\Delta</math>(E7p10)</u>	r.2587_2596del	p.(Asn863Valfs*5)	 $\Delta$ (E7p10)
<u><math>\Delta</math>(E7)</u>	r.2587_2748del	p.(Asn863_Glu916del)	 $\Delta$ (E7)
$\Delta$ (E8p7)	r.2749_2755del	p.(Val917Tyrfs*4)	 $\Delta$ (E8p7)
$\Delta$ (E8p15)	r.2749_2763del	p.(Val917_Gln921del)	 $\Delta$ (E8p15)

<u>Δ(E8)</u>	r.2749_2834del	p.(Val917Glyfs*7)	
▼(E8q69)	r.2834_2835ins2834+1_2834+69	p.(Ala946Tyrfs*7)	
<u>Δ(E9)</u>	r.2835_2996del	p.(Ala946_Gly999del)	
<u>Δ(E10p2)</u>	r.2997_2998del	p.(Gly1000_Glnfs*10)	
<u>Δ(E10)</u>	r.2997_3113del	p.(Gly1000_Trp1038del)	
<u>Δ(E10q31)</u>	r.3083_3113del	p.(Thr1029Ilefs*2)	
<u>Δ(E10q41)</u>	r.3073_3113del	p.(Ala1025Glufs*14)	
<u>Δ(E10q65)</u>	r.3049_3113del	p.(Ala1017Glufs*14)	
<u>Δ(E9_10)</u>	r.2835_3113del	p.(Ala946_Trp1038del)	

<u><math>\Delta</math>(E11)</u>	r.3114_3201del	p.(Asn1039Glyfs*7)	
<u><math>\Delta</math>(E12)</u>	r.3202_3350del	p.(Gly1068Valfs*5)	
<u><math>\Delta</math>(E12q4)</u>	r.3347_3350del	p.(Arg1117Serfs*6)	
<u><math>\nabla</math>(E12q4)</u>	r.3350_3351ins3350-4_3351-1	p.(Phe1118Glnfs*6)	

<sup>1</sup>Los eventos de *splicing* alternativo natural están subrayados.

<sup>2</sup>Los eventos de *splicing* aberrante están marcados en rojo. Los nuevos donadores y aceptores alternativos están marcados también en rojo.

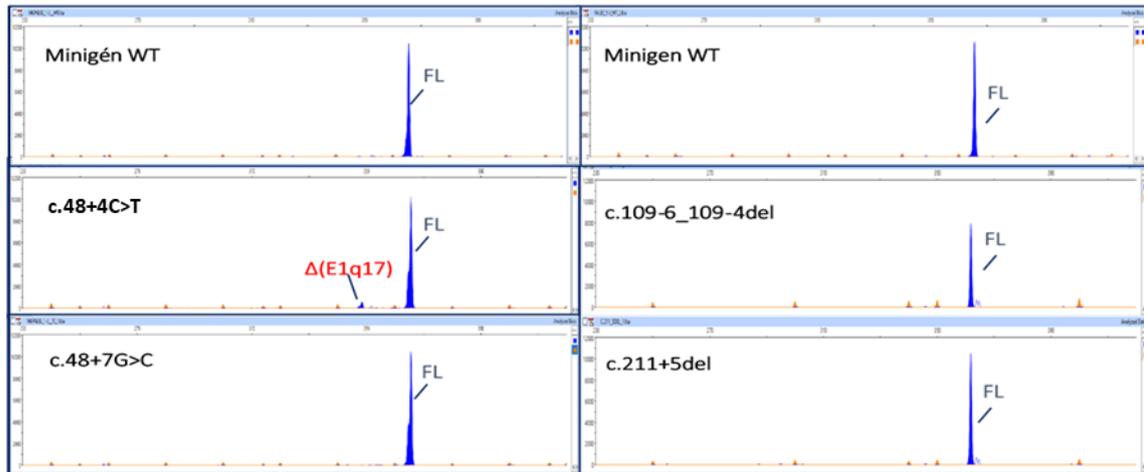
Las variantes de sitio de *splicing* **c.108+1G>A** y **c.108+2T>C**, son variantes registradas en la base de datos ClinVar y catalogadas como posiblemente patogénicas. Según la bioinformática, estas 2 variantes, eliminan el sitio donador del exón 2 (Tabla 14). Los resultados de los ensayos funcionales con **c.108+1G>A** revelaron la presencia de un único transcrito aberrante ( $\Delta(E2)$ , 100%) (Figura 23; Tablas 17-18). Sin embargo, **c.108+2T>C** generó el transcrito *full-length* en un 88,5%, por el uso de un donador atípico GC y  $\Delta(E2)$  en un 14,5% (Figura 23; Tablas 17-18).

### 1.3.1.3 Resultados de las variantes del exón 3 de *PALB2*.

Un total de 4 variantes BRIDGES fueron ensayadas funcionalmente en el minigén *mgPALB2\_ex1-3*. La variante **c.109-2A>G** está registrada y catalogada en ClinVar como posiblemente patogénica. El estudio bioinformático pronostica la eliminación del aceptor del exón 3 y creación de un nuevo sitio de *splicing* muy fuerte 11 nucleótidos *downstream* (Tabla 14). Los datos obtenidos en los ensayos funcionales mostraron la producción de  $\Delta(E3)$ , en un 41% y de  $\Delta(E3p11)$  en un 59% (Figura 23; Tablas 17-18).

Por otro lado, **c.211+1G>A**, es una variante registrada en la base de datos ClinVar y catalogada como posiblemente patogénica, que afecta al donador del exón 3 (Tabla 14). Los resultados del análisis bioinformático determinan que la presencia de esta variante elimina el donador natural del exón 3. Los resultados observados en los ensayos de *splicing* determinaron la presencia de transcritos aberrantes  $\Delta(E3)$ , en un 45,7% y  $\nabla(E3q48)$  en un 54,3%, a causa del uso de un donador alternativo situado 48 nucleótidos *downstream* del natural (Figura 23; Tablas 17-18).

No todas las variantes ensayadas en este exón tuvieron efectos sobre el proceso de *splicing*. El estudio bioinformático de las variantes de significado clínico incierto **c.109-6\_109-4del** y de **c.211+5del**, sugería un debilitamiento de los sitios aceptor y donador del exón 3, respectivamente (Tabla 14). Sin embargo, los ensayos funcionales revelaron la presencia única del transcrito canónico en ambos casos (Figura 25).



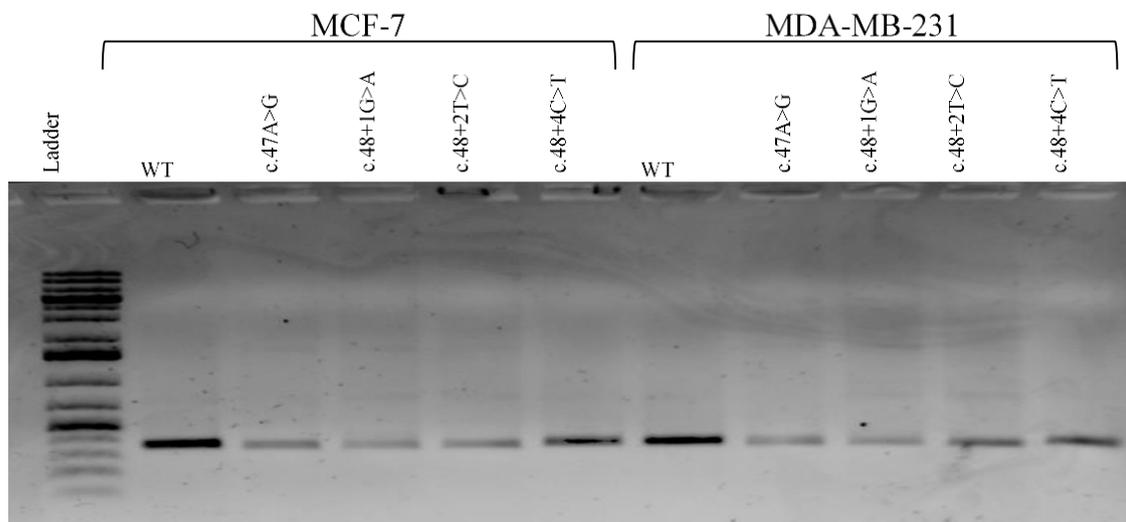
**Figura 25. Electroforesis capilar de c.48+4C>T y c.48+7G>C, c.109-6\_109-4del y c.211+5del en el minigén mgPALB2\_ex1-3.** Se amplificaron con los oligonucleótidos FAM-RT-PSPL3-Fw y RT-PSAD-rv (Transcripción completa VI-ex1-ex2-ex3-V2 = 366 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ600).

### 1.3.1.4 Recapitulación de resultados de los exones 1 a 3.

Trece variantes del gen *PALB2*, fueron introducidas en el minigén mgPALB2\_ex1-3 por mutagénesis dirigida y ensayadas funcionalmente en células MCF-7. Además, 4 de estas variantes fueron ensayadas en células de cáncer de mama triple negativo MDA-MB-231, replicando los resultados obtenidos en MCF-7, como podemos observar en la Figura 26.

Diez de estas variantes alteraban los patrones de *splicing*, de las cuales 8 mostraron una falta completa, o cantidades mínimas del transcrito WT (c.48G>A, 0,9%). Siete variantes afectan a las posiciones clásicas  $\pm 1,2$  del 5' o 3'ss (c.48+1G>A, c.48+2T>C, c.49-2A>T, c.108+1G>A, c.108+2T>C, c.109-2A>G y c.211+1G>A). Todas estas variantes, a excepción de c.108+2T>C, produjeron patrones aberrantes completos.

Las variantes c.108+2T>C, y c.48+2T>C, transforman un donador clásico GT en un donador GC atípico. Sorprendentemente, mientras que c.48+2T>C solo produjo transcritos aberrantes, c.108+2T>C produjo un 85,5% del transcrito canónico. Además, 2 variantes afectan a sitios diferentes del 5'ss, incluyendo los últimos 3 nucleótidos exónicos (c.47A>G, c.48G>A).



**Figura 26. Reproducibilidad de los eventos de *splicing* en células de cáncer de mama triple negativas MDA-MB-231 en el minigén mgPALB2\_ex1-3.**

### 1.3.1.5 Análisis de transcritos producidos por las variantes BRIDGES.

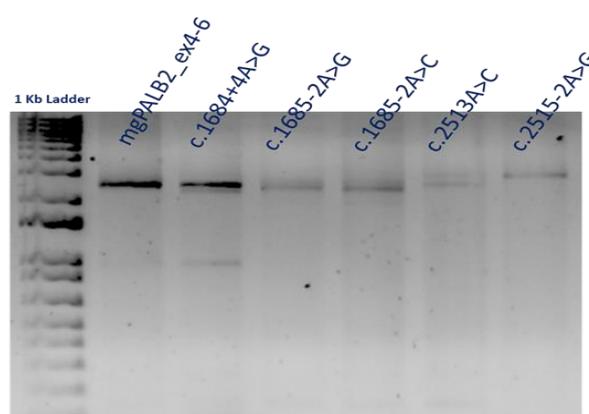
El análisis de fragmentos fluorescentes o estudio de las isoformas mediante electroforesis capilar reveló la existencia de 9 transcritos diferentes, incluyendo el transcrito *wild type*, donde el uso de un sitio alternativo fue el evento de *splicing* más frecuente:  $\Delta$ (E1q17),  $\nabla$ (E1q9),  $\nabla$ (I1<sup>mg</sup>),  $\Delta$ (E2p6),  $\Delta$ (E3p11) y  $\nabla$ (E3q48), aunque también encontramos dos eventos de exón *skipping*:  $\Delta$ (E2),  $\Delta$ (E3) (Tablas 17-18). Cabe destacar, que 4 transcritos incluyen codones de parada prematuros, alterando la pauta de lectura, y propiciando probablemente la generación de una proteína disfuncional. Mientras tanto  $\nabla$ (E1q9),  $\Delta$ (E2p6),  $\Delta$ (E2) y  $\nabla$ (E3q48) mantienen la pauta de lectura. Además, cuatro de las isoformas mencionadas anteriormente han sido previamente caracterizadas como eventos naturales de *splicing*:  $\nabla$ (E1q9),  $\Delta$ (E1q17),  $\Delta$ (E2p6),  $\Delta$ (E2) (Lopez-Perolio et al., 2019).

### 1.3.2 Resultados del minigén mgPALB2\_ex4-6.

De las 15 variantes procedentes del proyecto BRIDGES y comprendidas en los exones 4, 5 y 6 de *PALB2*, 5 fueron seleccionadas para su estudio funcional (Tabla 15), siguiendo los criterios mostrados en el apartado 1.2 de Materiales y Métodos.

Mediante mutagénesis dirigida, se han generado minigenes mutantes a partir del mgPALB2\_ex4-6 (Figura 21). De las 5 variantes ensayadas, 1 afecta al exón 4, (c.1684+4A>G), 3 al exón 5 (c.1695-2A>C, c.1685-2A>G y c.2513A>C) y 1 al exón 6 (c.2515-2A>G).

Las células MCF-7 fueron transfectadas con los 5 minigenes mutantes, y debido al elevado tamaño de los transcritos generados por mgPALB2\_ex4-6, éstos fueron evaluados por diferentes técnicas: amplificaciones con diferentes parejas de oligonucleótidos, electroforesis en gel de agarosa, electroforesis capilar y secuenciación capilar. La Figura 27 muestra los resultados de las RT-PCRs de las 5 variantes obtenidos mediante electroforesis en gel de agarosa.



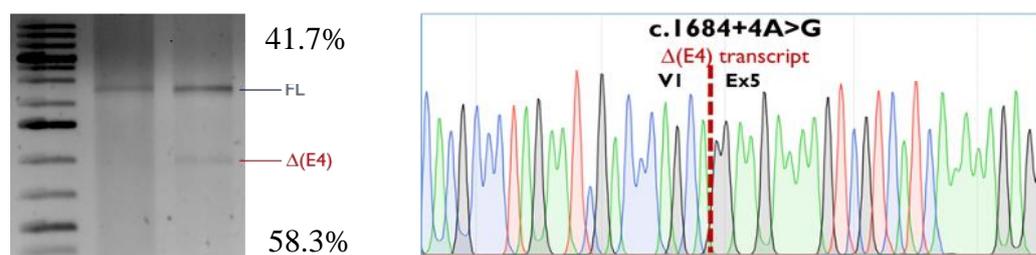
**Figura 27. Transcrito canónico V1-ex4-ex5-ex6-V2 con un tamaño de 2.556 nt y transcritos generados por las variantes: c.1684+4A>G, c.1685-2A>C, c.1685-2A>G, c.2513A>C y c.2515-2A>G.**

El RNA de las variantes c.1685-2A>C, c.1685-2A>G, c.2513A>C y c.2515-2A>G fue retrotranscrito y amplificado con oligonucleótidos marcados fluorescentemente situados en los exones 4 y 6 (FAM-RT-PALB2\_EX6-Rv y RT-PALB2\_ex4-Fw) produciendo un producto de 918 nucleótidos. Posteriormente las diferentes isoformas generadas por estas variantes fueron

analizadas por electroforesis y secuenciación capilar. Por otro lado, debido al elevado tamaño del exón 4 (1.473 pb), se decidió hacer un estudio alternativo de la variante c.1684+4A>G mediante purificación de bandas de gel y un análisis posterior mediante secuenciación capilar y densitometría de imagen.

### 1.3.2.1 Resultados de las variantes del exón 4 de *PALB2*.

La variante **c.1684+4A>G**, no está registrada en ClinVar, aunque sí podemos encontrar la variante c.1684+4A>C, catalogada como posiblemente benigna. Según el estudio bioinformático, **c.1684+4A>G**, altera débilmente el donador del exón 4; aunque lo suficiente para ser llevada a estudio (Tabla 15). Los resultados obtenidos mediante secuenciación capilar y densitometría de imagen, representados en la Figura 28, muestran  $\Delta(E4)$  en un 58,3%, mientras que el 41,7% corresponden al transcrito *full-length* (Figura 28; Tablas 17-18).



**Figura 28. Análisis de la variante c.1684+4A>G, mediante electroforesis en gel de agarosa con los oligonucleótidos: RT-PSPL3-Fw y RT-PALB2-ex6-rv. (A) y secuenciación Sanger del transcrito después de la purificación de la banda de gel (B). Esta variante no pudo analizarse mediante electroforesis capilar debido al gran tamaño del exón 4 (1473 pb).**

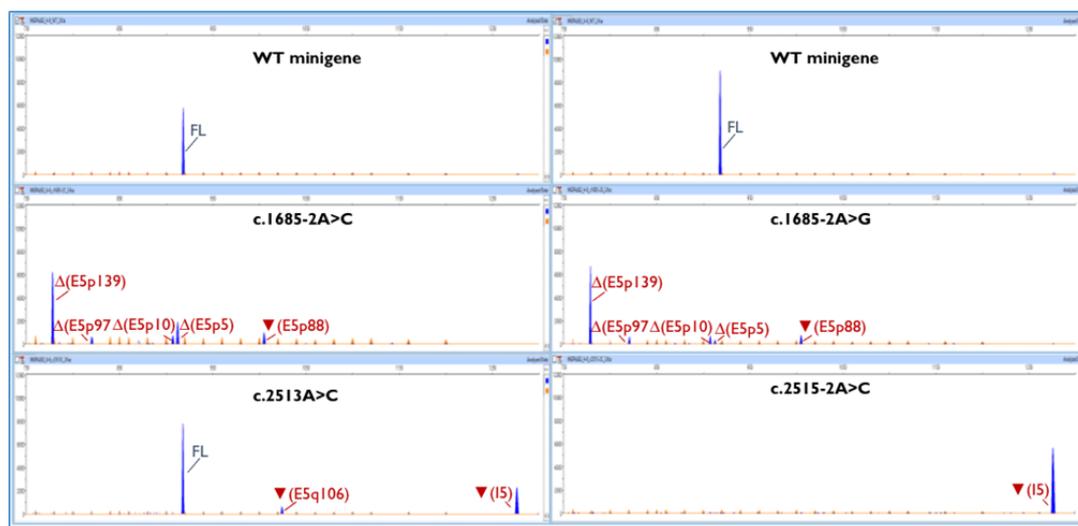
### 1.3.2.2 Resultados de las variantes de los exones 5 y 6 de *PALB2*.

Las variantes de sitio de *splicing* **c.1685-2A>C** y **c.1685-2A>G** están registradas y catalogadas en ClinVar como VUS y como posiblemente patogénica o VUS, respectivamente, creando conflictos de interpretación. Según los estudios bioinformáticos, ambas variantes eliminan el sitio aceptor del exón 5 (Tabla 15). Los resultados de los ensayos funcionales con **c.1685-2A>C**

revelaron la producción de 5 transcritos aberrantes:  $\Delta$ (E5p139)-67%,  $\Delta$ (E5p5)-15,4%,  $\Delta$ (E5p10)-5,7%,  $\Delta$ (E5p97)-5,5% y  $\nabla$ (E5p88)-6,4% (Figura 29; Tablas 17-18). La variante **c.1685-2A>G** produjo exactamente los mismos transcritos, pero en diferentes porcentajes:  $\Delta$ (E5p139)-73,3%,  $\nabla$ (E5p88)-8,7%,  $\Delta$ (E5p10)-7,3%,  $\Delta$ (E5p97)-6,7%,  $\Delta$ (E5p5)-4% (Figura 29; Tablas 17-18).

La variante *missense* **c.2513A>C** p.(Gln838Pro), no está registrada en ClinVar, aunque sí la variante c.2513A>G, clasificada como VUS. Los resultados bioinformáticos determinan que esta variante elimina el sitio donador del exón 5 (Tabla 15). Sin embargo, los resultados de los ensayos funcionales revelaron la producción del transcrito *full-length*, en un 73%. Además, esta variante generó  $\nabla$ (I5) en un 21,2% y  $\nabla$ (E5q106) en un 5,7% (Figura 29; Tablas 17-18).

La última variante ensayada en este minigén, fue **c.2515-2A>G**. No está registrada en la base de datos ClinVar y los resultados bioinformáticos de la Tabla 15 muestran como **c.2515-2A>G**, elimina el sitio aceptor del exón 6. En los ensayos funcionales de **c.2515-2A>G**, se pudo comprobar que esta variante impide el reconocimiento del aceptor del exón 6, generando un único transcrito aberrante con la retención completa del intrón 5 ( $\nabla$ (I5), 100%) (Figura 29; Tablas 17-18).



**Figura 29. Electroforesis capilar de c.1685-2A>C, c.1685-2A>G, c.2513A>C y 2515-2A>G en el minigén mgPALB2\_ex4-6.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

### 1.3.2.3 Recapitulación de resultados de los exones 4 a 6.

Cinco variantes del gen *PALB2*, fueron introducidas en el minigén mgPALB2\_ex4-6 por mutagénesis dirigida y ensayadas funcionalmente en células MCF-7. Todas las variantes alteraban el *splicing*, de las cuales 3 mostraron una falta completa del transcrito WT. Tres variantes afectan a las posiciones clásicas -1, 2 del 3'ss (c.1685-2A>C, c.1685-2A>G y c.2515-2A>G). Además, una variante afecta al penúltimo nucleótido exónico (c.2513A>C) y otra a la posición +4 (c.1684+4A>G).

### 1.3.2.4 Análisis de transcritos producidos por las variantes BRIDGES.

El estudio de las isoformas mediante electroforesis capilar reveló la existencia de 9 transcritos diferentes, incluyendo el transcrito *full-length*, donde el uso de sitios alternativos fue el evento de *splicing* más frecuente:  $\Delta$ (E5p5),  $\Delta$ (E5p10),  $\Delta$ (E5p97),  $\Delta$ (E5p139),  $\blacktriangledown$ (E5p88) y  $\blacktriangledown$ (E5q106), aunque también encontramos un evento de exón *skipping*  $\Delta$ (E4) y otro de retención total del intrón 5,  $\blacktriangledown$ (I5) (Tabla 16). Dos de estas isoformas,  $\Delta$ (E4) y  $\Delta$ (E5p139), habían sido previamente caracterizadas por López-Perolio y colaboradores como eventos de *splicing* alternativo natural (Lopez-Perolio et al., 2019). Además, todos los eventos de *splicing* producidos a causa de estas variantes, a excepción de  $\Delta$ (E4), alteran la pauta de lectura e introducen codones de parada prematuros, propiciando probablemente la producción de proteínas disfuncionales.

### 1.3.3 Resultados del minigén mgPALB2\_ex5-12.

De las 42 variantes procedentes del proyecto BRIDGES comprendidas entre los exones 6 a 12 de *PALB2*, 26 fueron seleccionadas para su estudio funcional siguiendo los criterios bioinformáticos indicados en el apartado 1.2 de Materiales y Métodos. Mediante mutagénesis dirigida se han generado minigenes mutantes a partir del mgPALB2\_ex5-12 (Figura 22). De las 26 variantes ensayadas, encontramos una que afecta a los sitios de *splicing* del exón 6 (c.2586+4A>T), 5 al exón 7 (c.2587-2A>G, c.2748+1G>T, c.2748+1G>A, c.2748+2dup y c.2748+4A>T), 8 al exón

8 (c.2749-1G>T, c.2750T>C, c.2834G>C, c.2834+1G>A, c.2834+2T>C, c.2834+3A>G, c.2834+5G>A y c.2834+6T>C), 1 al exón 9, (c.2996+4A>G), 5 al exón 10 (c.2997-1G>A, c.2997-2del, c.3113G>A, c.3113+3A>G y c.3113+5G>C), 3 al exón 11 (c.3201+1G>A, c.3201+6T>A y c.3201+3\_3201+4insTG) y 3 al exón 12 ( c.3348C>T, c.3350+4A>G y c.3350+5G>A).

Estos minigenes mutantes se introdujeron en células MCF-7 y los resultados se comprobaron por electroforesis en gel de agarosa, electroforesis capilar y secuenciación capilar. A continuación, se muestran los resultados obtenidos.

### **1.3.3.1 Resultados de las variantes del exón 6 de *PALB2*.**

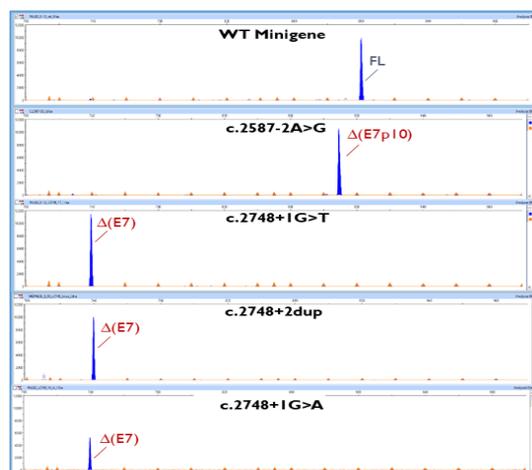
La variante **c.2586+4A>T**, está registrada en la base de datos ClinVar y catalogada como VUS. Según el estudio informático, la variante elimina el sitio donador del exón 6 (Tabla 16). Los resultados del ensayo funcional representados en la Figura 33, mostraron la presencia única del transcrito *full-length*.

### **1.3.3.2 Resultados de las variantes del exón 7 de *PALB2*.**

La variante, **c.2587-2A>G**, no está registrada en la base de datos ClinVar. Sin embargo, la variante c.2587-2A>C, se encuentra registrada y catalogada como variante de *splicing* y estudios desarrollados por López-Perolio y colaboradores en 2019, demuestra su implicación en la generación de transcritos aberrantes en *PALB2* (Lopez-Perolio et al., 2019). Los estudios bioinformáticos, determinan que esta variante elimina el sitio aceptor del exón 7 de *PALB2* (Tabla 16). Los resultados del ensayo funcional mostraron como único transcrito la isoforma aberrante ( $\Delta(E7)$ ,100%) (Figura 30; Tablas 17-18); similar al resultado obtenido con la variante c.2587-2A>C, citada anteriormente.

Por otro lado, **c.2748+1G>A**, **c.2748+1G>T** y **c.2748+2dup** son variantes clasificadas en ClinVar como patogénicas o posiblemente patogénicas, que afectan al donador del exón 7 de *PALB2*. Según el estudio bioinformático, estas variantes eliminan el sitio donador del exón 7 (Tabla

16). El análisis de fragmentos fluorescente reveló la presencia única del transcrito aberrante ( $\Delta(E7)$ ) en todos los ensayos (Figura 30; Tablas 17-18).



**Figura 30. Electroforesis capilar de c.2587-2A>G, c.2748+1G>T, c.2748+2dup y c.2748+1G>A en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

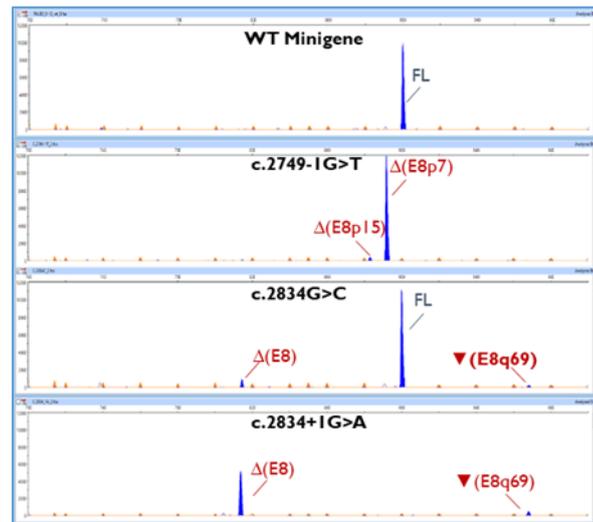
No todas las variantes ensayadas en este exón tuvieron efectos sobre el proceso de *splicing*. El estudio bioinformático de la variante **c.2748+4A>T** predecía alteraciones en *splicing* (Tabla 16), sin embargo, los ensayos funcionales pusieron de manifiesto, la producción única del transcrito canónico (Figura 33).

### 1.3.3.3 Resultados de las variantes del exón 8 de *PALB2*.

La variante de sitio de *splicing*, **c.2749-1G>T** está registrada en la base de datos ClinVar y catalogada como posiblemente patogénica. Según el estudio informático, **c.2749-1G>T**, causa la eliminación del sitio aceptor del exón 8 (Tabla 16). Los resultados del ensayo funcional revelaron la producción de forma mayoritaria de  $\Delta(E8p7)$  en un 97% y de ( $\Delta(E8p15)$ , 3%) (Figura 31; Tablas 17-18).

Las variantes *missense* y de sitio de *splicing*, **c.2834G>C** p.(Arg945Thr), y **c.2834+1G>A**, se encuentran registradas en ClinVar y están catalogadas como VUS y patogénica, respectivamente. Los resultados de los ensayos funcionales con **c.2834G>C** mostraron la producción de transcritos

*full-length* en un 90%. El resto transcritos, aunque eran producidos de forma minoritaria, procedían de eventos anómalos de *splicing*: ( $\Delta$ (E8), 7,8%) y ( $\nabla$ (E8q69), 2,2%) (Figura 31; Tablas 17-18). Por otro lado, la variante **c.2834+1G>A** generó en su totalidad transcritos aberrantes. El 82,2% no incluían el exón 8 y el 17,9% se generaban por el uso de un donador alternativo 69 nucleótidos *downstream* del natural (Figura 31; Tablas 17-18).

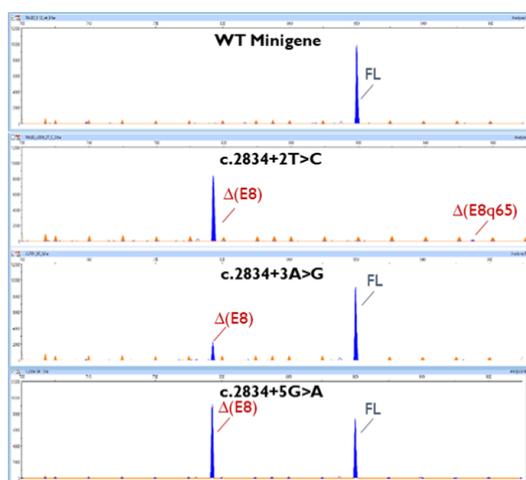


**Figura 31. Electroforesis capilar de c.2749-1G>T, c.2834G>C y c.2834+1G>A en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

La variante **c.2834+2T>C** de sitio de *splicing* se encuentra registrada en la base de datos ClinVar y está clasificada como posiblemente patogénica. Los resultados de los ensayos funcionales demostraron que este donador GC, creado *de novo* no es reconocido por la maquinaria de *splicing* (Figura 32). Los resultados obtenidos con esta variante revelaron la presencia única del transcrito aberrante ( $\Delta$ (E8), 100%) (Figura 32; Tablas 17-18).

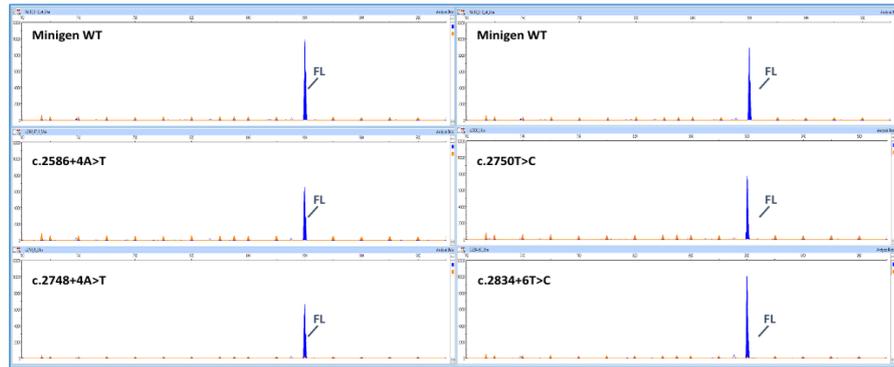
Las variantes **c.2834+3A>G** y **c.2834+5G>A** también están registradas en la base de datos ClinVar y clasificadas como VUS o posiblemente patogénica y como variante de significado clínico incierto, respectivamente. El análisis *in silico* de ambas variantes, predice la eliminación del sitio donador del exón 8 (Tabla 16). Los resultados de los ensayos funcionales determinaron que

**c.2834+3A>G** y **c.2834+5G>A** generaban el transcrito canónico (80,2% y 40,7%, respectivamente) y el aberrante  $\Delta(E8)$  (19,8% y 59,3%, respectivamente) (Figura 32; Tablas 17-18)



**Figura 32. Electroforesis capilar de c.2834+2T>C c.2834+3A>G y c.2834+5G>A en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

No todas las variantes ensayadas en este exón tuvieron efectos sobre el proceso de *splicing*. Así, la variante *missense* **c.2750T>C** p.(Val917Ala), clasificada en ClinVar como VUS y **c.2834+6T>C** (no registrada) alteraban el sitio donador del exón 8 de acuerdo al estudio bioinformático (Tabla 16). Sin embargo, los ensayos funcionales no mostraron ninguna alteración del *splicing* (Figura 33; Tablas 17-18).



**Figura 33. Electroforesis capilar de c.2586+4A>T, c.2748+4A>T, c.2750T>C y 2834+6T>C en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

#### 1.3.3.4 Resultados de las variantes del exón 9 de *PALB2*.

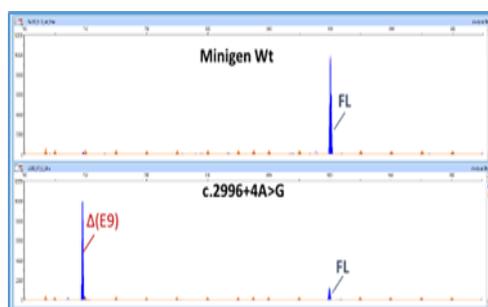
La variante **c.2996+4A>G** no está registrada en la base de datos ClinVar. Los resultados bioinformáticos predicen la eliminación del sitio donador del exón 9 (Tabla 16). Los ensayos funcionales con esta variante determinaron la existencia de alteraciones en los patrones de *splicing* y se pudo comprobar mediante electroforesis capilar la presencia mayoritaria del transcrito anómalo  $\Delta(E9)$  (87,1%; exclusión del exón 9), y también del transcrito canónico (12,9%) (Figura 34; Tablas 17-18).

#### 1.3.3.5 Resultados de las variantes del exón 10 de *PALB2*.

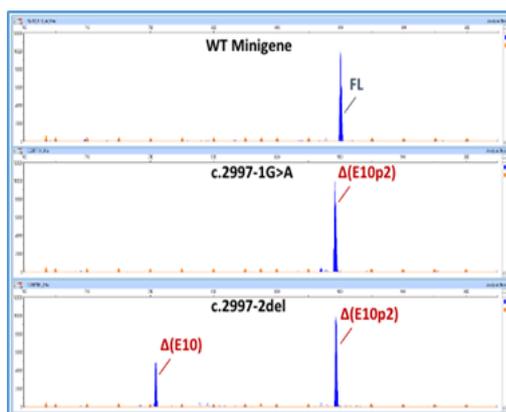
Existen 2 variantes BRIDGES que afectan a los sitios de *splicing* del exón 10: **c.2997-1G>A** y **c.2997-2del**. La primera de éstas, está representada en ClinVar y catalogada como posiblemente patogénica y como VUS, creando conflictos de interpretación, sin embargo, **c.2997-2del** no está registrada. Según los resultados bioinformáticos, ambas variantes eliminan el sitio aceptor del exón 10. Además, según MES, **c.2997-1G>A** generaba un aceptor alternativo *de novo* 2 nt *downstream* (Tabla 16). La electroforesis capilar reveló que la variante **c.2997-1G>A** generaba

un único transcrito aberrante ( $\Delta(E10p2)$ , 100%). Por otro lado, **c.2997-2del** también producía ( $\Delta(E10p2)$ , 66,6%) y el transcrito  $\Delta(E10)$ , 33,3%) (Figura 35; Tablas 17-18).

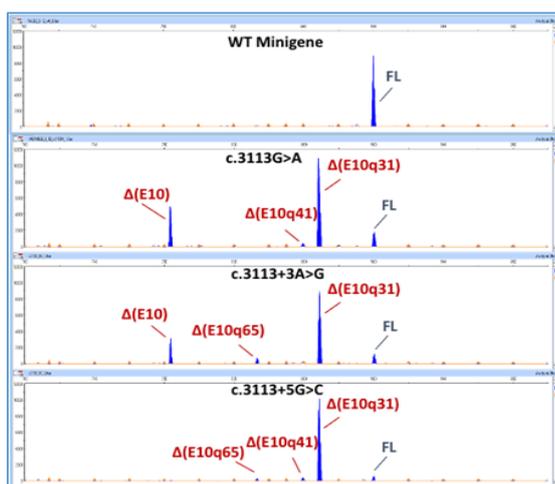
La variante *nonsense* **c.3113G>A** p.(Trp1038Ter) y **c.3113+5G>C**, clasificadas como patogénicas o posiblemente patogénicas, y **c.3113+3A>G** (no registrada) afectan al sitio donador del exón 10. El estudio bioinformático de las 3 variantes predice la eliminación del sitio donador del exón 10 (Tabla 16). Los datos obtenidos en los ensayos funcionales revelaron que **c.3113G>A** producía los siguientes transcritos aberrantes:  $\Delta(E10q31)$ -52,9%,  $\Delta(E10)$ -28,1%,  $\Delta(E9-10)$ -6,3% y  $\Delta(E10q41)$ -2,2%, además de un 10,5% del transcrito *full-length* (Figura 36). Por otro lado, la variante **c.3113+3A>G** mostró 4 transcritos diferentes:  $\Delta(E10q31)$ -62,7%,  $\Delta(E10)$ -19,9%,  $\Delta(E9-10)$ -4,5% y  $\Delta(E10q65)$ -4,9%). Además, produjo el transcrito *full-length* en un 8% (Figura 36; Tablas 17-18). Por último, la variante **c.3113+5G>C** indujo los transcritos ( $\Delta(E10q31)$ , 90%), ( $\Delta(E10q41)$ , 2,8%), ( $\Delta(E10q65)$ , 2,3%) y el transcrito canónico esperado (4,9%) (Figura 36; Tablas 17-18).



**Figura 34. Electroforesis capilar de c.2996+4A>G en el minigen mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).



**Figura 35. Electroforesis capilar de c.2997-1G>A y c.2997-2del en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

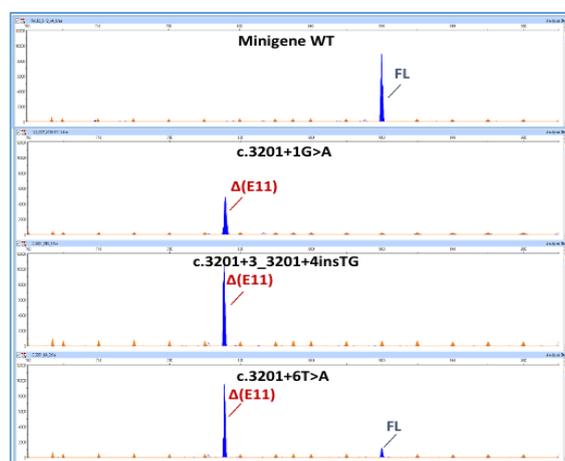


**Figura 36. Electroforesis capilar de c.3113G>A, c.3113+3A>G y c.3113+5G>C en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

### 1.3.3.6 Resultados de las variantes del exón 11 de PALB2.

Las variantes **c.3201+1G>A** y **c.3201+3\_3201+4insTG** aún no han sido reportadas en ClinVar. El estudio bioinformático de ambas variantes predice la rotura del sitio donador del exón 11 (Tabla 16). El análisis de fragmentos fluorescente reveló en ambas variantes la presencia de un único transcrito aberrante ( $\Delta(E11)$ , 100%) (Figura 37; Tablas 17-18). Por otro lado, la variante

**c.3201+6T>A**, reportada en ClinVar, disminuía levemente la fortaleza del sitio donador del exón 11, pero en una proporción menor que en los anteriores casos. Los resultados de *splicing* revelaron  $\Delta(E11)$  en un 88,5% y transcrito *full-length* en un 11,5% (Figura 37; Tablas 17-18).



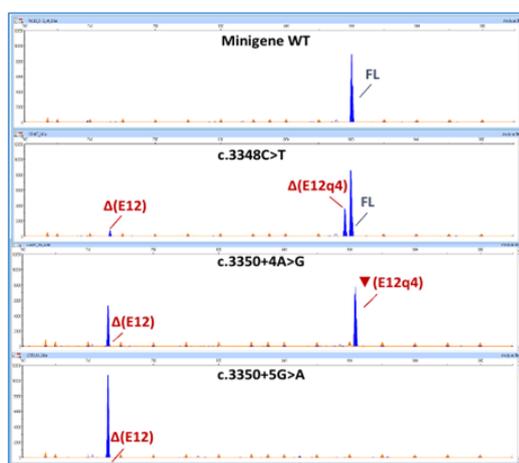
**Figura 37. Electroforesis capilar de c.3201+1G>A, c.3201+3\_3201+4insTG y c.3201+6T>A en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

### 1.3.3.7 Resultados de las variantes del exón 12 de *PALB2*.

La variante sinónima **c.3348C>T** (p.(Gly1116=)) está reportada en ClinVar como VUS. Afecta al antepenúltimo nucleótido del exón 12 de *PALB2*, y según los estudios bioinformáticos, esta variante debilita el sitio donador del exón 12 (Tabla 16). Los ensayos funcionales de esta variante revelaron la presencia del transcrito canónico de forma mayoritaria (68%) (Figura 38; Tablas 17-18). Además, se detectaron los transcritos  $\Delta(E12q4)$  (uso de un donador alternativo 4 nt *upstream*, 26,7%) y  $\Delta(E12)$  (pérdida del exón 12, 5,3%) (Figura 38; Tablas 17-18).

Las variantes **c.3350+4A>G** y **c.3350+5G>A** están registradas en la base de datos ClinVar, la primera clasificada como patogénica/posiblemente patogénica y la segunda presenta interpretaciones conflictivas de su patogenicidad. Ambas variantes se sitúan en la secuencia consenso del sitio donador del exón 12, y según los resultados bioinformáticos, estas dos variantes eliminan el sitio donador del exón 12 y crean un sitio crítico *de novo* (Tabla 16). El ensayo

funcional de **c.3350+5G>A** reveló exclusivamente el transcrito aberrante ( $\Delta(E12)$ , 100%), mientras que **c.3350+4A>G** generó  $\nabla(E12q4)$  (inserción de los primeros 4 nt del intrón 12 por el uso de un nuevo donador; 56,1%) y ( $\Delta(E12)$ , 43,9%) (Figura 38; Tablas 17-18).

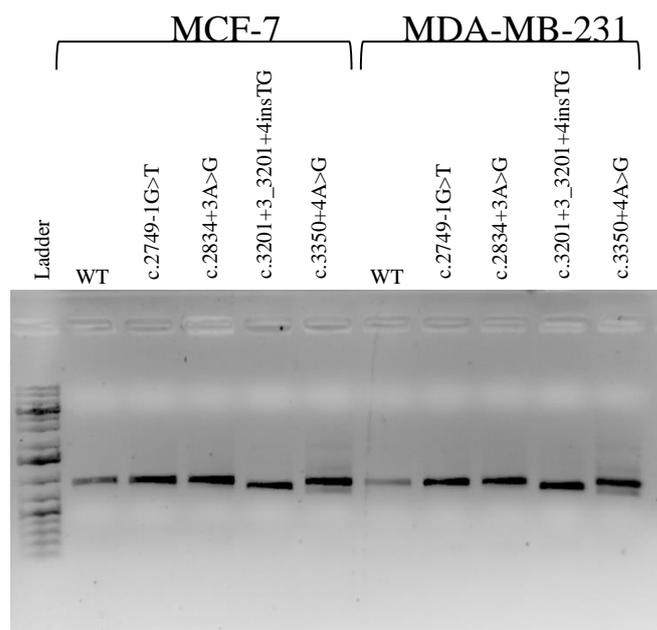


**Figura 38. Electroforesis capilar de c.3348C>T, c.3350+4A>G y c.3350+5G>A en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

### 1.3.3.8 Recapitulación de resultados de los exones 6 a 12.

Veintiséis variantes del gen *PALB2*, fueron introducidas en el minigén mgPALB2\_ex5-12 por mutagénesis dirigida y ensayadas funcionalmente en células MCF-7. Además, 4 de estas variantes fueron ensayadas en células de cáncer de mama triple negativo MDA-MB-231, en las cuales replicaron los resultados obtenidos en células MCF-7, como podemos observar en la Figura 39.

Veintidós de estas variantes alteraban el *splicing*, de las cuales 14 mostraron una falta completa, o cantidades mínimas del transcrito canónico esperado (c.3113+5G>C-4.9%). Diez variantes afectan a las posiciones clásicas  $\pm 1, 2$  y ninguna de ellas produce trazas del transcrito canónico. Además, 12 variantes afectan a sitios diferentes del 5'ss, incluyendo los últimos 3 nucleótidos exónicos (c.2834G>C, c.3113G>A y c.3348C>T) y posiciones intrónicas +3 (c.2834+3A>G, c.3113+3A>G y c.3201+3\_3201+4insTG), +4 (c.2996+4A>G y c.3350+4A>G), +5 (c.2834+5G>A, c.3113+5G>C y c.3350+5G>A), y +6 (c.3201+6T>A).



**Figura 39. Reproducibilidad de los eventos de *splicing* en células de cáncer de mama triple negativas MDA-MB-231 en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv (Tamaño = 906 nt).

### 1.3.3.9 Análisis de transcritos producidos por las variantes en el minigén mgPALB2\_ex5-12.

El análisis de fragmentos fluorescentes reveló la existencia de 21 transcritos diferentes, incluido el transcrito canónico. El uso de sitios alternativos fue el mecanismo de *splicing* más frecuente (10 transcritos):  $\Delta$ (E7p10),  $\Delta$ (E8p7),  $\Delta$ (E8p15),  $\blacktriangledown$ (E8q69),  $\Delta$ (E10p2),  $\Delta$ (E10q31),  $\Delta$ (E10q41),  $\Delta$ (E10q65),  $\Delta$ (E12q4) y  $\blacktriangledown$ (E12q4). Tres de ellos,  $\blacktriangledown$ (E8q69),  $\Delta$ (E10q41) y  $\Delta$ (E10q65), se generan por el uso de donadores GC atípicos. El segundo mecanismo más frecuente fue el exón *skipping*, que fue detectado en 7 isoformas:  $\Delta$ (E7),  $\Delta$ (E8),  $\Delta$ (E9),  $\Delta$ (E10),  $\Delta$ (E9\_10),  $\Delta$ (E11) y  $\Delta$ (E12). Sin embargo, no se pudieron caracterizar tres transcritos: uno de 888 nt originado por c.2997-1G>A y dos de 592 nt y 710 nt producidos por c.2748+2dup, con porcentajes menores del 9%.

Cabe destacar, que de los 20 transcritos producidos a causa de estas variantes, únicamente 5 mantienen la pauta de lectura:  $\Delta$ (E7),  $\Delta$ (E8p15),  $\Delta$ (E9),  $\Delta$ (E10) y  $\Delta$ (E9\_10). El resto de estos

eventos de *splicing* introducen codones de terminación prematuros, propiciando probablemente la generación de una proteína disfuncional.

Finalmente, diez de las isoformas mencionadas anteriormente han sido previamente caracterizadas como eventos naturales de *splicing*:  $\Delta(E7p10)$ ,  $\Delta(E7)$ ,  $\Delta(E8)$ ,  $\Delta(E9)$ ,  $\Delta(E10)$ ,  $\Delta(E9\_10)$ ,  $\Delta(E10p2)$ ,  $\Delta(E10q31)$ ,  $\Delta(E11)$  y  $\Delta(E12)$  (Lopez-Perolio et al., 2019).

---

## **2. Análisis *in silico* y funcional de variantes de *PALB2* reportadas en la base de datos ClinVar.**

### **2.1. Estudio bioinformático de variantes ClinVar.**

Se han analizado todas las variantes reportadas en ClinVar del gen *PALB2*, en los límites intrón-exón con el algoritmo MaxEntScan (MES) (Yeo and Burge, 2004), para identificar posibles variantes *esplíceogénicas*. Las variantes candidatas se seleccionaron de acuerdo a los criterios indicados en el apartado 1.1 de Resultados. Cuando existe más de una variante que afecta al mismo sitio de *splicing* y con la misma predicción bioinformática, nuestra selección se fundamentará, en la importancia clínica de la variante y el número de registros observados. Así, se han analizado **151** variantes ClinVar de *PALB2* localizadas en las fronteras intrón/exón. Las Tablas 19, 20 y 21 muestran, ordenadas por minigenes y exones, todas las variantes de DNA analizadas.

Tabla 19. Variantes de *PALB2* estudiadas *in silico* de la base de datos ClinVar en el minigén *mgPALB2\_ex1-3*.

VARIANTES <i>PALB2</i> <sup>1</sup>	EXÓN/ INTRÓN	MES wt	MES mut	% Cambio MES <sup>2</sup>	Número de registros
<b>c.46A&gt;G</b>	Ex1	5.74	4.05	-29.48%	1
<b>c.48+1G&gt;C</b>	IVS1	5.74	-2.53	-144%	3
<b>c.48+1del</b>	IVS1	5.74	-12.45	-316.7%	1
<b>c.48+1G&gt;T</b>	IVS1	5.74	-2.76	-148.1%	1
<b>c.48+2T&gt;G</b>	IVS1	5.74	-1.9	-133.15%	1
<b>c.48+5C&gt;T</b>	IVS1	5.74	4.05	-29.53%	3
<b>c.48+6G&gt;C</b>	IVS1	5.74	5.96	+3.81%	1
<b>c.49-1G&gt;A</b>	IVS1	9.28	0.53	-94.26%	1
<b>c.49-1del</b>	IVS1	9.28	-7.59	-181.8%	1
<b>c.49-2del</b>	IVS1	9.28	1.16	-87.54%	1
<b>c.50T&gt;G</b>	Ex2	9.28	9.02	-2.8%	2
<b>c.50dup</b>	Ex2	9.28	9.25	-0.3%	1
<b>c.106C&gt;T</b>	Ex2	10.86	9.66	-11%	4
<b>c.108G&gt;A</b>	Ex2	10.86	10.08	-7.19%	1
<b>c.108+1_108+2insC</b>	IVS2	10.86	-4.16	-138.3%	1
<b>c.108+4A&gt;G</b>	IVS2	10.86	10.28	-5.3%	3
<b>c.108+5G&gt;A</b>	IVS2	10.86	8.76	-19.3%	1
<b>c.108+6T&gt;C</b>	IVS2	10.86	9.88	-9%	1
<b>c.109-2A&gt;C</b>	IVS2	10.06	2.02	-79.9%	3
<b>c.109C&gt;G</b>	Ex3	10.06	11.86	+17.9%	2
<b>c.109C&gt;T</b>	Ex3	10.06	9.82	-2.5%	5
<b>c.109C&gt;A</b>	Ex3	10.06	10.18	+1.1%	7
<b>c.110G&gt;T</b>	Ex3	10.06	9.69	-3.7%	2
<b>c.110G&gt;A</b>	Ex3	10.06	9.64	-4.3%	9
<b>c.111T&gt;C</b>	Ex3	10.06	10.3	+2.3%	3

VARIANTES <i>PALB2</i> <sup>1</sup>	EXÓN/ INTRÓN	MES wt	MES mut	% Cambio MES <sup>2</sup>	Número de registros
<b>c.210A&gt;G</b>	Ex3	8.76	6.15	-29.9%	2
<b>c.210A&gt;C</b>	Ex3	8.76	6.89	-21.4%	1
<b>c.211+1G&gt;T</b>	IVS3	8.76	0.26	-97%	1
<b>c.211+2T&gt;C</b>	IVS3	8.76	1.01	-88.5%	1
<b>c.211+4A&gt;G</b>	IVS3	8.76	7.25	-17.2%	1
<b>c.211+6T&gt;A</b>	IVS3	8.76	8.59	-1.9%	1

<sup>1</sup>Las variantes seleccionadas se encuentran marcadas en **rojo** <sup>2</sup>El símbolo [+] implica ganancia, mientras que [-] pérdida.

Tabla 20. Variantes de *PALB2* estudiadas *in silico* de la base de datos ClinVar en el minigén mgPALB2\_ex4-6.

VARIANTES <i>PALB2</i> <sup>1</sup>	EXÓN/ INTRÓN	MES wt	MES mut	% Cambio MES <sup>2</sup>	Número de registros
c.212-3C>T	IVS3	10.02	9.85	-1.7%	2
<b>c.212-2A&gt;G</b>	IVS3	10.02	2.07	-79.4%	6
c.212-2A>T	IVS3	10.02	1.66	-83.5%	1
c.212-2A>C	IVS3	10.02	1.98	-80.2%	1
c.212-1G>T	IVS3	10.02	1.42	-85.8%	1
<b>c.212-1G&gt;A</b>	IVS3	10.02	1.27	-87.3%	1
c.212A>G	Ex4	10.02	10.78	+7.6%	1
c.213A>C	Ex4	10.02	10.13	+1.1%	1
c.213A>T	Ex4	10.02	11.03	+10.1%	6
c.213A>G	Ex4	10.02	10.10	+2.8%	1
<b>c.1684G&gt;A</b>	Ex4	8.88	4.06	-54.3%	2
<b>c.1684+1G&gt;A</b>	IVS4	8.88	0.7	-92.1%	2
c.1684+1G>T	IVS4	8.88	0.38	-95.7%	1
c.1684+4A>C	IVS4	8.88	7.87	-11.5%	1
c.1684+6del	IVS4	8.88	8.14	-8.4%	2
c.1684+6_1684+9del	IVS4	8.88	9.06	+2%	1
<b>c.1685-3T&gt;C</b>	IVS4	11.15	10.96	-1.8%	1
<b>c.1685-3T&gt;G</b>	IVS4	11.15	-3.65	-132.8%	1
<b>c.1685-1G&gt;A</b>	IVS4	11.15	2.4	-78.5%	2
<b>c.1685-1G&gt;C</b>	IVS4	11.15	3.09	-72.3%	2
c.1686G>A	Ex4	11.15	11.56	+3.6%	2
c.2512C>G	Ex5	7.09	5.13	-27.7%	3
c.2512C>T	Ex5	7.09	3.23	-54.4%	3
<b>c.2513A&gt;G</b>	Ex5	7.09	2.62	-63.1%	3
<b>c.2514G&gt;A</b>	Ex5	7.09	-1.09	-115.3%	1

VARIANTES <i>PALB2</i> <sup>1</sup>	EXÓN/ INTRÓN	MES wt	MES mut	% Cambio MES <sup>2</sup>	Número de registros
<b>c.2514+1G&gt;C</b>	IVS5	7.09	-1.18	-116.6%	2
<b>c.2514+1G&gt;A</b>	IVS5	7.09	-1.09	115.4%	1
<b>c.2514+2T&gt;C</b>	IVS5	7.09	-0.66	-109.3%	2
<b>c.2514+3dup</b>	IVS5	7.09	8.88	+25.2%	3
<b>c.2514+3A&gt;G</b>	IVS5	7.09	2.83	-60.1%	5
<b>c.2514+4C&gt;T</b>	IVS5	7.09	7.46	+5.2%	1
<b>c.2514+5A&gt;G</b>	IVS5	7.09	10.03	+41.4%	1
<b>c.2515-1G&gt;A</b>	IVS5	9.47	0.72	-92.4%	3
<b>c.2515-1G&gt;C</b>	IVS5	9.47	1.41	-85.2%	4
<b>c.2515-1G&gt;T</b>	IVS5	9.47	0.87	-90.8%	4
<b>c.2517T&gt;G</b>	Ex6	9.47	9.55	+0.85%	1
<b>c.2517_2518del</b>	Ex6	9.47	9.47	+1.64%	1

<sup>1</sup>Las variantes seleccionadas se encuentran marcadas en **rojo** <sup>2</sup>El símbolo [+] implica ganancia, mientras que [-] pérdida.

Tabla 21. Variantes de PALB2 estudiadas in silico de la base de datos ClinVar en el minigén mgPALB2\_ex5-12.

VARIANTES PALB2 <sup>1</sup>	EXÓN/ INTRÓN	MES wt	MES mut	% Cambio MES <sup>2</sup>	Número de registros
<b>c.2586G&gt;C</b>	Ex6	6.80	1.06	-84.4%	3
<b>c.2586+1G&gt;T</b>	IVS6	6.80	-1.70	-125%	1
<b>c.2586+3C&gt;G</b>	IVS6	6.80	8.68	+27.6%	1
<b>c.2586+3C&gt;T</b>	IVS6	6.80	6.12	-10.1%	2
<b>c.2587-2A&gt;G</b>	IVS6	3.31	-4.64	-240.2%	1
<b>c.2587-1G&gt;C</b>	IVS6	3.31	-4.75	-243.5%	3
<b>c.2587A&gt;G</b>	Ex7	3.31	4.27	+28.8%	1
<b>c.2747_c.2748+4del</b>	IVS7	11.08	-12.91	-39.5%	4
<b>c.2748+1G&gt;A</b>	IVS7	11.08	2.90	-73.85%	2
<b>c.2748+1G&gt;C</b>	IVS7	11.08	3.32	-74.67%	1
<b>c.2748+1G&gt;T</b>	IVS7	11.08	2.57	-76.76%	5
<b>c.2748+1_2748+2del</b>	IVS7	11.08	-11.89	-207.3%	1
<b>c.2748+2T&gt;C</b>	IVS7	11.08	2.81	-70%	1
<b>c.2748+3A&gt;C</b>	IVS7	11.08	7.7	-30.5%	1
<b>c.2749-2A&gt;C</b>	IVS7	9.58	1.54	-83.9%	1
<b>c.2749-1G&gt;A</b>	IVS7	9.58	0.83	-91.3%	1
<b>c.2749-1G&gt;C</b>	IVS7	9.58	1.52	-84.2%	1
<b>c.2749G&gt;A</b>	Ex8	9.58	8.16	-14.83	2
<b>c.2832C&gt;G</b>	Ex8	9.8	9.81	+0.11%	1
<b>c.2832del</b>	Ex8	9.8	7.76	-20.8%	2
<b>c.2834+1G&gt;T</b>	IVS8	9.8	1.3	-86.8%	7
<b>c.2834+1G&gt;C</b>	IVS8	9.8	1.53	-84.4%	6
<b>c.2834+2T&gt;C</b>	IVS8	9.8	2.05	-79.1%	2
<b>c.2834+2T&gt;G</b>	IVS8	9.8	2.15	-78%	1
<b>c.2834+4T&gt;C</b>	IVS8	9.8	10.65	+8.7%	5

VARIANTES <i>PALB2</i> <sup>1</sup>	EXÓN/ INTRÓN	MES wt	MES mut	% Cambio MES <sup>2</sup>	Número de registros
<b>c.2834+5G&gt;A</b>	IVS8	9.8	7.88	-19.6%	3
<b>c.2835-3T&gt;C</b>	IVS8	8.14	7.2	-11.5%	2
<b>c.2835-2A&gt;C</b>	IVS8	8.14	0.09	-98.8%	1
<b>c.2835-1G&gt;C</b>	IVS8	8.14	0.07	-99.1%	5
<b>c.2835-1G&gt;A</b>	IVS8	8.14	-0.61	-107.5%	3
<b>c.2835G&gt;A</b>	Ex9	8.14	6.73	-17.2%	1
<b>c.2836G&gt;A</b>	Ex9	8.14	7.99	-1.7%	1
<b>c.2836G&gt;T</b>	Ex9	8.14	8.43	+3.6%	4
<b>c.2837del</b>	Ex9	8.14	7.74	-4.8%	1
<b>c.2837C&gt;G</b>	Ex9	8.14	7.44	-9.9%	1
<b>c.2995G&gt;A</b>	Ex9	9.21	10.57	+14.8%	1
<b>c.2996G&gt;C</b>	Ex9	9.21	6.21	-32.5%	1
<b>c.2996G&gt;A</b>	Ex9	9.21	5.73	-37.7%	1
<b>c.2996+1G&gt;T</b>	IVS9	9.21	0.7	-92.4%	2
<b>c.2996+2T&gt;C</b>	IVS9	9.21	1.45	-84.2%	1
<b>c.2996+3A&gt;G</b>	IVS9	9.21	6.65	-27.8%	2
<b>c.2996+3A&gt;C</b>	IVS9	9.21	3.16	-65.6%	1
<b>c.2996+5G&gt;T</b>	IVS9	9.21	5.47	-40.6%	7
<b>c.2997-3C&gt;G</b>	IVS10	6.69	-5.39	-180.5%	1
<b>c.2997-2A&gt;T</b>	IVS10	6.69	-1.67	-125%	1
<b>c.2997A&gt;T</b>	Ex10	6.69	6.32	-5.6%	1
<b>c.2998G&gt;A</b>	Ex10	6.69	6.78	+1.3%	1
<b>c.2999_3001del</b>	Ex10	6.69	6.95	+3.9%	1
<b>c.2999del</b>	Ex10	6.69	7.45	+11.33	1
<b>c.2999G&gt;T</b>	Ex10	6.69	7.09	+5.9%	1
<b>c.2999G&gt;A</b>	Ex10	6.69	6.95	+3.9%	1

VARIANTES <i>PALB2</i> <sup>1</sup>	EXÓN/ INTRÓN	MES wt	MES mut	% Cambio MES <sup>2</sup>	Número de registros
c.3111T>C	Ex10	8.72	8.69	-0.33%	1
c.3113+1G>A	IVS10	8.72	0.54	-93.8%	1
c.3113+5G>C	IVS10	8.72	3.53	-59.5%	1
c.3114-3A>G	IVS10	4.13	-0.17	-104.2%	2
c.3114-1G>A	IVS10	4.13	-4.62	-211.7%	3
c.3114del	Ex11	4.13	2.68	-35.3%	1
c.3114G>A	Ex11	4.13	2.11	-49.1%	3
c.3116A>C	Ex11	4.13	4.24	+2.4%	1
c.3116del	Ex11	4.13	4.55	+10.1%	13
c.3201G>A	Ex11	11.01	8.65	-21.4%	2
c.3201+1G>C	IVS11	11.01	2.74	-75.1%	7
c.3201+1G>T	IVS11	11.01	2.51	-77.2%	2
c.3201+2dup	IVS11	11.01	-1.78	-116.1%	2
c.3201+3A>C	IVS11	11.01	6.57	-40.4%	1
c.3201+4A>G	IVS11	11.01	8.60	-21.9%	1
c.3201+4del	IVS11	11.01	0.55	-95%	2
c.3201+5G>T	IVS11	11.01	6.49	-41%	1
c.3201+5_3201+8del	IVS11	11.01	5.46	-50.4%	2
c.3201+6T>A	IVS11	11.01	9.48	-14%	2
c.3202-2A>C	IVS11	6.42	-1.62	-125.2%	3
c.3202-2A>G	IVS11	6.42	-1.53	-123.8%	1
c.3202-1G>C	IVS11	6.42	-1.64	-125.5%	4
c.3202-1G>A	IVS11	6.42	-2.33	-136.2%	4
c.3202G>T	Ex12	6.42	4.22	-34.4%	2
c.3203G>A	Ex12	6.42	5.94	-3.2%	3
c.3204G>T	Ex12	6.42	7.33	+14.1%	2

VARIANTES <i>PALB2</i> <sup>1</sup>	EXÓN/ INTRÓN	MES wt	MES mut	% Cambio MES <sup>2</sup>	Número de registros
<b>c.3204G&gt;A</b>	Ex12	6.42	6.84	+6.4%	4
<b>c.3350G&gt;A</b>	Ex12	3.1	2.32	-25.6%	5
<b>c.3350+1G&gt;A</b>	IVS12	3.1	-5.08	-263.6%	1
<b>c.3350+2C&gt;G</b>	IVS12	3.1	3.21	+3.44%	1
<b>c.3350+4A&gt;C</b>	IVS12	3.1	2.13	-6.56%	4
<b>c.3350+6T&gt;C</b>	IVS12	3.1	2.9	-31.4%	1

<sup>1</sup>Las variantes seleccionadas se encuentran marcadas en **rojo** <sup>2</sup>El símbolo [+] implica ganancia, mientras que [-] pérdida.

En total, se seleccionaron 67 variantes (en rojo en las Tablas 19, 20 y 21) para los estudios funcionales, debido a que podrían estar asociadas a un potencial efecto deletéreo sobre el *splicing*. Un total de 30 y 37 variantes alteraron los sitios 3'ss y 5'ss, respectivamente. La variante c.1685-3T>C, fue seleccionada a pesar del pequeño cambio en el valor de MES (-1,8%), porque afectaba a posiciones conservadas del 3'ss.

## **2.2. Estudio funcional de variantes ClinVar.**

### **2.2.1. Resultados del minigén mgPALB2\_ex1-3.**

Los minigenes híbridos utilizados para desarrollar este estudio se han generado de la forma y con la estructura detallada en el apartado **2.2** de Materiales y Métodos. Los ensayos funcionales se realizaron de la manera que ha sido descrita en el apartado **2** de Materiales y Métodos. La idea inicial era realizar ensayos de RNA de todas las variantes ClinVar seleccionadas y así generar un “catálogo funcional” de gran utilidad para su interpretación clínica. Desafortunadamente, esta etapa de la tesis doctoral estuvo marcada por la pandemia COVID-19 y únicamente se han obtenido resultados de las variantes ClinVar en el mgPALB2\_ex1-3, que, aun así, constituyen una muestra representativa de esta base de datos.

De las 151 variantes preseleccionadas, procedentes de la base de datos ClinVar, 31 están situadas en los exones 1, 2 y 3 y en los intrones adyacentes. Once variantes fueron seleccionadas para ser introducidas en el minigén mgPALB2\_ex1-3 mediante mutagénesis dirigida: **c.46A>G, c.48+1G>C, c.48+2T>G, c.48+1del, c.49-1del, 49-2del, c.49-1G>A, c.108+1\_108+2insC, c.109-2A>C, c.211+1G>T y c.211+2T>C.**

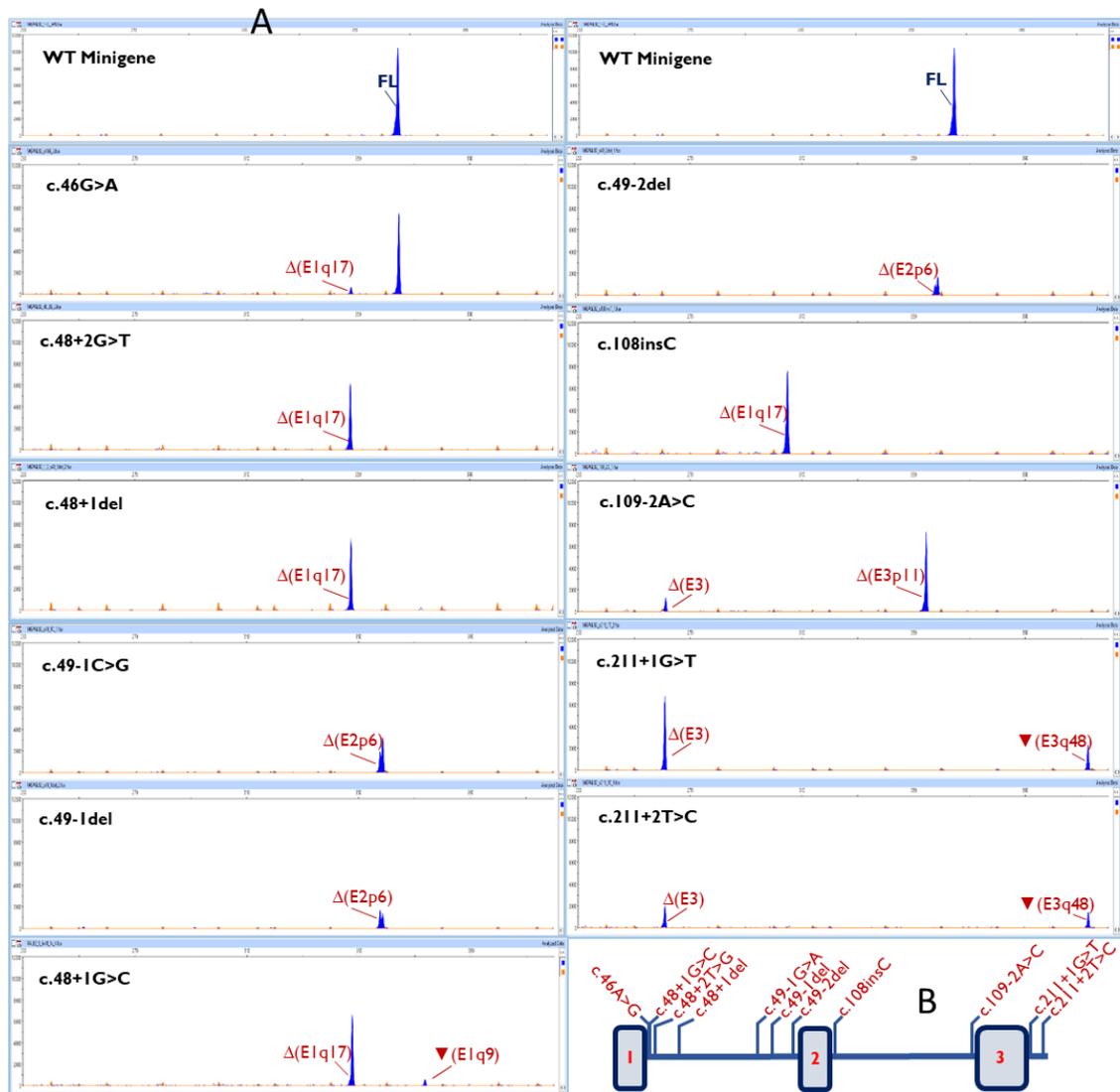
Las células MCF-7 fueron transfectadas con los minigenes mutantes, y los productos de la amplificación se obtuvieron con los oligonucleótidos (RTPSPL3-FW y RT-PSAD-RV-FAM). Finalmente, se comprobaron mediante las técnicas indicadas en apartados anteriores.

### 2.2.2. Resultados de las variantes ClinVar del exón 1 de *PALB2*.

La variante **c.46A>G** *missense* p.(Lys16Glu) es interpretada por ClinVar como benigna. Según el estudio informático, **c.46A>G**, altera el reconocimiento del sitio donador del exón 1 (Tabla 19). En los resultados del ensayo funcional se han podido observar los transcritos *full-length* en un 92,5%, y  $\Delta$ (E1q17), en un 7,5% (Tablas 22-23; Figura 40).

Las variantes, **c.48+1del** y **c.48+2T>G** están catalogadas en ClinVar como de sitio de *splicing* y han sido interpretadas como posiblemente patogénicas. Sin embargo, no existe evidencia funcional en ClinVar para estas variantes. El estudio bioinformático mostró que las 2 variantes eliminaban el sitio donador del exón 1 (Tabla 19). Los resultados de los ensayos funcionales revelaron un único transcrito anómalo, ( $\Delta$ (E1q17), 100%) (Tablas 22-23; Figura 40).

La variante **c.48+1G>C** está catalogada en ClinVar como de sitio de *splicing*, existiendo un conflicto de interpretación, ya que es considerada como benigna y posiblemente patogénica, aunque no existen evidencias funcionales acerca de esta variante. Según el estudio bioinformático, **c.48+1G>C**, causa la eliminación del sitio donador del exón 1 (Tabla 19). Los ensayos funcionales revelaron la generación de los transcritos aberrantes  $\Delta$ (E1q17) (90,8%) y  $\nabla$ (E1q9) (retención de 9 nt del intrón 1; 9,2%) (Tablas 22-23; Figura 40).



**Figura 40. Ensayos funcionales de *splicing* de variantes *PALB2* seleccionadas de la base de datos ClinVar. (A) Electroforesis capilar de fluorescencia de productos RT-PCR de los minigén *wild type* mgPALB2\_ex1-3 y mutantes. Se amplificaron con los oligonucleótidos FAM-RT-PSPL3-Fw y RT-pSAD-rv (Transcripción completa VI-ex1-ex2-ex3-V2 = 366 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ600). (B) Mapa de variantes ensayadas.**

**Tabla 22. Resultados de los estudios de *splicing* de variantes ClinVar de los exones 1 a 3 de *PALB2*.**

<b>Variantes (HGVS)<sup>1</sup></b>	<b>Resumen Bioinformático<sup>2</sup></b>	<b>Transcrito canónico</b>	<b>Transcritos-PTC<sup>3</sup></b>	<b>Transcritos <i>In-frame</i><sup>3</sup></b>
<i>Wild type</i> mgPALB2_ex1-3		100%		
c.46A>G p.(Lys16Glu)	[-]5'ss	92.5%±0.1%	Δ(E1q17) [7.5%±0.1%]	
<b>c.48+1G&gt;C</b>	[↓]5'ss	-	Δ(E1q17) [90.8%±0.6%]	▼(E1q9) [9.2%±0.6%]
<b>c.48+2T&gt;G</b>	[-]5'ss (5.74→-2.43)	-	Δ(E1q17) [100%]	
<b>c.48+1del</b>	[-]5'ss (5.74→-2)	-	Δ(E1q17) [100%]	
<b>c.49-1del</b>	[-]5'ss (5.74→2.88)	-		Δ(E2p6) [100%]
<b>c.49-2del</b>	5'ss (5.74→5.74) [+]5'ss (6.49), 9-nt <i>downstream</i>	-		Δ(E2p6) [100%]
<b>c.49-1G&gt;A</b>	[-]3'ss (9.28→0.92) [+]3'ss (8.49), 6-nt <i>downstream</i>	-		Δ(E2p6) [100%]
<b>c.108+1_108+2 insC</b>	[-]5'ss(10.86→2.68)	-		Δ(E2) [100%]
<b>c.109-2A&gt;C</b>	[-]5'ss(10.86→3.1)	-	Δ(E3p11) [85%±0.5%] Δ(E3) [15%±0.5%]	
<b>c.211+1G&gt;T</b>	[↓]3'ss(10.06→7.27)	-	Δ(E3) [73.3% ±0.6%]	▼(E3q48)[26.7%±0.6%]
<b>c.211+2T&gt;C</b>	[-]3'ss (10.06→2.11)	-	Δ(E3) [48.1% ±7.4%]	▼(E3q48) [51.9% ±7.4%]

<sup>1</sup> Fuente en negrita: sin rastros o <5% del transcrito *full-length*. <sup>2</sup> [-] Eliminación del sitio de *splicing*; [+] Nuevo sitio. [↓] Reducción de la puntuación MES. <sup>3</sup> PTC: codón de terminación prematura; <sup>3</sup> Δ, pérdida de secuencias exónicas; ▼ inclusión de secuencias intrónicas; E (exón), p (cambio de aceptor), q (cambio de donador). Cuando es necesario, se indica el número exacto de nt insertados o eliminados. Por ejemplo, la transcripción ▼ (E1q9) denota el uso de un sitio donador alternativo que se localiza 9 nucleótidos *downstream* del exón 1, provocando la adición de 9-nt al mRNA maduro.

**Tabla 23. Anotación de los transcritos de acuerdo con la secuencia de *PALB2* NM\_024675.4.**

Transcritos <sup>1</sup>	RNA-HGVS	PROTEINA-HGVS	Diagramas de eventos de <i>splicing</i> <sup>2</sup>
$\Delta$ (E1q17)	r.32_48del	p.Cys11Phe*26	
$\nabla$ (E1q9)	r.48_49ins48+1_48+9	p.Lys16_Leu17insVP G	
$\Delta$ (E2p6)	r.49_54del	p.Leu17_Lys18del	
$\Delta$ (E2)	r.49_108del	p.Leu17_Asn36del	
$\Delta$ (E3p11)	r.109_117del	p.Arg37Serfs*2	
$\nabla$ (E3q48)	r.211_212ins211+1_211+48	p.Ser70_Glu71ins16	
$\Delta$ (E3)	r.109_211del	p.Arg37Asnfs*106	

<sup>1</sup> Los eventos de *splicing* alternativo fisiológico, están subrayados. <sup>2</sup> Los exones y los eventos anómalos de *splicing* se indican en rojo. Los sitios de *splicing de novo* / sitios crípticos AG o GT se muestran en rojo.

### 2.2.3. Resultados de las variantes ClinVar del exón 2 de *PALB2*.

Las variantes, **c.49-1del**, **c.49-1G>A** y **c.49-2del** han sido catalogadas en la base de datos ClinVar como variantes de sitio de *splicing* y están interpretadas clínicamente como patogénica, posiblemente patogénica y variante de significado clínico desconocido, respectivamente. Según el estudio bioinformático, las 3 variantes eliminan el sitio aceptor del exón 2. Además, **c.49-1G>A** y **c.49-2del** crean un sitio de *splicing de novo* 6 nucleótidos *downstream* del natural (Tabla 19). Los ensayos funcionales de estas tres variantes en experimentos independientes, determinaron que se producía un único transcrito aberrante ( $\Delta$ (E2p6), 100%), que no incluía los 6 primeros nucleótidos del exón 1, debido al uso del sitio aceptor alternativo (Tablas 22-23; Figura 40).

La variante **c.108+1\_108+2insC**, está registrada como variante de sitio de *splicing* en ClinVar, y ha sido interpretada como posiblemente patogénica. Sin embargo, no existen ensayos

funcionales registrados, que puedan confirmar esta premisa. Los estudios bioinformáticos previos mostraron que **c.108+1\_108+2insC**, eliminaba el sitio donador del exón 2 de *PALB2* (Tabla 19). Los resultados de los ensayos funcionales de esta variante revelaron  $\Delta(E2)$  (delección del exón 2) como único transcrito (Tablas 22-23; Figura 40).

#### **2.2.4. Resultados de las variantes ClinVar del exón 3 de *PALB2*.**

La variante **c.109-2A>C**, está registrada como variante de sitio de *splicing* en ClinVar, y ha sido interpretada como posiblemente patogénica, aunque no existen evidencias funcionales al respecto. Según los estudios bioinformáticos, esta variante elimina el sitio aceptor del exón 3 (Tabla 19). Los ensayos funcionales de esta variante revelaron los transcritos aberrantes  $\Delta(E3)$  (delección del exón 3; 15%) y  $\Delta(E3p11)$  (uso de un aceptor alternativo situado 11 nucleótidos *downstream* del natural; 85%) (Tablas 22-23; Figura 40).

Las variantes **c.211+1G>T** y **c.211+2T>C**, están clasificadas en la base de datos ClinVar como variantes de sitio de *splicing*. La variante **c.211+1G>T**, ha sido interpretada como patogénica o posiblemente patogénica y **c.211+2T>C**, como posiblemente patogénica. Sin embargo, no existen evidencias funcionales que puedan confirmar estas interpretaciones. Según el estudio bioinformático estas variantes eliminan el sitio donador del exón 3 (Tabla 19). La variante **c.211+1G>T** generó los transcritos  $\Delta(E3)$  (pérdida del exón 3; 73,3%) y  $\blacktriangledown$  (E3q48) (retención de 48 nt del intrón 3; 26,7%). Por otro lado, **c.211+2T>C** generó los mismos transcritos en diferentes proporciones ( $\Delta(E3)$ , 48,1%) y ( $\blacktriangledown$  (E3q48), 51,9 %) obviando el uso del donador GC atípico generado a causa de la variante (Tablas 22-23; Figura 40).

#### **2.2.5. Recapitulación de los resultados en el minigén mgPALB2\_ex1-3 de las variantes ClinVar.**

De las 31 variantes de los exones 1 a 3 del gen *PALB2* reportadas en la base de datos ClinVar, 11 fueron seleccionadas bioinformáticamente e introducidas en el minigén mgPALB2\_ex1-3 por mutagénesis dirigida y ensayadas funcionalmente en células MCF-7. Todas

las variantes ensayadas alteraban el *splicing*, de las cuales 10 mostraron una falta completa del transcrito *full-length*. Diez variantes afectan a las posiciones clásicas  $\pm 1,2$  del 5' ss o 3' ss y ninguna de ellas produce trazas del transcrito canónico.

### 2.2.6. Análisis de transcritos producidos por el minigén mgPALB2\_ex1-3.

El análisis de fragmentos fluorescentes reveló la existencia de 8 transcritos diferentes, incluido el transcrito WT, donde el uso de sitios alternativos fue el mecanismo de *splicing* más frecuente:  $\Delta(E1q17)$ ,  $\nabla(E1q9)$ ,  $\Delta(E2p6)$ ,  $\Delta(E3p11)$  y  $\nabla(E1q48)$ . Además, observamos 2 eventos de exón *skipping*:  $\Delta(E2)$  y  $\Delta(E3)$ . También, es importante destacar, que de los 7 transcritos producidos a causa de estas variantes, 4 mantienen la pauta de lectura:  $\nabla(E1q9)$ ,  $\Delta(E2)$ ,  $\Delta(E2p6)$  y  $\nabla(E1q48)$ . Los demás eventos de *splicing* producidos a causa de estas variantes, alteran la pauta de lectura, introduciendo codones prematuros de terminación, y probablemente estén asociados a la generación de una proteína disfuncional. Por último, 4 de las isoformas mencionadas anteriormente han sido previamente caracterizadas como eventos naturales de *splicing*:  $\nabla(E1q9)$ ,  $\Delta(E1q17)$ ,  $\Delta(E2p6)$  y  $\Delta(E2)$  (Lopez-Perolio et al., 2019).

---

### 3. Mapeo funcional de elementos reguladores del *splicing* de los exones 11 y 12 de PALB2.

El exón 12 tiene un donador atípico GC (MES = 3,1) que aparece en aproximadamente el 1% de los intrones de mamíferos. El cambio en la posición crítica +2 del sitio donador añade complejidad al reconocimiento de dicho exón a través de un mecanismo poco conocido. Los exones 11 y 12 del gen *PALB2* sufren *splicing* alternativo fisiológico, habiéndose caracterizado previamente hasta 8 transcritos diferentes:  $\nabla(E11p23)$ ,  $\Delta(E11p2)$ ,  $\Delta(E11)$ ,  $\Delta(E11)+\nabla(E12p446)$ ,  $\nabla(E12p65)$ ,  $\Delta(E12p136)$ ,  $\Delta(E12)$  y  $\Delta(E11\_E12)$ . Por lo tanto, y con el objetivo de localizar SREs, que presumiblemente regulan su reconocimiento, se procedió con una estrategia de mapeo funcional. Para ello, se introdujeron 8 microdeleciones independientes de aproximadamente 30

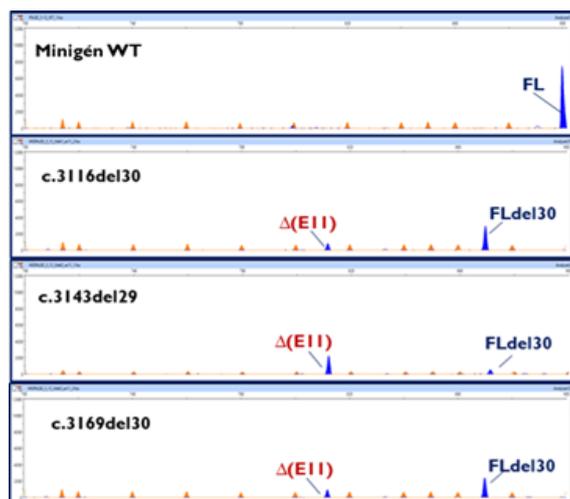
nucleótidos, en los exones 11 y 12 del minigén mgPALB2\_ex5-12. Estas microdeleciones cubrían toda la secuencia codificante de ambos exones, a excepción de los dos primeros y los tres últimos nucleótidos de cada exón, debido a que forman parte de las secuencias consenso de los sitios aceptor y donador, respectivamente.

### 3.1. Estudio de microdeleciones del exón 11.

Se introdujeron tres microdeleciones solapantes de 29-30 nucleótidos en el exón 11 para su posterior ensayo funcional en células MCF-7: **c.3116del30**, **c.3143del29** y **c.3169del30**. La microdelección **c.3116del30** es la más próxima al sitio aceptor del exón 11. Los resultados del ensayo funcional el minigén-c.3116del30, mostraron la producción de  $\Delta(E11)$ , en un 23,7% y transcrito *full-length* en un 76,3% (Figura 41; Tabla 24). La microdelección **c.3143del29**, se encuentra en la mitad del exón 11 solapando con las otras dos deleciones introducidas en este exón. Los ensayos funcionales de esta microdelección revelaron que un alto porcentaje de los transcritos no habían incorporado el exón 11, ( $\Delta(E11)$ , 76,2%) y que el resto de ellos se correspondían con el transcrito canónico (Figura 41; Tabla 24). Por último, la microdelección **c.3169del30** es la más próxima al donador del exón 11 de *PALB2*. Los resultados funcionales mostraron la producción de  $\Delta(E11)$  en un 26,3% y de transcritos *full-length* en un 73,7% (Figura 41; Tabla 24).

**Tabla 24. Eventos de *splicing* producidos a causa de las microdeleciones seleccionadas en los exones 11 y 12 de *PALB2*.**

Microdeleciones	Transcritos <i>full-length</i>	Transcritos-PTC
<b>mgPALB2_ex5-12</b>	100%	
<b>Exón 11</b>		
<b>c.3116del30</b>	76.3%±0.2%	$\Delta(E11)$ [23.7%±0.2%]
<b>c.3143del29</b>	23.8%±0.1%	$\Delta(E11)$ [76.2%±0.1%]
<b>c.3169del30</b>	73.7±0.2%	$\Delta(E11)$ [26.3%±0.2%]
<b>Exón 12</b>		
c.3204del30	100%	
<b>c.3229del30</b>	46.2%±2.5%	$\Delta(E12)$ [53.8%±2.5%]
<b>c.3254del44</b>	92.5%±0.4%	$\Delta(E12)$ [7.5%±0.4%]
<b>c.3293del30</b>	43.5%±0.9%	$\Delta(E12)$ [56.5%±0.9%]
c.3318del30	100%	



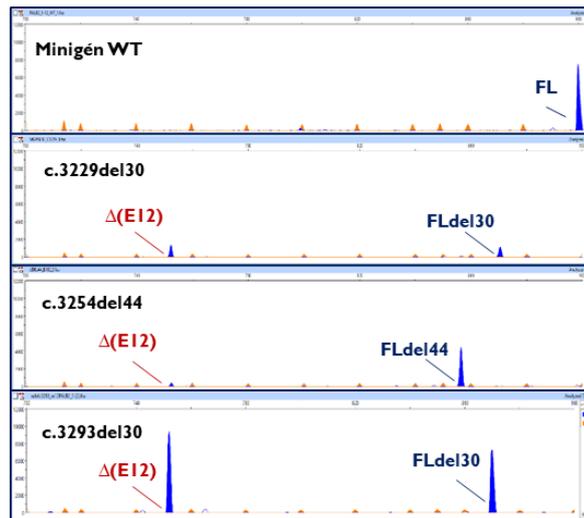
**Figura 41. Electroforesis capilar de las microdeleciones c.3116del30, c.3143del29 y c.3169del30 en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

### 3.2. Estudio de microdeleciones del exón 12.

Se introdujeron cinco microdeleciones solapantes de 30 a 44 nucleótidos en el exón 12 para su posterior ensayo funcional en células MCF-7: **c.3204del30**, **c.3229del30**, **c.3254del44**, **c.3293del30** y **c.3318del30**. Estas microdeleciones cubrían toda la secuencia codificante de este exón, a excepción de los dos primeros y los tres últimos nucleótidos de cada exón, como hemos visto anteriormente.

Dos microdeleciones, **c.3229del30** y **c.3293del30**, tuvieron impacto moderado en el *splicing*, observando diferentes niveles de exón 12 *skipping* (53,8% y 56,5% de  $\Delta(E12)$ , respectivamente), y sugiriendo que estos intervalos pueden contener motivos reguladores (Figura 42; Tabla 24).

Por otro lado, los resultados obtenidos con la microdelección **c.3254del44** mostraron un impacto débil en el *splicing* ya que únicamente el 8% de los transcritos generados no incluían el exón 12 de *PALB2* mientras que el transcrito canónico era mayoritario (92%) (Figura 42; Tabla 24).



**Figura 42. Electroforesis capilar de las microdeleciones c.3229del30, c.3254del44, y c.3293del30 en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

No todas las microdeleciones ensayadas en este exón tuvieron efectos sobre el proceso de *splicing*. Los resultados de los ensayos funcionales realizados con los minigenes mutantes portadores de las microdeleciones **c.3204del30** y **c.3318del30** revelaron únicamente la presencia de transcritos *full-length*.

### 3.3. Selección de variantes ClinVar de los exones 11 y 12 de *PALB2* situadas en las microdeleciones **c.3143del29**, **c.3229del30** y **c.3293del30**.

Los intervalos positivos **c.3143del29**, **c.3229del30** y **c.3293del30** son candidatos a contener variantes de SREs que afecten al *splicing*. Como prueba de concepto, se decidió ensayar funcionalmente variantes reportadas por ClinVar localizadas en estos intervalos (total = 84 variantes), utilizando el programa HEXplorer como herramienta de selección. Para ello, se seleccionaron al azar tres variantes de cada intervalo entre aquellas con los valores HEXplorer más negativos (Punto de corte < -40; 16 variantes ClinVar sobrepasaron el umbral, datos no mostrados). Se seleccionaron las siguientes:

- Intervalo c.3143-c.3171: c.3155A>T (HEXplorer = -95,54), c.3157G>T (-106,87) y c.3161C>T (-87,39).

- Intervalo c.3229-c.3258: c.3242A>G (-67,77), c.3244A>G (-65,81) y c.3251C>T (-45,61).
- Intervalo c.3293-c.3322: c.3294\_3298del (-63,22), c.3306C>G (-143,8) y c.3306C>T (-53,46).

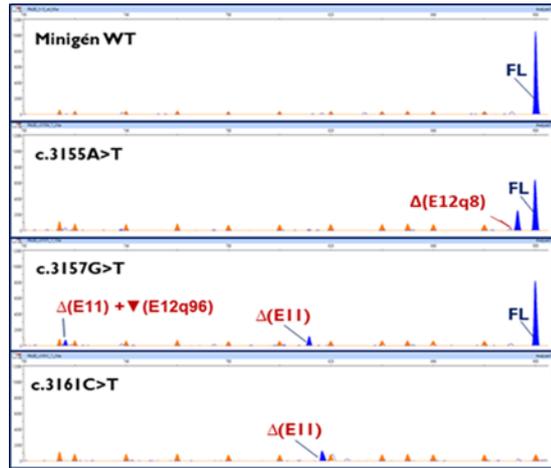
Todas las variantes fueron estudiadas cómo se ha indicado previamente en otros apartados.

### 3.3.1.1. Estudio de variantes exónicas ClinVar del intervalo c.3143del30.

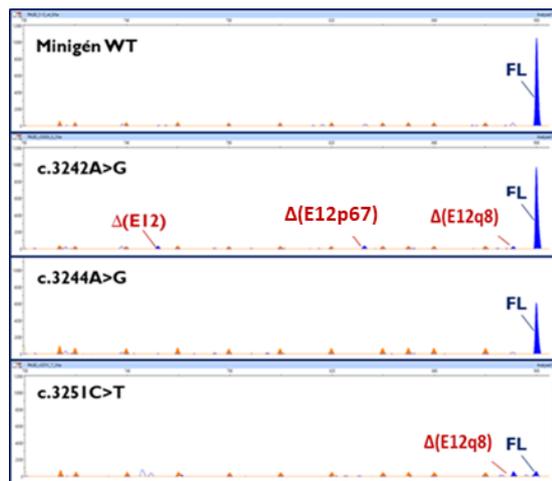
Las variantes *missense* **c.3155A>T** p.(Asp1052Val), **c.3157G>T** p.(Asp1053Tyr) y **c.3161C>T** p.(Ser1054Phe), están catalogadas en ClinVar como variantes de significado clínico desconocido. Los resultados observados en los ensayos de *splicing* revelaron alteraciones en este proceso. La variante c.3155A>T generaba un 73,6% de transcritos *full-length*, y  $\Delta$ (E12q8) en un 26,4% (Tabla 25, Figura 43). Además, c.3157G>T producía transcritos *full-length* en un 82,4% así como  $\Delta$ (E11) +  $\blacktriangledown$ (E12q96) (6,3%) y  $\Delta$ (E11) (11,3%) (Tabla 25, Figura 43). Por último, los ensayos de *splicing* de la variante c.3161C>T revelaron una alteración completa del *splicing* con la producción única de  $\Delta$ (E11) (Tabla 25, Figura 43).

### 3.3.1.2. Estudio de variantes exónicas ClinVar del intervalo c.3229del29.

Las variantes *missense* **c.3242A>G** p.(Glu1081Gly) y **c.3244A>G** p.(Glu1081Gly), y la variante sinónima **c.3251C>T** p.(Ser1087=) están catalogadas en ClinVar como VUS, VUS y posiblemente benigna, respectivamente. Los ensayos funcionales con la variante c.3242A>G revelaron la presencia mayoritaria del transcrito canónico (91,6%) y minoritariamente la presencia de transcritos anómalos en diferentes proporciones: ( $\Delta$ (E12), 2,6%), ( $\Delta$ (E12q8), 2,4%) y ( $\Delta$ (E12p67), 3,4%) (Tabla 25; Figura 44). Por otro lado, los resultados de la variante c.3251C>T determinaron la presencia mayoritaria del transcrito canónico (52,9%) y de  $\Delta$ (E12q8) (47,1%; exclusión de los últimos 8 nucleótidos del exón 12) (Tabla 25; Figura 44). Los resultados de c.3244A>G pusieron de manifiesto la presencia única de transcritos *full-length* (Tabla 25; Figura 44).



**Figura 43. Electroforesis capilar de las variantes c.3155A>T, c.3157G>T y c.3161C>T en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).



**Figura 44. Electroforesis capilar de las variantes c.3242A>G, c.3244A>G y c.3251C>T en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

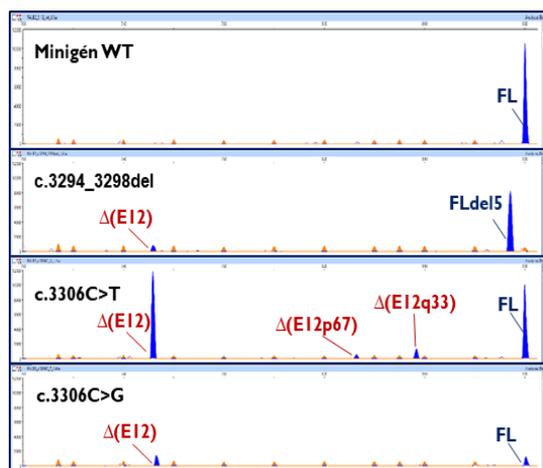
**Tabla 25. Eventos de *splicing* producidos por variantes exónicas ClinVar y situadas en nuestras deleciones positivas.**

Variantes <sup>1</sup>	HEXplorer	Full-length	Transcritos-PTC <sup>2-3</sup>	Transcritos In-Frame <sup>2</sup>
Minigén <i>Wild type</i>		100%		
<b>Exón 11</b>				
c.3155A>T p.(Asp1052Val)	-95.54	73.6%±0.8%	Δ(E12q8) [26.4%±0.8%]	
c.3157G>T p.(Asp1053Tyr)	-106.87	82.4%±1.0%	Δ(E11) [11.3%±0.4%] Δ(E11)+▼(E12q96) [6.3%±0.6%]	
c.3161C>T p.(Ser1054Phe)	-87.39	-	Δ(E11) [100%]	
<b>Exón12</b>				
c.3242A>G p.(Glu1081Gly)	-67.77	91.6±0.3	Δ(E12) [2.6%±0,5%] Δ(E12q8) [2.4%±0,3%] Δ(E12p67) [3.4%]	
c.3244A>G p.(Ser1082Gly)	-65.81	100%		
c.3251C>T p.(Ser1084Leu)	-45.61	52.9%±0.5%	Δ(E12q8) [47.1±0.5%]	
c.3294_3298del p.(Lys1098fs)	-63.22	94.4±2.8%	Δ(E12) [5.6%±2.8%]	
c.3306C>G p.(Ser1102Arg)	-143.8	35.6±1.1%	Δ(E12) [57.6%±3.0%] Δ(E12p67) [2.1±0.1%]	Δ(E12q33) [5.1±0.3%]
c.3306C>T p.(Ser1102=)	-53.46	44.5%±2.3%	Δ(E12) [55.5%±2.3%]	

<sup>1</sup> Fuente en negrita: sin rastros o <5% de la transcripción completa. <sup>2</sup> PTC: codón de terminación prematura; <sup>3</sup> Δ, pérdida de secuencias exónicas; ▼ inclusión de secuencias intrónicas; E (exón), p (cambio de aceptor), q (cambio de donador). Cuando es necesario, se indica el número exacto de nt insertados o eliminados. Por ejemplo, el transcrito Δ(E12q33) denota el uso de un sitio donador alternativo que se localiza 33 nucleótidos *upstream* del donador del exón 12, provocando la eliminación de 33-nt en el mRNA maduro.

### 3.3.1.3. Estudio de variantes exónicas ClinVar del intervalo c.3293del30.

La variante *frameshift* c.3294\_3298del p.(Lys1098fs) está catalogada en ClinVar como patogénica. Esta variante indujo la isoforma **canónica** de forma mayoritaria (94,4%) y la pérdida del exon 12 (Δ(E12), 5,6%) (Tabla 25; Figura 45). La variante sinónima c.3306C>T p.(Ser1102=) y la *missense* c.3306C>G p.(Ser1102Arg), están catalogadas en ClinVar como posiblemente benigna y VUS, respectivamente. La variante c.3306C<G generaba un 35,6% de transcritos *full-length* (Tabla 25; Figura 45), Δ(E12) en un 57,2%, Δ(E12q33) (uso de un donador alternativo situado 33 nucleótidos *upstream*) en un 5,1% y Δ(E12p67) (uso de un aceptor alternativo situado 67 nucleótidos *downstream*) en un 2,1%. Finalmente, los resultados de la variante c.3306C>T revelaron la presencia de Δ (E12), 55,5%) y del transcrito canónico (44,5%) (Tabla 25; Figura 45).



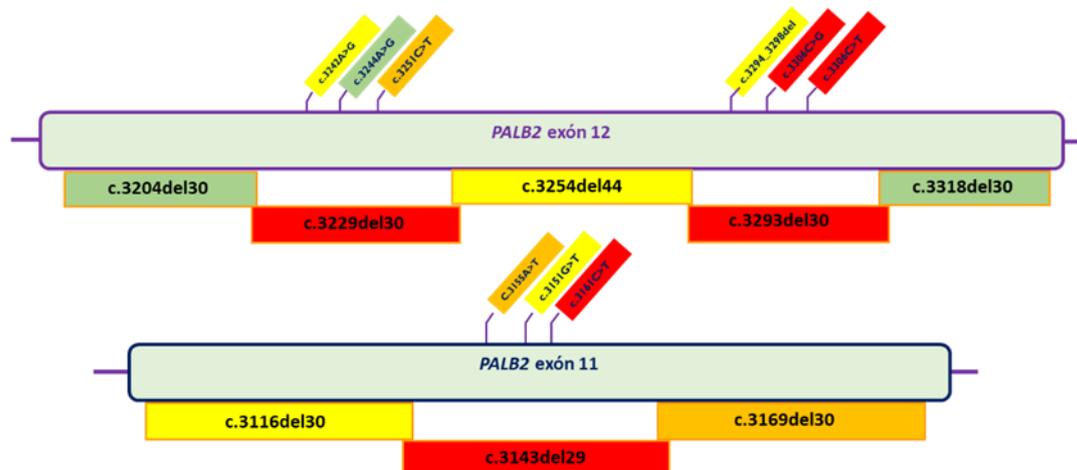
**Figura 45. Electroforesis capilar de las variantes c.3294\_3298del, c.3306C>T y c.3306C>G en el minigén mgPALB2\_ex5-12.** Se amplificaron con los oligonucleótidos RT-PALB2\_ex6-fw y RT-pSAD-rv-FAM (Tamaño = 906 nt). Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

#### 3.3.1.4. Recapitulación de resultados de las microdeleciones y de las variantes ClinVar de los exones 11 y 12 de *PALB2*.

Se han estudiado 8 microdeleciones de 29-44 pares de bases que cubrían toda la secuencia de los exones 11 y 12 del gen *PALB2* en el minigén mgPALB2\_ex5-12 en células MCF-7. Seis de estas microdeleciones alteraban los patrones de *splicing* (c.3116del30, c.3143del29, c.3169del30, c.3229del30, c.3254del44 y c.3293del30), con un mayor impacto por parte de c.3143del29, c.3229del30 y c.3293del30

Además, se ensayaron funcionalmente 9 variantes exónicas, registradas en la base de datos ClinVar y situadas en los intervalos donde las microdeleciones habían producido alteraciones notables en el *splicing*. Interesantemente, 8 de las 9 variantes ensayadas procedentes de estos intervalos tuvieron impacto en el *splicing*, destacando el efecto de la variante c.3161C>T que sólo produjo el transcrito aberrante  $\Delta(E11)$  (con codón de parada prematuro) sin ninguna traza del transcrito *full-length* (Tabla 25, Figura 43-46). Estos resultados avalan la estrategia utilizada. Las variantes con impacto en el *splicing* corresponden a 5 *missense*, 2 sinónimas, y 1 *frameshift*, indicando que cualquier variante puede alterar el *splicing*. Finalmente, no existe una clara

correlación entre los valores Hexplorer y el grado de alteración del *splicing*, medido como tanto por ciento de transcritos aberrantes.



**Figura 46. Mapa de microdeleciones del exón 11 Y 12 de *PALB2* y variantes exónicas ensayadas.** Cajas: rojo (>50% skipping), naranja (25-50%), amarillo (<25%) y verde (0%).

### 3.3.1.5. Análisis de transcritos producidos por las microdeleciones y las variantes exónicas.

El análisis de fragmentos fluorescentes obtenido de los ensayos funcionales desarrollados con las microdeleciones, reveló la existencia de 3 transcritos diferentes, incluyendo el transcrito *wild type*, donde el *exón skipping* fue el único evento:  $\Delta(E11)$  y  $\Delta(E12)$  (Tabla 24). Además, todos los eventos de *splicing* producidos a causa de estas microdeleciones introducen codones de terminación prematuros, alterando la pauta de lectura y propiciando probablemente la producción de proteínas disfuncionales.

Por otro lado, el estudio de electroforesis capilar de los ensayos funcionales desarrollados con las variantes exónicas mostró la existencia de 6 transcritos diferentes, donde encontramos dos eventos de *exón skipping* ( $\Delta(E11)$  y  $\Delta(E12)$ ) que fueron mayoritarios, y 4 transcritos debidos al uso de un sitio alternativo:  $\Delta(E12q8)$ ,  $\Delta(E11)+\blacktriangledown(E12q96)$ ,  $\Delta(E12p67)$  y  $\Delta(E12q33)$  (Tabla 25). Uno de ellos,  $\Delta(E12q8)$ , se generó por el uso de un donador GC atípico. Cabe destacar que todos los eventos de *splicing* producidos a causa de estas variantes, a excepción de  $\Delta(E12q33)$ , introducen

codones de parada. prematuro, alterando la pauta de lectura y propiciando probablemente la generación de una proteína disfuncional.

# DISCUSIÓN

Durante más de dos décadas el diagnóstico genético del cáncer de mama ha funcionado en un entorno, donde la identificación de variantes patogénicas como causa hereditaria de susceptibilidad a cáncer de mama sucedía en una pequeña proporción de las mujeres que buscaban asesoramiento genético. En esta etapa, *BRCA1* y *BRCA2*, han sido los únicos genes que han proporcionado una información útil para el asesoramiento genético de familias sospechosas de padecer este síndrome y para el manejo clínico de pacientes y familias. Sin embargo, el desarrollo de la tecnología de secuenciación masiva (NGS) está permitiendo mejorar el conocimiento existente y el diagnóstico molecular de las enfermedades hereditarias.

Hoy en día sabemos que el espectro de predisposición genética del cáncer de mama hereditario es muy complejo y se ha observado como variantes de pérdida de función en al menos 8 genes (*BRCA1*, *BRCA2*, *PALB2*, *BARD1*, *RAD51C*, *RAD51D*, *ATM* y *CHEK2*) muestran una asociación significativa con un aumento en el riesgo de padecer cáncer de mama en la población general, como se ha reportado recientemente en varios estudios de secuenciación a gran escala (Yang et al., 2020; Dorling et al., 2021; Hu et al., 2021; Narod, 2021). De acuerdo a los estudios citados, *BRCA1*, *BRCA2*, *PALB2*, *ATM* y *CHEK2* constituyen el núcleo central de la susceptibilidad hereditaria a cáncer de mama, a los cuales, según algunos autores, habría que unir *TP53*, aunque la frecuencia de variantes patogénicas es muy baja en este gen (Foulkes, 2021).

De hecho, estos estudios, dentro de los cuales se incluye el proyecto BRIDGES, en el cual está inmerso nuestro laboratorio, han establecido firmemente a *PALB2* como un gen de susceptibilidad de alto riesgo a cáncer de mama, confiriéndole un riesgo absoluto superior al 30%, de modo que las variantes patogénicas en *PALB2* son clínicamente accionables. Así, las familias portadoras de variantes patogénicas germinales en *PALB2* pueden favorecerse del programa de prevención de cáncer de mama, con acciones propuestas que van desde la vigilancia intensificada hasta la cirugía profiláctica (Tischkowitz et al., 2021).

Además, las alteraciones genéticas en la línea germinal de *PALB2*, que son relativamente prevalentes, representan cerca del 10% de todas las variantes patogénicas encontradas en los ocho genes mencionados anteriormente y representa el 0,5% de todos los casos de cáncer de mama (422

variantes patogénicas en 87158 pacientes), lo que subraya el papel de este gen en el cáncer de mama hereditario (Dorling et al., 2021; Hu et al., 2021). Sin embargo, es importante destacar que una cantidad sustancial de los pacientes con cáncer de mama porta una variante de significado clínico desconocido en alguno de estos genes de susceptibilidad (1,65% en *PALB2*) (Hu et al., 2021), lo que plantea un desafío en el asesoramiento genético. Generalmente, son consideradas patogénicas todas las variantes que generan codones de parada prematuros, como variantes *nonsense* o *frameshift* y aquellas que afectan a aminoácidos esenciales de la proteína. Esta interpretación, pasa por alto aquellas variantes que podrían alterar mecanismos esenciales de la expresión génica, como la transcripción y el *splicing*, entre otros. En este contexto, los estudios funcionales de variantes de *splicing* aportan información relevante para su interpretación clínica y, por tanto, para el manejo tanto de pacientes como de portadores asintomáticos.

El *splicing* aberrante es un mecanismo frecuente de inactivación génica y está asociado a variantes de la línea germinal en genes de susceptibilidad a cáncer de mama. Así, por ejemplo, el grupo de *Splicing* y Susceptibilidad Genética a Cáncer del IBGM, donde se ha realizado esta tesis doctoral, ha realizado más de 300 estudios funcionales de variantes del gen *BRCA2*, de las cuales aproximadamente el 50% alteraron el *splicing* (Sanz et al., 2010; Acedo et al., 2012, 2015, Fraile-Bethencourt et al., 2017, 2018, 2019). Más aún, la participación en el proyecto BRIDGES como parte del *Work Package 4* “Análisis Funcional de variantes” ha posibilitado la realización de estudios sistemáticos de *splicing* de los principales genes implicados en cáncer de mama hereditario. Hasta la fecha ya están publicados los relativos a los genes *RAD51C*, *RAD51D* y (Fraile-Bethencourt et al., 2019; Sanoguera-Miralles et al., 2020; Bueno-Martínez et al., 2021). En *RAD51C* se han analizado 20 variantes candidatas en las fronteras intrón-exón, de las que 19 (95%) indujeron un *splicing* aberrante. No sólo eso, las variantes *esplíceogénicas* de *RAD51C* constituyen el ~45% de todas las variantes patogénicas detectadas en pacientes BRIDGES. Esta frecuencia triplica la inicialmente estimada (~15%) en enfermedades hereditarias (Krawczak et al., 1992), demostrando la importancia de las alteraciones de *splicing* como mecanismo primario de patogenicidad. Por otro lado, en *RAD51D* se encontró que más del 75% de las variantes afectaban

al *splicing*, y se demostró que el exón 3 está regulado por SREs, convirtiéndolo en un punto caliente de variantes que afectan a estos elementos y, consecuentemente, podrían inducir anomalías en el *splicing*. Otros estudios funcionales de variantes BRIDGES en los principales genes *CHEK2* y *ATM* serán publicados en breve. De este modo, se está generando un catálogo funcional de variantes que puede ser consultado por cualquier genetista para la interpretación clínica de variantes y la elaboración de un consejo genético.

En esta tesis doctoral se ha realizado el mayor estudio funcional de *splicing* de *PALB2* reportado hasta la fecha. Nos hemos centrado en el análisis funcional de 82 variantes de sitios de *splicing*, detectadas en la cohorte BRIDGES y 151 variantes reportadas en la base de datos ClinVar situadas en los límites exón/intrón. Además, se han mapeado los elementos reguladores de los exones 11 y 12, mediante el testado de 8 microdeleciones, que cubren prácticamente toda la secuencia de los exones 11 y 12 de *PALB2*. Por último, y como prueba de concepto, se decidió ensayar funcionalmente 9 variantes reportadas por ClinVar localizadas en los intervalos críticos determinados por aquellas microdeleciones con impacto en el *splicing*. Todas las variantes y microdeleciones se introdujeron en los 3 minigenes clonados en el transcurso de esta tesis, que incluían de los exones 1 a 12 de *PALB2* para su posterior ensayo funcional. Los minigenes mgPALB2\_ex1-3, mgPALB2\_EX4-6 y mgPALB2\_ex5-12 son herramientas simples, robustas y de alta sensibilidad; características que, en ausencia de muestra de RNA de paciente, respaldan su idoneidad para la caracterización preliminar de posibles variantes de *splicing*. Los minigenes presentan importantes ventajas sobre otros ensayos de RNA: i) análisis aislado del efecto de un alelo sin la interferencia del alelo *wild type* que ocurriría en muestras de portadoras; ii) identificación y cuantificación de todos los transcritos producidos por la variante; así, el uso de inhibidores del NMD (cicloheximida) impide la degradación de transcritos con PTC permitiendo una cuantificación precisa y la detección de transcritos raros; iii) estudio de múltiples variantes con una o pocas construcciones, ya que cada una de ellas puede ser introducida mediante mutagénesis dirigida sin necesidad de muestras de pacientes; iv) ensayos en una línea celular relevante para la enfermedad en estudio; v) alta reproducibilidad de los patrones de *splicing*, ya que en la gran

mayoría de las variantes analizadas los minigenes replican los resultados obtenidos en muestras de pacientes. A todo ello hay que añadir la alta sensibilidad y resolución de la electroforesis capilar fluorescente que facilita tanto la detección de transcritos raros como la resolución de transcritos que difieren en unos pocos nucleótidos.

## 1. Variantes BRIDGES.

En este trabajo, hemos realizado uno de los estudios más completos de la correlación entre el *splicing* aberrante y el cáncer de mama, mediante el cual evaluamos el impacto del *splicing* de 44 variantes de DNA de los exones 1 a 12 de *PALB2* con los minigenes mgPALB2\_ex1-3, mgPALB2\_ex4-6 y mgPALB2\_ex5-12. Sorprendentemente, una gran proporción de variantes BRIDGES (37/44, 84,1%) alteraron el *splicing*, respaldando la sensibilidad de las predictores *in silico* de variantes de *splicing*. Además, 25 de estas variantes no mostraron ningún rastro, o porcentajes mínimos del transcrito *full-length*, apoyando su interpretación como alelos de alto riesgo para cáncer de mama. Asimismo, la alta sensibilidad y resolución del análisis fluorescente de fragmentos, nos permitió detectar 36 transcritos aberrantes (23 transcritos PTC, 10 transcritos *in-frame* y 3 isoformas no caracterizadas). Por último, destacar que la sensibilidad de este método nos ha permitido caracterizar transcritos con deleciones de tan solo 2 nucleótidos, como  $\Delta(E10p2)$ .

En general, los resultados de *splicing* de estas variantes o de otras que afectan a los mismos sitios de *splicing* respaldan la reproducibilidad de los minigenes comparándolos con los análisis de RNA de paciente, descritos anteriormente (Catucci et al., 2014; Lopez-Perolio et al., 2019; Landrith et al., 2020; Ryu et al., 2020). Por tanto, c.48G>A, c.49-2A>T, c.2748+1G>T (frente a c.2748+2T>G), c.2834+2T>C, c.3113+5G>C y c.3350+5G>A arrojaron resultados muy similares o idénticos. Sin embargo, se observaron algunas discrepancias entre los ensayos de RNA de pacientes y los ensayos con minigenes, o incluso entre diferentes ensayos de una misma variante a partir de RNA de pacientes, como, por ejemplo, c.3113G>A y c.3113+5G>C (Reid et al., 2007; Casadei et al., 2012, 2019; Lopez-Perolio et al., 2019; Landrith et al., 2020). Esto puede deberse a la complejidad de los patrones *splicing* alternativo que se produce en ciertas regiones genómicas como se describe

para la variante *BRCA2* c.7976+5G>T (Montalban et al., 2018), o a problemas técnicos, como el uso de inhibidores del NMD en los ensayos funcionales con minigenes que permiten detectar transcritos-PTC menores, al tipo de célula o a los métodos utilizados para la caracterización de transcritos, como RT-PCR o RNA-seq del RNA del paciente.

Generalmente, las variantes de sitios de *splicing* situadas en las posiciones  $\pm 1,2$  son consideradas patogénicas, incluso son clasificadas como variantes truncantes de proteína en ausencia de un análisis funcional (Dorling et al., 2021). Sin embargo, algunas de ellas producen isoformas que mantienen la pauta de lectura o incluso producen transcrito *full-length* como c.108+2T>C (Tabla 17). Además, las variantes que afectan a otros nucleótidos conservados de los sitios de *splicing* son potencialmente perjudiciales. Por lo tanto, mientras que la mayoría de las variantes de sitio de *splicing* de nuestro estudio (20/37) afectaban a posiciones  $\pm 1,2$  nucleótidos, 17 variantes afectaban a otras posiciones (los últimos tres nucleótidos exónicos y posiciones intrónicas +3, +4, +5 y +6). Estas variantes, también alteraron los patrones de *splicing*, confirmando su relevancia en el reconocimiento del sitio de *splicing* y sugiriendo un posible papel en la patogenicidad de la enfermedad.

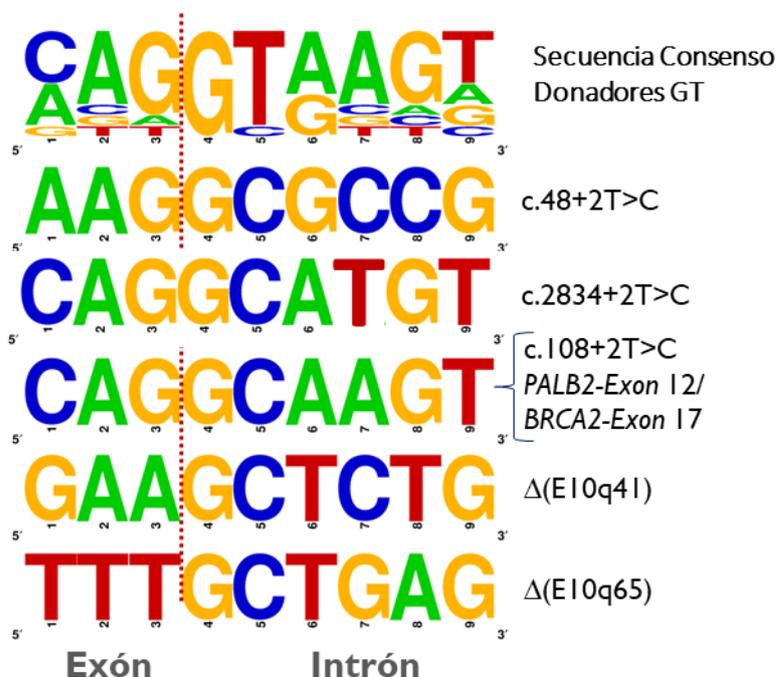
En este sentido, las alteraciones de *splicing* producidas a causa de las variantes situadas en estas posiciones (de +3 a +6), fueron especialmente difíciles de predecir por MES, observando impactos muy débiles o nulos en el *splicing* para variantes como: c.48+4C>T o c.2834+3A>G (Tablas 14-16). Además, otras dos variantes de sitio de *splicing*, c.3113+3A>G y c.3201+6T>A, fueron seleccionadas porque cambiaban nucleótidos muy conservados a pesar de su baja puntuación MES (Tabla 16). Sin embargo, ambas variantes tuvieron impactos relevantes en el *splicing* (92% y 88,5% de transcritos aberrantes, respectivamente). Por tanto, el criterio de selección de cambios en nucleótidos conservados es acertado para la selección de variantes potencialmente *spliceogénicas*.

También es importante recordar los resultados obtenidos con las variantes +2T>C (c.48+2T>C, c.108+2T>C y c.2834+2T>C) que convierten un donador GT canónico en un GC atípico. Aproximadamente el 1% de los donadores humanos GC están asociados con mecanismos de *splicing* alternativo fisiológico (Thanaraj and Clark, 2001; Kralovicova et al., 2011). Las

variantes c.48+2T>C, c.108+2T>C y c.2834+2T>C inducen diferentes impactos en el *splicing* (0%, 86% y 0% del transcrito FL, respectivamente; Figuras 23-32; Tabla 16). De hecho, se ha estimado previamente que entre el 15% y el 18% de las sustituciones +2T>C son capaces de generar cantidades variables del transcrito canónico (1-84%) (Lin et al., 2019). En consecuencia, todos estos cambios deben interpretarse con cautela porque no siempre están asociados con alteraciones del *splicing*. Asimismo, se han observados dos transcritos anómalos menores  $\Delta(E10q41)$  y  $\Delta(E10q65)$  generados a causa del uso de un donador GC atípico. Los sitios de *splicing* GC-5' son débiles por el cambio de un nucleótido esencial en la posición +2, pero, se ha propuesto que para compensar este desajuste, la secuencia consenso alrededor del donador GC ha evolucionado para maximizar la interacción en otras posiciones del snRNA de U1 (Jackson, 1991; Burge et al., 1999; Bursset et al., 2000; Mount, 2000).

La secuencia de los donadores GC naturales del exón 12 de *PALB2* y del exón 17 de *BRCA2*, así como el donador GC creado *de novo* a causa de la variante c.108+2T>C (Transcrito FL, 86%) es idéntica (CAG | gcaagt) (Figura 47). Estos datos sugieren que este donador creado *de novo* se comporta igual que los donadores naturales del exón 12 de *PALB2* y del exón 17 de *BRCA2*, y que podría estar regulado de manera similar. De hecho, varios estudios de regulación con siRNA han sugerido que el reconocimiento del exón 17 de *BRCA2* está mediado por SC35, SF2/ASF y Tra2 $\beta$  (Fraile-Bethencourt et al., 2017), mientras que 9G8, Tra2 $\beta$  y SC35 participan en el reconocimiento de un donador GC anómalo utilizado en el gen *BTK* (Kralovicova et al., 2011). Finalmente, también debe destacarse el resultado obtenido con la variante c.3348C>T que afecta al sitio 5'ss del exón 12. Según los resultados bioinformáticos generados por MES, c.3348C>T debilita el reconocimiento del donador GC natural (puntuación MES: 3.1 $\rightarrow$ 1.9) y al mismo tiempo crea un sitio GT más fuerte (puntuación MES: 6.5). Sin embargo, los resultados obtenidos mediante el análisis de fragmentos fluorescentes revelaron que la isoforma canónica era la más abundante (68%). Además, se detectaron los transcritos ( $\Delta(E12q4)$ , 26,7%) y ( $\Delta(E12)$ , 5,3%) de forma minoritaria, lo que indica un uso preferencial del donador GC, respaldando la existencia de una

regulación específica del exón 12 de *PALB2*, mediante factores de *splicing*, que mejoran el reconocimiento de este sitio.



**Figura 47. Secuencias del sitio de *splicing* GT canónico y los sitios GC atípicos de las variantes c.48+2T>C, c.108+2T>C, y c.2834+2T>C, del exón 12 de *PALB2*, del exón 17 de *BRCA2* exón 17 y de los transcritos Δ(E10q41) y Δ(E10q65). El tamaño de cada letra representa la frecuencia del nucleótido en cada posición. Los pictogramas se obtuvieron con WebLogo (<https://weblogo.berkeley.edu/logo.cgi>).**

### 1.1. Clasificaciones Clínica de variantes BRIDGES.

Para desarrollar esta clasificación clínica provisional ha sido fundamental conocer los transcritos biológicamente relevantes producidos por *PALB2* y las regiones codificantes esenciales de esta proteína para su correcta función. Tradicionalmente, las variantes han sido clasificadas sobre la base de los indicadores biológicos de patogenicidad: i) impacto sobre el *splicing* (abundancia de transcrito(s) aberrante(s)); ii) impacto sobre la función de la proteína: truncamiento de la proteína o afectación de dominios esenciales en los transcritos que mantienen la pauta de lectura. Sin embargo, en los últimos años se ha destinado un considerable esfuerzo a la elaboración de reglas

estrictas y universales que permitan una clasificación de todas las variantes identificadas. En este sentido, la ACMG/AMP ha elaborado una serie de reglas que han sido modificadas en función del gen, la enfermedad o el tipo de variante (Richards et al., 2015). Nuestro grupo ha colaborado estrechamente con el dr. Miguel de la Hoya (Hospital Clínico de San Carlos, Madrid) para la aplicación de dichas reglas ACMG/AMP sobre los resultados de *splicing* de variantes obtenidos en minigenes (Sanoguera-Miralles et al., 2020; Bueno-Martínez et al., 2021; Valenzuela-Palomo et al., 2021). Un aspecto muy importante de dicha clasificación es la evaluación del impacto de los transcritos aberrantes en la funcionalidad del gen en cuestión para estimar la patogenicidad de una variante.

## **1.2. Transcritos de *PALB2* biológicamente relevantes. Regiones codificantes críticas de *PALB2*.**

Aunque se desconoce el número exacto de isoformas generadas por *PALB2* a causa del *splicing* alternativo, estudios recientes han caracterizado un mínimo de 88 eventos diferentes de *splicing* alternativo (Hofstatter et al., 2011; Blanco et al., 2013; Lopez-Perolio et al., 2019). Sin embargo, únicamente la expresión combinada de  $\Delta(E11)$ ,  $\Delta(E12)$  y  $\Delta(E11\_E12)$ ; transcritos PTC que codificarán presumiblemente una proteína no funcional, representan una proporción significativa de la expresión total del gen *PALB2* (Hofstatter et al., 2011; Blanco et al., 2013; Lopez-Perolio et al., 2019). En consecuencia, creemos que NM\_000251.3/ENST00000261584.9 (13 exones codificantes y 1186 aminoácidos), es el único transcrito biológicamente relevante para propósitos de clasificación de variantes. Con esto, estamos asumiendo que las otras isoformas de *PALB2* generadas de forma natural no pueden rescatar la funcionalidad de esta proteína en la ruta de reparación del DNA de doble cadena (es decir, que no codifican una proteína funcional y/o se expresa en niveles muy bajos). Esto último tiene implicaciones para la clasificación de variantes y por ello proponemos que una variante de pérdida de función ubicada en el exón 11, es una variante que califica para un código de patogenicidad muy fuerte PVS1, a pesar de haberse caracterizado

previamente eventos de *splicing* alternativo natural que carecían de los exones 11 y 12 (Hofstatter et al., 2011; Blanco et al., 2013; Lopez-Perolio et al., 2019).

La interacción de PALB2 con BRCA1 y BRCA2 es fundamental para la recombinación homóloga (HR) (Blanco et al., 2012; Ramus et al., 2015). PALB2 interactúa con BRCA1 a través de un dominio superenrollado en su extremo amino terminal (dominio CC). Además, PALB2 tiene una serie de repeticiones WD en su extremo C-terminal. Estas repeticiones WD forman juntas una estructura  $\beta$ -propulsora que interacciona directamente con BRCA2 (Oliver et al., 2017) y están involucradas en la interacción con la DNA polimerasa  $\eta$ , RAD51, RAD51C y la ubiquitina ligasa RNF168 (Buisson et al., 2014; Park et al., 2014; Luijsterburg et al., 2017). Es importante destacar que algunas variantes *missense* situadas en esta región reducen la eficiencia de la recombinación homóloga en más de un 50% (Blanco et al., 2012; Fewings et al., 2018), situándola como una región crítica para la correcta función de PALB2 (Dorling et al., 2021). Del mismo modo, variantes *missense* situadas en el dominio superenrollado (dominio CC), reducen la eficiencia en las mismas proporciones (Blanco et al., 2012; Fewings et al., 2018).

Por último recalcar la presencia de otros tres dominios PALB2 fundamentales para su correcta función; el dominio de unión a RAD51, que abarca de los residuos 100 a 184, el motivo asociado a cromatina (CHAM) que abarca de los residuos 394 a 450, y el gen relacionado con MORF, en el dominio de interacción del cromosoma 15 (MRG15) que abarca de los residuos 611 a 764 (Ramus et al., 2015; Fewings et al., 2018). Hasta la fecha, no se han observado alteraciones en la eficiencia de recombinación homóloga, ocasionada por variantes *missense* en estos dominios (Blanco et al., 2012; Fewings et al., 2018). Por tanto, nuestro análisis clínico se centrará en los dominios CC y WD40, ignorando el papel de los 3 mencionados anteriormente.

Mientras que se asume la pérdida de función de PALB2 por parte de la gran mayoría de los transcritos que incluyen codones prematuros de terminación, el impacto de los transcritos *in-frame* en la actividad del gen y la susceptibilidad a cáncer de mama es más complicada de analizar. A continuación, se va a analizar resumidamente el impacto de los transcritos que mantienen la pauta

de lectura. Hay 5 transcritos asociados a pérdida completa de exones:  $\Delta(\mathbf{E2})$ ,  $\Delta(\mathbf{E7})$ ,  $\Delta(\mathbf{E9})$ ,  $\Delta(\mathbf{E9\_E10})$  and  $\Delta(\mathbf{E10})$ . Todos ellos eliminan regiones críticas para la función de PALB2 y por tanto se califican con un código patogénico ACMG/AMP muy fuerte PS3\_VS (Richards et al., 2015; Valenzuela-Palomo et al., 2021).

Así, la proteína codificada por  $\Delta(\mathbf{E2})$ , p.(Leu17\_Gln36del), elimina la región central de la  $\alpha$ -hélice del dominio CC. Siete residuos (17, 21, 24, 28, 31, 35 y 35) median interacciones importantes del homodímero de PALB2 y el heterodímero PALB2/BRCA1. Además, los estudios funcionales de varias *missense* han mostrado que afectan a la actividad HR (Boonen et al., 2019, 2020; Rodrigue et al., 2019; Wiltshire et al., 2020; Brnich et al., 2021; Nepomuceno et al., 2021). El transcrito  $\Delta(\mathbf{E7})$  codifica para la proteína p.(Asn863\_Glu916del) que ha perdido la hoja 1 de WD40 (Oliver et al., 2017). Al menos una variante *missense* situada en este intervalo, c.2734T>G p.(Thr912Gly) tiene impacto funcional severo en la actividad HR (Boonen et al., 2020; Nepomuceno et al., 2021). La proteína p.(Ala946\_Gly999del) codificada por  $\Delta(\mathbf{E9})$  no incluye las cadenas  $\beta$  (2C, 2D, 3A, 3B y 3C) de las hojas 2 y 3 del dominio WD40. Además, variantes *missense* situadas en este exón como c.2841G>C p.(Leu947Ser) y c.2882T>C p.(Leu961Pro), tienen un impacto negativo sobre la recombinación homóloga (Boonen et al., 2020; Nepomuceno et al., 2021). La proteína p.(Gly1000\_Trp1038del) codificada por  $\Delta(\mathbf{E10})$  carece de las cadenas (3D, 4A, 4B, and 4C) de las hojas 3 y 4 del dominio WD40. Los aminoácidos Val1019, Met1022, Ala1025 e Ile1037 se encuentran en el núcleo hidrofóbico que interactúa con *BRCA2* (Oliver et al., 2017). Además, se ha observado como c.3073G>A p.(Ala1025Arg), c.3089C>T p.(Thr1030Ile) y c.3110T>C p.(Ile1037Thr) tienen un impacto funcional severo en la actividad HR (Boonen et al., 2020; Nepomuceno et al., 2021). Las predicciones funcionales para  $\Delta(\mathbf{E9\_E10})$  son incluso más drásticas que las de  $\Delta(\mathbf{E9})$  y  $\Delta(\mathbf{E10})$ .

Por último, el transcrito *in-frame*  $\Delta(\mathbf{E4})$  (pérdida completa del exón 4) no codifica ni para el dominio CC ni para WD40. La proteína que codifica p.(Glu71\_Lys561del), carece del motivo CHAM. Sin embargo, se desconoce la relevancia funcional de este dominio y no hay ninguna *missense* descrita e identificada en esta región con impacto funcional (Boonen et al., 2020;

Nepomuceno et al., 2021). Aunque la relevancia del tamaño de la deleción (se pierde más del 10% de la secuencia codificante de *PALB2*) nos podría indicar una fuerte evidencia a favor de su patogenicidad.

Por otro lado, **▼(E1q9)**, **Δ(E2p6)** y **Δ(E8p15)** generan pequeñas alteraciones en la pauta de lectura en regiones críticas para la función de *PALB2*. La proteína codificada por **▼(E1q9)**, p.(Lys16\_Leu17insValProGly) incluye 3 aminoácidos extra en el dominio superenrollado, entre Lys16 y Leu17. Lys16 está involucrado en numerosas interacciones dentro de la hélice  $\alpha$ , con los residuos circundantes y Leu17 interacciona directamente con Leu35 en la interfaz del Homodímero *PALB2* (Song et al., 2018). Probablemente, la inserción de 3 aminoácidos entre estos 2 residuos altere la estructura de la hélice  $\alpha$ , pero no podemos predecir el efecto funcional de la proteína sin desarrollar un ensayo funcional de reparación del DNA. La isoforma **Δ(E2p6)**, caracterizada previamente por Lopez-Perolio y colaboradores (Lopez-Perolio et al., 2019), codifica para p.(Leu17\_Lis18del), que no incluye 2 residuos del dominio superenrollado. Como ya se indicó, Leu17 interacciona con Leu35 en la interfaz del homodímero *PALB2*. Un estudio funcional reciente con las *missense*, c.50C>T p.(Leu17Pro) (Brnich et al., 2021), mostró una bajada severa en la eficiencia de la recombinación homóloga. Sin embargo, no creemos que este resultado sea una prueba de que Leu17 sea fundamental en esta ruta, ya que la prolina es capaz de romper las hélices  $\alpha$  (Li et al., 1996). La proteína p.(Val917\_Gln921del) codificada por **Δ(E8p15)** carece de la cadena  $\beta$  2A de WD40. Probablemente, esto desestabilizará la hoja 2 de este dominio y, como resultado, el plegamiento general de WD40. Ninguna *missense* situada en la hebra  $\beta$  2A ha demostrado alterar la eficiencia de HR en un ensayo funcional, sin embargo, la *missense* c.2755G>A p.(Val919Ile) no produjo ningún efecto (Boonen et al., 2020; Nepomuceno et al., 2021). Comparando estos datos con los resultados bioinformáticos, creemos que **▼(E1q9)**, **Δ(E2p6)** y **Δ(E8p15)** proporcionan una evidencia moderada a favor de su patogenicidad (Richards et al., 2015; Valenzuela-Palomo et al., 2021).

Por último, la proteína p.(Ser70\_Glu71ins16) codificada por **▼(E3q48)** incluye 16 aminoácidos adicionales (GKSRPFTYACFIIHFP) en una región de *PALB2* en la que no se ha

descrito ningún dominio funcional. Hasta donde sabemos, la ACMG no ha proporcionado recomendaciones específicas sobre inserciones en proteínas. Por tanto, creemos que, en la actualidad, no existe una justificación estructural y/o funcional que sustente una evidencia a favor de la patogenicidad (Richards et al., 2015).

Debido a que no existe un panel de especificaciones de clasificación para el gen *PALB2* en la guía ACMG/AMP (<https://clinicalgenome.org/>, ultimo acceso 16 June 2021), el dr. Miguel de la Hoya (Hospital Clínico San Carlos, Madrid) en colaboración con nuestro laboratorio ha desarrollado sus propias especificaciones provisionales para clasificar clínicamente las variantes de *PALB2* ensayadas funcionalmente en minigenes (ver apartado: **3.1** de Materiales y métodos) (Valenzuela-Palomo et al., 2021). Para abordar esta clasificación, se encontraron algunas dificultades como la interpretación de variantes que inducen dos o más transcritos que pronosticaban diferentes impactos en la secuencia codificante.

En resumen, de acuerdo a los criterios ACMG/AMP específicos de *PALB2*, las variantes ensayadas en los minigenes quedan clasificadas según se indica en la Tablas 26 y S2 (Valenzuela-Palomo et al., 2021). De acuerdo a estos estándares, 5 variantes fueron clasificadas como patogénicas, 20 como probablemente patogénicas, 11 como VUS y 8 como probablemente benignas. No se pretende proporcionar una clasificación clínica definitiva de las variantes de *PALB2*, sino ilustrar también las dificultades asociadas a la aplicación de los criterios ACMG/AMP para la interpretación de los resultados complejos de variantes de *splicing*.

**Tabla 26. Clasificación ACMG/AMP de las variantes BRIDGES ensayadas.**

Clasificación ACMG/AMP	Variantes
<b>Patogénicas</b>	c.108+1G>A, c.3113G>A, c.3113+5G>C, c.3350+4A>G y c.3350+5G>A.
<b>Probablemente patogénicas</b>	c.47A>G, c.48G>A, c.48+1G>A, c.48+2T>C, c.109-2A>G, c.1685-2A>C, c.1685-2A>G, c.2515-2A>G, c.2587-2A>G, c.2748+1G>T, c.2748+1G>A, c.2748+2dup, c.2749-1G>T, c.2834+1G>A, c.2834+2T>C, c.2997-2del, c.2997-1G>A, c.3113+3A>G, c.3201+1G>A y c.3201+3_+4insTG.
<b>VUS</b>	c.49-2A>T, c.108+2T>C, c.211+1G>A, c.1684+4A>G, c.2513A>C, c.2834G>C, c.2834+3A>G, c.2834+5G>A, c.2996+4A>G, c.3201+6T>A y c.3348C>T.
<b>Probablemente benignas</b>	c.48+4C>T, c.48+7G>C, c.109-6_109-4del, c.211+5del, c.2586+4A>T, c.2748+4A>T, c.2750T>C y c.2834+6T>C

Según estos datos, 25 variantes serían clasificadas como patogénicas/probablemente patogénicas, lo cual implica que son clínicamente accionables. Interesantemente, los datos de minigenes reducen significativamente la proporción de VUS reportadas por ClinVar, pasando del 57% de VUS-ClinVar al 23% incluyendo nuestros resultados. Finalmente, las variantes patogénicas/probablemente patogénicas de *splicing* representarían aproximadamente el 15% de todas las variantes patogénicas notificadas en sujetos BRIDGES, subrayando la importancia de las alteraciones de *splicing* como mecanismo primario de patogenicidad.

## 2. Variantes ClinVar.

Nuestro estudio previo centrado en el análisis funcional de variantes BRIDGES representa el análisis más completo de variantes de *splicing* de *PALB2* en CMOH, permitiendo, además, con las tres construcciones generadas para ese estudio, ensayar cualquier variante potencialmente *espliceogénica* de este gen.

La implementación de la secuenciación masiva en el entorno clínico, ha permitido descubrir varios miles de variantes de la línea germinal en *PALB2*. Actualmente, en la base de datos ClinVar y Simple-ClinVar (último acceso 3 de agosto de 2021; <https://www.ncbi.nlm.nih.gov/clinvar/?term=palb2%5Bgene%5D>; Figura 3), se han reportado 3820 variantes diferentes en *PALB2*, de las cuales el 56% han sido clasificadas como VUS. Se analizaron con MES 151 variantes de las fronteras intrón-exón de *PALB2*, de las que 67 fueron seleccionadas para estudios funcionales. La intención inicial era realizar ensayos funcionales de todas las variantes ClinVar seleccionadas en los tres minigenes clonados anteriormente, y así generar “un catálogo funcional” de gran valor para su interpretación clínica, sin embargo, la limitación del tiempo a causa de la COVID-19 nos ha impedido completar el trabajo.

A pesar de esto, las 11 variantes seleccionadas de los primeros 3 exones pudieron ser testadas en el minigén mgPALB2\_ex1-3. Todas las variantes seleccionadas alteraron el *splicing*, subrayando la especificidad y sensibilidad de nuestros criterios de selección. Además, MES predijo correctamente el efecto de las variantes sobre la maquinaria de *splicing* (ya sea por la eliminación del sitio de *splicing* o por reducciones significativas de la puntuación) (Tabla 19). Diez variantes afectan a las posiciones clásicas  $\pm 1,2$  del 5'ss o 3'ss y ninguna de ellas produce trazas del transcrito canónico, además, la variante c.46A>G afecta al antepenúltimo nucleótido del exón 1. El análisis de fragmentos fluorescentes reveló la existencia de 7 transcritos aberrantes ya descritos en las variantes BRIDGES (3 transcritos PTC:  $\Delta(E1q17)$ ,  $\Delta(E3p11)$  y  $\Delta(E3)$  y 4 transcritos *in-frame*:  $\nabla(E1q9)$ ,  $\Delta(E2)$ ,  $\Delta(E2p6)$  y  $\nabla(E3q48)$ ). Por último, cabe destacar, que 4 de las isoformas mencionadas anteriormente habían sido previamente caracterizadas como eventos naturales de *splicing*:  $\nabla(E1q9)$ ,  $\Delta(E1q17)$ ,  $\Delta(E2p6)$  y  $\Delta(E2)$  (Lopez-Perolio et al., 2019), lo que sugiere que los eventos de *splicing* alternativo fisiológico pueden predecir de alguna manera los perfiles de *splicing* de las distintas variantes. Asimismo, los ensayos funcionales con minigenes híbridos han sido capaces de imitar los patrones aberrantes de las diferentes variantes, reproduciendo fielmente resultados previos de ensayos de RNA de pacientes como es el caso de la variante c.49-1G>A (Lopez-Perolio et al., 2019).

Las variantes que alteran los sitios 5'ss o 3'ss de *splicing* suelen considerarse perjudiciales, según las recomendaciones de la ACMG para la interpretación y notificación de variaciones de secuencia (Richards et al., 2008). No obstante, algunas variantes de sitio de *splicing* pueden producir un riesgo intermedio si: a) generan un porcentaje suficiente de transcritos que codifiquen para una proteína funcional, a pesar de producir a su vez transcritos PTC, como ocurre con algunas mutaciones que afecta al exón 17 de *BRCA1* (Casadei et al., 2019), b) alteran el marco de lectura como consecuencia de un *splicing* aberrante, que conduce a la pérdida de dominios no esenciales para la proteína, como ocurre con la variantes c.3154-1G>A de *ATM* (Casadei et al., 2019), o c) tienen efectos moderados en la función de la proteína (Vallée et al., 2016). El impacto de los transcritos *in-frame* sobre la funcionalidad de *PALB2* se ha discutido en la sección anterior de la Discusión.

Si intentáramos realizar una clasificación ACMG/AMP de forma análoga a la sección anterior (Valenzuela-Palomo et al., 2021), de las 11 variantes ensayadas: 5 variantes serían clasificadas como, posiblemente patogénicas (c.48+1G>C, c.48+2T>G, c.48+1del, c.108insC y c.109-2A>C), 5 como VUS (c.49-1del, c.49-2del, c.49-1G>A, c.211+1G>T y c.211+2T>C) y 1 como posiblemente benigna (c.46A>G). Así, las variantes c.48+1G>C, c.48+2T>G y c.48+1del presentan unos resultados similares a las variantes BRIDGES c.47A>G, c.48G>A, c.48+1G>A y c.48+2T>C que fueron clasificadas como posiblemente patogénicas (Tabla 26). Por otro lado, las variantes c.108insC y c.109-2A>C se asemejan a los resultados de las variantes BRIDGES c.108+1G>A y c.109-2A>G, respectivamente, que se clasificaron como posiblemente patogénicas. Asimismo, las variantes c.49-1del, c.49-2del y c.49-1G>A presentan como transcrito único  $\Delta(E2p6)$  al igual que la variante BRIDGES c.49-2A>T que fue clasificada como VUS. Las variantes c.211+1G>T y c.211+2T>C produjeron resultados similares a la variante BRIDGES c.211+1G>A, catalogada como VUS. Por último, la variante c.46A>G produjo más de un 90% de transcrito *full-length* que es el umbral acordado para descartar la pérdida de función asociada a una variante que afecta al *splicing*. La clasificación definitiva de una variante como patogénica o posiblemente

patogénica facilita que las portadoras y familias puedan beneficiarse de los protocolos de prevención indicados, así como de la aplicación de terapias personalizadas.

### **3. Mapeo Funcional de elementos reguladores del *splicing* de los exones 11 y 12 de *PALB2*.**

Los exones 11 y 12 de *PALB2*, como se ha indicado en el apartado 3 de Resultados, son diana de eventos de *splicing* alternativo fisiológico habiéndose caracterizado hasta 8 transcritos diferentes: ▼(E11p23), Δ(E11p2), Δ(E11), Δ(E11) + ▼(E12p446), ▼(E12p65), Δ(E12p136), Δ(E12) y Δ(E11\_E12) (Lopez-Perolio et al., 2019). El exón 12 es especialmente interesante ya que posee un donador atípico GC, cuyo reconocimiento requiere la participación de proteínas SR que se unirán a activadores de la transcripción (Kralovicova et al., 2011; Fraile-Bethencourt et al., 2017). El reconocimiento de exones diana de *splicing* alternativo requiere de la actuación de enhancers y silenciadores de *splicing*. Asimismo, en trabajos anteriores se ha demostrado la participación de factores de *splicing* en la regulación de los donadores-GC (Kralovicova et al., 2011; Fraile-Bethencourt et al., 2017). Por tanto, todo ello sugería que el reconocimiento de ambos exones requería de elementos reguladores adicionales. Para ello, se decidió mapear funcionalmente los exones 11 y 12, evaluando el impacto en el *splicing* de 8 microdeleciones que cubrían toda la secuencia codificante de los exones 11 y 12. Esta estrategia ya había sido aplicada con éxito por nuestro grupo para el estudio de diferentes exones de los genes *BRCA2* y *RAD51D* (Acedo et al., 2015; Fraile-Bethencourt et al., 2017; Bueno-Martínez et al., 2021). Gracias a ello, se identificaron intervalos clave en determinados exones, como los exones 17 y 18 de *BRCA2* o el exón 3 de *RAD51D*, que contenían motivos que promovían el reconocimiento de sus respectivos exones. De este modo, se observó que las deleciones c.3143del29, c.3229del30 y c.3293del30 de los exones 11 y 12 de *PALB2* tenían un impacto fuerte sobre el *splicing*, indicando que los intervalos c.3143-c.3171, c.3229-c.3258 y c.3293-c.3322 deben contener secuencias que controlan el reconocimiento de ambos exones. Consecuentemente, cualquier variante situada en estos fragmentos debe ser considerada candidata a inducir alteraciones de *splicing*.

Para probar esta hipótesis, con la ayuda del programa Hexplorer se seleccionaron 3 variantes de cada uno de estos intervalos para ser estudiadas funcionalmente en el minigén mgPALB2\_ex5-12. Los resultados confirmaron la hipótesis de partida, ya que 8 (6 *missense*, 1 *nonsense* y 1 *frameshift*) de las 9 variantes ensayadas (89%) causaron alteraciones de *splicing*, apoyando la sensibilidad de predicción in silico del programa HEXplorer como herramienta de selección. La electroforesis capilar fluorescente de los productos RT-PCR nos permitió observar la presencia de 7 transcritos aberrantes (6 PTC y 1 *in-frame*), 2 de los cuales,  $\Delta(E11)$  y  $\Delta(E12)$ , habían sido previamente caracterizados.

Curiosamente, las variantes c.3155A>T p.(Asp1052Val) y c.3157G>T p.(Asp1053Tyr) situadas en el exón 11 produjeron  $\Delta(E12q8)$  en un 26,4%, y  $\Delta(E11)+\nabla(E12q96)$  en un 6,3%, respectivamente. Este dato, junto con la presencia de transcritos alternativos naturales  $\Delta(E11_12)$ , sugieren la existencia de una regulación conjunta del proceso de *splicing* de los exones 11 y 12 de *PALB2*, de igual forma que ocurre con los exones 17 y 18 de *BRCA2* (Fraile-Bethencourt et al., 2017). Asimismo, se han observado transcritos aberrantes  $\Delta(E12q33)$ , 5,1% y  $\Delta(E12q8)$ , (26,4%, 2,4% y 47,1%), generados a causa del uso de donadores GC atípicos creados de novo con las secuencias CAG | gctggc y GAT | gctgta, respectivamente.

Además del dominio superenrollado, que media la interacción con BRCA1, el dominio C-terminal WD40 de PALB2 (aminoácidos 853-1186) es fundamental en las interacciones con otras proteínas esenciales de la recombinación homóloga, como BRCA2 y RAD51. Estudios recientes han demostrado el efecto deletéreo de numerosas *missense* catalogadas como VUS y situadas en este dominio funcional, por medio de ensayos de reparación del DNA, como por ejemplo c.2734T>G p.(Trp912Gly), c.3089C>T p.(Thr1030Ile), c.3418T>G p.(Trp1140Gly) y c.3515T>C p.(Leu1172Pro) (Boonen et al., 2019). Asimismo, Pauty y colaboradores mostraron como el dominio C-terminal WD40 de PALB2 ocultaba una señal de exportación nuclear que podía ser expuesta por la presencia ciertas variantes como c.3113G>A, clasificada como patogénica en este estudio, que a pesar de seguir siendo competente para estimular la recombinación homóloga mediada por RAD51 *in vitro*, favorecía la translocación de la proteína del núcleo al citoplasma,

estableciendo un vínculo entre una mala localización celular de PALB2 y la aparición del cáncer (Pauty et al., 2017). Por último, Caleca y col. observaron mediante ensayos funcionales de interacción proteica con GFP, como la variante *PALB2*: c.3262C>T p.(Pro1088Ser), procedente de pacientes enfermos, impedía la interacción BRCA2-PALB2 (Caleca et al., 2018).

Estos resultados deberían hacernos pensar que las variantes deletéreas situadas en el dominio WD40 de PALB2 y más concretamente en los exones 11 y 12, obstaculizaran la interacción del heterodímero BRCA2-PALB2, bajando la eficiencia de la recombinación homóloga. Sin embargo, no podemos descartar la posibilidad de pensar que las variantes patogénicas situadas los exones 11 y/o 12 confieren un riesgo asociado al cáncer inferior a las de otros exones, debido al elevado porcentaje de isoformas  $\Delta(E11)$ ,  $\Delta(E12)$  y  $\Delta(E11\_E12)$  generadas por *splicing* alternativo fisiológico (8%-32%). Además, todas variantes exónicas ensayadas, a excepción de c.3161C>T ( $\Delta(E11)$ ; 100%), producen porcentajes superiores al 35% del transcrito *full-length*. Por lo tanto, creemos que sería necesario realizar ensayos funcionales de reparación del DNA para poder clasificar estas variantes de forma satisfactoria.

Por último, una de las variantes, c.3161C>T, produjo sólo el transcrito aberrante  $\Delta(E11)$  sin ninguna traza del transcrito *full-length*. Como  $\Delta(E11)$  contiene un PTC (p.(Asn1039Glyfs\*7)), podríamos concluir que c.3161C>T podría ser clasificada como probablemente patogénica, al igual que las variantes c.3201+1G>A y c.3201+3\_3201+4insTG (Tablas 25-26) que presentan exactamente el mismo impacto sobre el *splicing* ( $\Delta(E11)$  – 100%). Por tanto, podemos concluir que el mapeo funcional es una estrategia valiosa que permite detectar nuevos alelos de riesgo que afectan a otro tipo de elementos esenciales que regulan el *splicing*. Dicho esto, creemos que todas las variantes situadas en los intervalos c.3143-c.3171, c.3229-c.3258 y c.3293-c.3322 deberían ser ensayadas funcionalmente debido a sus posibles efectos adversos sobre el proceso de *splicing* y a su asociación con la susceptibilidad a cáncer de mama. El uso de los minigenes aquí descritos facilitará esta tarea, debido a que, en nuestras manos, se trata de un sistema sencillo, robusto, sensible y de alta capacidad.



# **CONCLUSIONES**

## Conclusiones Generales.

1. Se han desarrollado en nuestro laboratorio los minigenes: mgPALB2\_ex-1-3, mgPALB2\_ex-4-6 y mgPALB2\_ex-5-12 del gen de susceptibilidad a cáncer de mama *PALB2*.

## Variantes sitios de *splicing* (BRIDGES y ClinVar).

2. El cribado Bioinformático mediante MES, nos permitió seleccionar 109 variantes (44 variantes BRIDGES y 67 variantes ClinVar) para su posterior análisis.
3. La mayoría de las variantes ensayadas (48/55, 87,3%) alteraban el *splicing* de *PALB2* y 35 no mostraba trazas del transcrito canónico o menos de un 5%, subrayando la alta sensibilidad de MES en la selección de variantes.
4. Hemos encontrado 34 transcritos, incluyendo el full-length de los cuales 16 habían sido previamente caracterizados como eventos naturales de *splicing*.
  - Veintitrés transcritos predijeron la introducción de codones stop prematuros que probablemente inactiven *PALB2*, mientras que otros 10 transcritos mantenían la pauta de lectura.
  - El evento de *splicing* aberrante más común fue el uso de un sitio alternativo (22) *splicing*, seguido por el *exon skipping* (10)
  - Veintitrés transcritos predijeron la introducción de codones de parada prematuros que probablemente inactiven *PALB2*, mientras que otros 10 transcritos mantenían la pauta de lectura.
5. El reconocimiento de los donadores GC requiere una conservación de las bases muy estricta.
  - Las variantes c.108+2T>C, c.48+2T>C, y c.2834+2T>C transformaban un donador GT en un donador atípico GC. Las variantes c.48+2T>C y c.2834T>C

solo producen transcritos aberrantes, mientras que c.108+2T>C, que muestra la secuencia más conservada, genera un 85.5% de transcrito canónico.

6. Variantes situadas en la misma posición de la secuencia consenso muestran impactos muy diferentes en *splicing* (c.2834+6T>C; 0% *full-length* y c.3201+6T>A, 88,5 % *full-length*).

#### **Elementos Reguladores del *Splicing* de los exones 11 y 12.**

7. El mapeo funcional mediante microdeleciones ha sido una valiosa estrategia para la identificación de 3 zonas ricas en ESEs de los exones 11 y 12 de *PALB2*: c.3143-c.3171, c.3229-c.3258 y c.3293-c.3322.
8. Ocho de las 9 variantes exónicas localizadas en los intervalos indicados alteraron el *splicing*. La variante c.3161C>T no produjo trazas del transcrito *full-length*.

#### **Conclusiones finales.**

9. Los ensayos funcionales de *splicing* con minigenes son una valiosa estrategia para la caracterización inicial de las variantes de *splicing* y la posterior interpretación clínica de las variantes de cualquier enfermedad genética.
10. La desregulación del *splicing* es uno de los principales mecanismos de inactivación génica en enfermedades hereditarias. Una alta proporción de las variantes estudiadas alteran el *splicing* (56/64, 87,5%). Interesantemente, 10 variantes missense, 3 sinónimas, 1 nonsense y 1 frameshift deberían ser reclasificadas como variantes de *splicing*.
11. Treinta y una variantes han sido clasificadas como patogénicas o probablemente patogénicas. Además, hemos reducido el porcentaje de variantes BRIDGES clasificadas como VUS reportadas por ClinVar, pasando del 57% de VUS al 23% incluyendo los resultados de minigenes. Curiosamente algunas variantes situadas en posiciones  $\pm 1,2$  han sido reclasificadas como VUS después de los resultados obtenidos.



# **BIBLIOGRAFÍA**

Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, Harrison SM. 2018a. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* 39:1517–1524.

Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, Harrison SM. 2018b. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* 39:1517–1524.

Acedo A, Hernández-Moro C, Curiel-García Á, Díez-Gómez B, Velasco EA. 2015. Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat* 36:210–221.

Acedo A, Sanz DJ, Durán M, Infante M, Pérez-Cabornero L, Miner C, Velasco EA. 2012. Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res* 14:.

Aebi M, Hornig H, Weissmann C. 1987. 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. *Cell* 50:237–246.

Allender Y, Arias A. 2008. Cáncer de mama y ovario hereditario. *Rev Per Ginecol Obs* 54:194–198.

Anantha RW, Alcivar AL, Ma J, Cai H, Simhadri S, Ule J, König J, Xia B. 2013. Requirement of Heterogeneous Nuclear Ribonucleoprotein C for BRCA Gene Expression and Homologous Recombination. *PLoS One* 8:.

Anczuków O, Rosenberg AZ, Akerman M, Das S, Zhan L, Karni R, Muthuswamy SK, Krainer AR. 2012. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat Struct Mol Biol* 19:220–228.

Andreopoulou E, Kelly CM, McDaid HM. 2017. Therapeutic Advances and New Directions for Triple-Negative Breast Cancer. *Breast Care* 12:21–28.

Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkäs K, Roberts J, Lee A, Subramanian D, Leeneer K De, Fostira F, Tomiak E, Neuhausen SL, et al. 2014. Breast-cancer risk in families with mutations in PALB2. *N Engl J Med* 371:497–506.

Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S. 2014. CircRNA Biogenesis competes with Pre-mRNA splicing. *Mol Cell* 56:55–66.

Auerbach AD, Allen RG. 1991. Leukemia and preleukemia in Fanconi anemia patients. A review of the literature and report of the International Fanconi Anemia Registry. *Cancer Genet Cytogenet* 51:1–12.

Baralle D, Lucassen A, Buratti E. 2009. Missed threads: The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep* 10:810–816.

Barry G, Briggs JA, Vanichkina DP, Poth EM, Beveridge NJ, Ratnu VS, Nayler SP, Nones K, Hu J, Bredy TW, Nakagawa S, Rigo F, et al. 2014. The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol Psychiatry* 19:486–494.

Beral V, Bull D, Doll R, Peto R, Reeves G. 2002. Breast cancer and breastfeeding: Collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease. *Lancet* 360:187–195.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of

adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* 74:3171–3175.

Beyer AL, Osheim YN. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev* 2:754–765.

Biamonti G, Maita L, Montecucco A. 2018. The Krebs Cycle Connection: Reciprocal Influence Between Alternative Splicing Programs and Cell Metabolism. *Front Oncol* 8:1–17.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72:291–336.

Blanco A, la Hoya M de, Balmaña J, Ramón Y Cajal T, Teulé A, Miramar MD, Esteban E, Infante M, Benítez J, Torres A, Tejada MI, Brunet J, et al. 2012. Detection of a large rearrangement in PALB2 in Spanish breast cancer families with male breast cancer. *Breast Cancer Res Treat* 132:307–315.

Blanco A, la Hoya M de, Osorio A, Diez O, Miramar MD, Infante M, Martinez-Bouzas C, Torres A, Lasa A, Llorca G, Brunet J, Graña B, et al. 2013. Analysis of PALB2 Gene in BRCA1/BRCA2 Negative Spanish Hereditary Breast/Ovarian Cancer Families with Pancreatic Cancer Cases. *PLoS One* 8:.

Blencowe BJ. 2000. Exonic splicing enhancers: Mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 25:106–110.

Bonnal SC, López-Oreja I, Valcárcel J. 2020. Roles and mechanisms of alternative splicing in cancer — implications for care. *Nat Rev Clin Oncol* 17:457–474.

Boonen RACM, Rodrigue A, Stoepker C, Wiegant WW, Vroliing B, Sharma M, Rother MB, Celosse N, Vreeswijk MPG, Couch F, Simard J, Devilee P, et al. 2019. Functional analysis of genetic variants in the high-risk breast cancer

susceptibility gene PALB2. *Nat Commun* 10:1–15.

Boonen RACM, Vreeswijk MPG, Attikum H van. 2020.

Briese M, Esmaeili B, Sattelle DB. 2005. Is spinal muscular atrophy the result of defects in motor neuron processes? *BioEssays* 27:946–957.

Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD, Kanavy DM, Luo X, McNulty SM, Starita LM, Tavtigian S V, Wright MW, et al. 2020. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework On behalf of the Clinical Genome Resource Sequence Variant Interpretation Working Group. *Genome Med* 12:1–12.

Brnich SE, Arteaga EC, Wang Y, Tan X, Berg JS. 2021. A Validated Functional Analysis of Partner and Localizer of BRCA2 Missense Variants for Use in Clinical Variant Interpretation. *J Mol Diagnostics* 23:847–864.

Broca P. 1866. *Traité des tumeurs*.

Brooks C, Soule D, Muel KI. 1973. Estrogen ( MCF Receptor from Breast in a Human Carcinoma \* Cell ( Received for publication , Michigan 48901 SUMMARY A stable cell line ( MCF-7 ) derived by pleural effusion from a breast cancer patient has been demonstrated to contain significant amounts . 6251–6253.

BRUWER A, BARGEN JA, KIERLAND RR. 1954. Surface pigmentation and generalized intestinal polyposis; (Peutz-Jeghers syndrome). *Proc Staff Meet Mayo Clin* 29:168–71.

Bryksin A V., Matsumura I. 2010. Overlap extension PCR cloning: A simple and reliable way to create recombinant plasmids. *Biotechniques* 48:463–465.

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 DF, Devilee P, et al. 2021. Rad51d aberrant splicing in breast cancer: Identification of splicing regulatory elements and minigene-based evaluation of 53 dna variants. *Cancers (Basel)* 13:.

Buisson R, Niraj J, Pauty J, Maity R, Zhao W, Coulombe Y, Sung P, Masson JY. 2014. Breast cancer proteins PALB2 and BRCA2 stimulate polymerase  $\eta$  in recombination-associated DNA Synthesis At Blocked Replication Forks. *Cell Rep* 6:553–564.

Burge C, Tuschl T, ., Sharp PA. 1999. Splicing of Precursors to mRNA by the Spliceosome. 303-345 p.

Burn TC, Connors TD, Klinger KW, Landes GM. 1995. Increased exon-trapping efficiency through modifications to the pSPL3 splicing vector. *Gene* 161:183–187.

Burset M, Seledtsov IA, Solovyev V V. 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res* 28:4364–4375.

Busch A, Hertel KJ. 2012. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip Rev RNA* 3:1–12.

Buyss SS, Sandbach JF, Gammon A, Patel G, Kidd J, Brown KL, Sharma L, Saam J, Lancaster J, Daly MB. 2017. A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. *Cancer* 123:1721–1730.

Caleca L, Catucci I, Figlioli G, Cecco L De, Pesaran T, Ward M, Volorio S, Falanga A, Marchetti M, Iacone M, Tondini C, Zambelli A, et al. 2018. Two missense

variants detected in breast cancer probands preventing BRCA2-PALB2 protein interaction. *Front Oncol* 8:.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat Rev Genet* 3:285–298.

Casadei S, Gulsuner S, Shirts BH, Mandell JB, Kortbawi HM, Norquist BS, Swisher EM, Lee MK, Goldberg Y, O'Connor R, Tan Z, Pritchard CC, et al. 2019. Characterization of splice-altering mutations in inherited predisposition to cancer. *Proc Natl Acad Sci U S A* 116:26798–26807.

Casadei S, Norquist BM, Walsh T, Stray S, Jessica B, Lee MK, Stamatoyannopoulos JA, King M. 2012. Mutations in the Brca2-Interacting Protein Palb2. *Cancer Res* 71:2222–2229.

Castéra L, Krieger S, Rousselin A, Legros A, Baumann JJ, Bruet O, Brault B, Fouillet R, Goardon N, Letac O, Baert-Desurmont S, Tinat J, et al. 2014. Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *Eur J Hum Genet* 22:1305–1313.

Catucci I, Peterlongo P, Ciceri S, Colombo M, Pasquini G, Barile M, Bonanni B, Verderio P, Pizzamiglio S, Foglia C, Falanga A, Marchetti M, et al. 2014. PALB2 sequencing in Italian familial breast cancer cases reveals a high-risk mutation recurrent in the province of Bergamo. *Genet Med* 16:688–694.

Cerasuolo A, Cammarota F, Duraturo F, Staiano A, Martinelli M, Miele E, Izzo P, Rosa M De. 2020. 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* 21:1–

12.

Chakarova CF, Hims MM, Bolz H, Abu-Safieh L, Patel RJ, Papaioannou MG, Inglehearn CF, Keen TJ, Willis C, Moore AT, Rosenberg T, Webster AR, et al. 2002. Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet* 11:87–92.

Chen J, Silver DP, Walpita D, Cantor SB, Gazdar AF, Tomlinson G, Couch FJ, Weber BL, Ashley T, Livingston DM, Scully R. 1998. Stable interaction between the products of the BRCA1 and BRCA2 tumor suppressor genes in mitotic and meiotic cells. *Mol Cell* 2:317–328.

Cho DH, Tapscott SJ. 2007. Myotonic dystrophy: Emerging mechanisms for DM1 and DM2. *Biochim Biophys Acta - Mol Basis Dis* 1772:195–204.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12:1–8.

Christofk HR, Heiden MG Vander, Harris MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL, Cantley LC. 2008. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* 452:230–233.

Churbanov A, Winters-Hilt S, Koonin E V., Rogozin IB. 2008. Accumulation of GC donor splice signals in mammals. *Biol Direct* 3:9–12.

Couch FJ, Rasmussen LJ, Hofstra R, Monteiro ANA, Greenblatt MS, Wind N de. 2008. Assessment of functional effects of unclassified genetic variants. *Hum Mutat* 29:1314–1326.

Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J, Hallberg E, Moore R,

Thomas A, Lilyquist J, Feng B, McFarland R, et al. 2017. Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncol* 3:1190–1196.

Crick F. 1970. Central Dogma of Molecular Biology. *Nature* 227:561–563.

David CJ, Manley JL. 2010. Alternative pre-mRNA splicing regulation in cancer: Pathways and programs unhinged. *Genes Dev* 24:2343–2364.

Denichenko P, Mogilevsky M, Cléry A, Welte T, Biran J, Shimshon O, Barnabas GD, Danan-Gotthold M, Kumar S, Yavin E, Levanon EY, Allain FH, et al. 2019. Specific inhibition of splicing factor activity by decoy RNA oligonucleotides. *Nat Commun* 10:.

Desterro J, Bak-Gordon P, Carmo-Fonseca M. 2020. Targeting mRNA processing as an anticancer strategy. *Nat Rev Drug Discov* 19:112–129.

Di C, Syafrizayanti, Zhang Q, Chen Y, Wang Y, Zhang X, Liu Y, Sun C, Zhang H, Hoheisel JD. 2019. Function, clinical application, and strategies of Pre-mRNA splicing in cancer. *Cell Death Differ* 26:1181–1194.

Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, Pooley KA, Parsons MT, Fortuno C, Wang Q, Bolla MK, Dennis J, et al. 2021. Breast Cancer Risk Genes — Association Analysis in More than 113,000 Women. *N Engl J Med* 384:428–439.

Ducy M, Sesma-Sanz L, Guitton-Sert L, Lashgari A, Gao Y, Brahiti N, Rodrigue A, Margailan G, Caron MC, Côté J, Simard J, Masson JY. 2019. The Tumor Suppressor PALB2: Inside Out. *Trends Biochem Sci* 44:226–240.

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, et al. 2012. An integrated

encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.

Dvinge H, Kim E, Abdel-wahab O, Bradley RK, Program B, Health P, Division S, Cancer FH, Division S, Hutchinson F, Program P, Sloan M, et al. 2016. 27282250 RNA splicing factors as oncoproteins and tumor suppressors *Nat Rev Cancer*. 2016 Jul.pdf. 16:413–430.

Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian S V., Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DGR, et al. 2015. Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* 372:2243–2257.

Erkko H, Dowty JG, Nikkilä J, Syrjäkoski K, Mannermaa A, Pylkäs K, Southey MC, Holli K, Kallioniemi A, Jukkola-Vuorinen A, Kataja V, Kosma VM, et al. 2008. Penetrance analysis of the PALB2 c.1592delIT founder mutation. *Clin Cancer Res* 14:4667–4671.

Erkko H, Xia B, Nikkilä J, Schleutker J, Syrjäkoski K, Mannermaa A, Kallioniemi A, Pylkäs K, Karppinen SM, Rapakko K, Miron A, Sheng Q, et al. 2007. A recurrent mutation in PALB2 in Finnish cancer families. *Nature* 446:316–319.

Ewing B, Green P. 2000. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet* 25:232–234.

Fackenthal JD, Yoshimatsu T, Zhang B, Garibay GR de, Colombo M, Vecchi G De, Ayoub SC, Lal K, Olopade OI, Vega A, Santamariña M, Blanco A, et al. 2016. Naturally occurring &em&BRCA2&/em& alternative mRNA splicing events in clinically relevant samples. *J Med Genet* 53:548 LP-558.

Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JWW, Comber H, Forman D, Bray F. 2013. Cancer incidence and mortality patterns

in Europe: Estimates for 40 countries in 2012. *Eur J Cancer* 49:1374–1403.

Fewings E, Larionov A, Redman J, Goldgraben MA, Scarth J, Richardson S, Brewer C, Davidson R, Ellis I, Evans DG, Halliday D, Izatt L, et al. 2018. Germline pathogenic variants in PALB2 and other cancer-predisposing genes in families with hereditary diffuse gastric cancer without CDH1 mutation: a whole-exome sequencing study. *Lancet Gastroenterol Hepatol* 3:489–498.

Fisher B, Costantino JP, Wickerham DL, Cecchini RS, Cronin WM, Robidoux A, Bevers TB, Kavanah MT, Atkins JN, Margolese RG, Runowicz CD, James JM, et al. 2005. Tamoxifen for the prevention of breast cancer: Current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J Natl Cancer Inst* 97:1652–1662.

Foulkes WD. 2008. Inherited Susceptibility to Common Cancers. *N Engl J Med* 359:2143–2153.

Foulkes WD. 2021. The ten genes for breast (and ovarian) cancer susceptibility. *Nat Rev Clin Oncol* 18:259–260.

Foulkes WD, Ghadirian P, Akbari MR, Hamel N, Giroux S, Sabbaghian N, Darnel A, Royer R, Poll A, Fafard E, Robidoux A, Martin G, et al. 2007. Identification of a novel truncating PALB2 mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. *Breast Cancer Res* 9:1–9.

Foulkes WD, Smith IE, Reis-Filho JS. 2010. Triple-Negative Breast Cancer. *N Engl J Med* 363:1938–1948.

Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V, Acedo A, Sanz DJ, Velasco EA. 2017. Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and

18. PLoS Genet 13:e1006691.

Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, Acedo A, Velasco EA. 2018. Identification of eight spliceogenic variants in BRCA2 Exon 16 by minigene assays. *Front Genet* 9:1–10.

Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, Goina E, Acedo A, Buratti E, Velasco EA. 2019. Mis-splicing in breast cancer: identification of pathogenic BRCA2 variants by systematic minigene assays. *J Pathol* 248:409–420.

Fu XD. 2004. Towards a splicing code. *Cell* 119:736–738.

Garibay GR de, Acedo A, García-Casado Z, Gutiérrez-Enríquez S, Tosar A, Romero A, Garre P, Llord G, Thomassen M, Díez O, Pérez-Segura P, Díaz-Rubio E, et al. 2014. Capillary Electrophoresis Analysis of Conventional Splicing Assays: IARC Analytical and Clinical Classification of 31 BRCA2 Genetic Variants. *Hum Mutat* 35:53–57.

Gatchel JR, Zoghbi HY. 2005. Diseases of unstable repeat expansion: mechanisms and common principles. *Nat Rev Genet* 6:743–755.

Geenen JJJ, Linn SC, Beijnen JH, Schellens JHM. 2018. PARP Inhibitors in the Treatment of Triple-Negative Breast Cancer. *Clin Pharmacokinet* 57:427–437.

Geuens T, Bouhy D, Timmerman V. 2016. The hnRNP family: insights into their role in health and disease. *Hum Genet* 135:851–67.

Gilbert W. 1978. Why genes in pieces? *Nature* 271:501–501.

Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS. 2008. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum Mutat* 29:1265–1272.

Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro ANA, Tavtigian S V, Couch FJ, Information C, Bic C. 2004. <AJHGv75p535.pdf>. 535–544.

Gueiderikh A, Rosselli F, Neto JBC. 2017. A never-ending story: The steadily growing family of the FA and FA-like genes. *Genet Mol Biol* 40:398–407.

Hall SL, Padgett RA. 1994.

Hemminki A, Markie D, Tomlinson I, Avizienyte E, Roth S, Loukola A, Bignell G, Warren W, Aminoff M, Höglund P, Järvinen H, Kristo P, et al. 1998. A serine/threonine kinase gene defective in Peutz–Jeghers syndrome. *Nature* 391:184–187.

Hofstatter EW, Domchek SM, Miron A, Garber J, Wang M, Componeschi K, Boghossian L, Miron PL, Nathanson KL, Tung N. 2011. PALB2 mutations in familial breast and pancreatic cancer. *Fam Cancer* 10:225–231.

Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, Gao C, Lilyquist J, Yadav S, Boddicker NJ, Samara R, Klebba J, et al. 2021. A Population-Based Study of Genes Previously Implicated in Breast Cancer. *N Engl J Med* 384:440–451.

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, et al. 2002. The Ensembl genome database project. *Nucleic Acids Res* 30:38–41.

Irigoyen MAA, García FV, Iturriagagoitia AC, Beroiz BI, Martínez MS, Grima FG. 2011. Subtipos moleculares del cáncer de mama : implicaciones pronósticas y características clínicas e inmunohistoquímicas Molecular subtypes of breast cancer : prognostic implications. *An Sist Sanit Navar* 34:219–233.

Jackson LJ. 1991. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res* 19:3795–3798.

Janknecht R. 2005. EWS-ETS oncoproteins: The linchpins of Ewing tumors. *Gene* 363:1–14.

Jeong S. 2017. SR Proteins: Binders, Regulators, and Connectors of RNA. *Mol Cells* 40:1.

Joenje H, Patel KJ. 2001. The emerging genetic and molecular basis of Fanconi anaemia. *Nat Rev Genet* 2:446–457.

Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, Lin JCH, Palmisano E, Brune K, Jaffee EM, Iacobuzio-Donahue CA, Maitra A, et al. 2009. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* (80- ) 324:217.

Jozwiak J, Jozwiak S, Wlodarski P. 2008. Possible mechanisms of disease development in tuberous sclerosis. *Lancet Oncol* 9:73–79.

Jung H, Lee D, Lee J, Park D, Kim YJ, Park WY, Hong D, Park PJ, Lee E. 2015. Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat Genet* 47:1242–1248.

Jurica MS, Moore MJ. 2003. Pre-mRNA Splicing. *Mol Cell* 12:5–14.

Karni R, Stanchina E de, Lowe SW, Sinha R, Mu D, Krainer AR. 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* 14:185–193.

Kastner B, Will CL, Stark H, Lührmann R. 2019. Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harb Perspect Biol* 11:.

Kervestin S, Jacobson A. 2012. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol* 13:700–12.

Knudson AG. 1971. Mutation and Cancer: Statistical Study of Retinoblastoma.

Proc Natl Acad Sci 68:820–823.

König H, Matter N, Bader R, Thiele W, Müller F. 2007. Splicing Segregation: The Minor Spliceosome Acts outside the Nucleus and Controls Cell Proliferation. *Cell* 131:718–729.

Kozlovski I, Siegfried Z, Amar-Schwartz A, Karni R. 2017. The role of RNA alternative splicing in regulating cancer metabolism. *Hum Genet* 136:1113–1127.

Kralovicova J, Hwang G, Asplund a. C, Churbanov A, Smith CIE, Vorechovsky I. 2011. Compensatory signals associated with the activation of human GC 5' splice sites. *Nucleic Acids Res* 39:7077–7091.

Krawczak M, Reiss J, Cooper DN. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Hum Genet* 90:41–54.

Krush AJ. 1979. Contributions of Pierre Paul Broca to Cancer Genetics. *Trans Nebraska Acad Sci Affil Soc VII:125–129.*

Kurian AW, Sigal BM, Plevritis SK. 2010. Survival analysis of cancer risk reduction strategies for BRCA1/2 mutation carriers. *J Clin Oncol* 28:222–231.

la Hoya M de, Gutiérrez-Enríquez S, Velasco E, Osorio A, Sanchez De Abajo A, Vega A, Salazar R, Esteban E, Llorca G, Gonzalez-Sarmiento R, Carracedo A, Benítez J, et al. 2006. Genomic rearrangements at the BRCA1 locus in Spanish families with breast/ovarian cancer. *Clin Chem* 52:1480–1485.

Ladomery M. 2013. Aberrant Alternative Splicing Is Another Hallmark of Cancer. *Int J Cell Biol* 2013:1–6.

Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, Malafosse A,

Antonarakis SE. 1997.

Landrith T, Li B, Cass AA, Conner BR, LaDuca H, McKenna DB, Maxwell KN, Domchek S, Morman NA, Heinlen C, Wham D, Koptiuch C, et al. 2020. Splicing profile by capture RNA-seq identifies pathogenic germline variants in tumor suppressor genes. *npj Precis Oncol* 4:.

Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, Hoffman D, Jang W, Kaur K, Liu C, Lyoshin V, Maddipatla Z, et al. 2020. ClinVar: Improvements to accessing data. *Nucleic Acids Res* 48:D835–D844.

Lee K, Krempely K, Roberts ME, Anderson MJ, Carneiro F, Chao E, Dixon K, Figueiredo J, Ghosh R, Huntsman D, Kaurah P, Kesserwan C, et al. 2018. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Hum Mutat* 39:1553–1568.

Lee Y, Rio DC. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* 84:291–323.

Lenoir G. 1991. Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet* 338:82–83.

Li FP, Fraumeni JF, Mulvihill JJ, Blattner WA, Dreyfus MG, Tucker MA, Miller RW. 1988. A Cancer Family Syndrome in Twenty-four Kindreds. *Cancer Res* 48:5358–5362.

Li S-C, Goto NK, Williams KA, Deber CM, Fasman GD. 1996. alpha-Helical, but not beta-3-sheet, propensity of proline is determined by peptide environment. *Proc Natl Acad Sci USA* 93:6676–6681.

Liaw D, Marsh DJ, Li J, Dahia PLM, Wang SI, Zheng Z, Bose S, Call KM, Tsou HC, Peacocke M, Eng C, Parsons R. 1997. Germline mutations of the PTEN gene

in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat Genet* 16:64–67.

Lin JH, Tang XY, Boulling A, Zou W Bin, Masson E, Fichou Y, Raud L, Tertre M Le, Deng SJ, Berlivet I, Ka C, Mort M, et al. 2019. First estimate of the scale of canonical 5' splice site GT>GC variants capable of generating wild-type transcripts. *Hum Mutat* 40:1856–1873.

Liu X, Zhou Y, Lou Y, Zhong H. 2016. Knockdown of HNRNPA1 inhibits lung adenocarcinoma cell proliferation through cell cycle arrest at G0/G1 phase. *Gene* 576:791–797.

Lloyd KM, Hospital HF. 2017. CASE STUDIES Cowden ' s Disease.

Lobitz S, Velleuer E. 2006. Guido Fanconi (1892-1979): A jack of all trades. *Nat Rev Cancer* 6:893–898.

Loh TJ, Moon H, Cho S, Jang H, Liu YC, Tai H, Jung DW, Williams DR, Kim HR, Shin MG, Liao DJ, Zhou J, et al. 2015. CD44 alternative splicing and hnRNP A1 expression are associated with the metastasis of breast cancer. *Oncol Rep* 34:1231–1238.

López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. 2005. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett* 579:1900–1903.

Lopez-Perolio I, Leman R, Behar R, Lattimore V, Pearson JF, Castéra L, Martins A, Vaur D, Goardon N, Davy G, Garre P, García-Barberán V, et al. 2019. Alternative splicing and ACMG-AMP-2015-based classification of PALB2 genetic variants: An ENIGMA report. *J Med Genet* 56:453–460.

Luijsterburg MS, Typas D, Caron MC, Wiegant WW, Heuvel D Van Den, Boonen

RA, Couturier AM, Mullenders LH, Masson JY, Attikum H Van. 2017. A PALB2-interacting domain in RNF168 couples homologous recombination to DNA break-induced chromatin ubiquitylation. *Elife* 6:1–26.

Ma J, Cai H, Wu T, Sobhian B, Huo Y, Alcivar A, Mehta M, Cheung KL, Ganesan S, Kong A-NT, Zhang DD, Xia B. 2012. PALB2 Interacts with KEAP1 To Promote NRF2 Nuclear Accumulation and Function. *Mol Cell Biol* 32:1506–1517.

Ma SL, Vega-Warner V, Gillies C, Sampson MG, Kher V, Sethi SK, Otto EA. 2015. Whole exome sequencing reveals novel PHEX splice site mutations in patients with hypophosphatemic rickets. *PLoS One* 10:1–12.

Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA, Friend SH. 1990. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* (80- ) 250:1233–1238.

Mamrak NE, Shimamura A, Howlett NG. 2017. Recent discoveries in the molecular pathogenesis of the inherited bone marrow failure syndrome Fanconi anemia. *Blood Rev* 31:93–99.

Maniatis T, Tasic B. 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* 418:236–243.

Manley JL, Krainer AR. 2010. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev* 24:1073–1074.

Mau C, Untch M. 2017. Prophylactic Surgery: For Whom, When and How? *Breast Care* 12:379–384.

McKie AB, McHale JC, Keen TJ, Tarttelin EE, Goliath R, Lith-Verhoeven JJC Van, Greenberg J, Ramesar RS, Hoyng CB, Cremers FPM, Mackey DA, Bhattacharya

SS, et al. 2001. Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet* 10:1555–1562.

Merkl PE, Pils M, Fremter T, Schwank K, Engel C, Längst G, Milkereit P, Griesenbeck J, Tschochner H. 2020. RNA polymerase i (Pol I) passage through nucleosomes depends on Pol i subunits binding its lobe structure. *J Biol Chem* 295:4782–4795.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266:66–71.

Montalban G, Fraile-Bethencourt E, López-Perolio I, Pérez-Segura P, Infante M, Durán M, Alonso-Cerezo MC, López-Fernández A, Diez O, la Hoya M de, Velasco EA, Gutiérrez-Enríquez S. 2018. Characterization of spliceogenic variants located in regions linked to high levels of alternative splicing: BRCA2 c.7976+5G > T as a case study. *Hum Mutat* 39:1155–1160.

Mordes D, Luo X, Kar A, Kuo D, Xu L, Fushimi K, Yu G, Sternberg P, Wu JY. 2006. Pre-mRNA splicing and retinitis pigmentosa. *Mol Vis* 12:1259–1271.

Mount SM. 2000. Genomic sequence, splicing, and gene annotation. *Am J Hum Genet* 67:788–792.

Narod SA. 2021. Which Genes for Hereditary Breast Cancer? *N Engl J Med* 384:471–473.

Nepomuceno TC, Carvalho MA, Rodrigue A, Simard J, Masson J-Y, Monteiro ANA. 2021. PALB2 Variants: Protein Domains and Cancer Susceptibility. *Trends in cancer* 7:188–197.

Nepomuceno TC, Gregoriis G De, Oliveira FMB de, Suarez-Kurtz G, Monteiro AN, Carvalho MA. 2017. The role of PALB2 in the DNA damage response and cancer predisposition. *Int J Mol Sci* 18:1–20.

Nielsen FC, Overeem Hansen T Van, Sørensen CS. 2016. Hereditary breast and ovarian cancer: New genes in confined pathways. *Nat Rev Cancer* 16:599–612.

Oliver AW, Swift S, Lord CJ, Ashworth A, Pearl LH. 2017. Structural basis for recruitment of BRCA2 by PALB2. *EMBO Rep* 18:1264–1264.

Oltean S, Bates DO. 2014. Hallmarks of alternative splicing in cancer. *Oncogene* 33:5311–5318.

Osborne RJ, Thornton CA. 2006. RNA-dominant diseases. *Hum Mol Genet* 15:162–169.

Pagani F, Raponi M, Baralle FE. 2005. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc Natl Acad Sci U S A* 102:6368–6372.

Pagenstecher C, Wehner M, Friedl W, Rahner N, Aretz S, Friedrichs N, Sengteller M, Henn W, Buettner R, Propping P, Mangold E. 2006. Aberrant splicing in MLH1 and MSH2 due to exonic and intronic variants. *Hum Genet* 119:9–22.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40:1413–1415.

Papasaikas P, Valcárcel J. 2016. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci* 41:33–45.

Park JY, Singh TR, Nassar N, Zhang F, Freund M, Hanenberg H, Meetei AR,

Andreassen PR. 2014. Breast cancer-associated missense mutants of the PALB2 WD40 domain, which directly binds RAD51C, RAD51 and BRCA2, disrupt DNA repair. *Oncogene* 33:4803–4812.

Pauty J, Couturier AM, Rodrigue A, Caron MC, Coulombe Y, Dellaire G, Masson JY. 2017. Cancer-causing mutations in the tumor suppressor PALB2 reveal a novel cancer mechanism using a hidden nuclear export signal in the WD40 repeat motif. *Nucleic Acids Res* 45:2644–2657.

Perou CM, Sørile T, Eisen MB, Rijn M Van De, Jeffrey SS, Resh CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge Ø, Pergammenschikov A, et al. 2000. Molecular portraits of human breast tumours. *Nature* 406:747–752.

Pickrell JK, Pai AA, Gilad Y, Pritchard JK. 2010. Noisy Splicing Drives mRNA Isoform Diversity in Human Cells. *PLoS Genet* 6:e1001236.

Ramus SJ, Song H, Dicks E, Tyrer JP, Rosenthal AN, Intermaggio MP, Fraser L, Gentry-Maharaj A, Hayward J, Philpott S, Anderson C, Edlund CK, et al. 2015. Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer. *J Natl Cancer Inst* 107:1–8.

Ranum LPW, Cooper TA. 2006. RNA-mediated neuromuscular disorders. *Annu Rev Neurosci* 29:259–277.

Rebbeck TR, Mitra N, Wan F, Sinilnikova OM, Healey S, McGuffog L, Chenevix-Trench G, Easton DF, Antoniou AC, Nathanson KL, Laitman Y, Kushnir A, et al. 2015. Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. *JAMA - J Am Med Assoc* 313:1347–1361.

Reddy R, Busch H. 1988. Small Nuclear RNAs: RNA Sequences, Structure, and Modifications. *Struct Funct Major Minor Small Nucl Ribonucleoprotein Part 1–*

37.

Reid S, Schindler D, Hanenberg H, Barker K, Hanks S, Kalb R, Neveling K, Kelly P, Seal S, Freund M, Wurm M, Batish SD, et al. 2007. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat Genet* 39:162–164.

Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE. 2008. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med* 10:294–300.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. 2015. 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* 17:405–424.

Rilem. 2020. Test results. RILEM Tech Recomm Test use Constr Mater 29:397–399.

Rodrigue A, Margaillan G, Torres Gomes T, Coulombe Y, Montalban G, Costa E Silva Carvalho S da, Milano L, Ducy M, De-Gregoriis G, Dellaire G, Araújo da Silva W, Monteiro AN, et al. 2019. A global functional analysis of missense mutations reveals two major hotspots in the PALB2 tumor suppressor. *Nucleic Acids Res* 47:10662–10677.

Ryu JS, Lee HY, Cho EH, Yoon KA, Kim MK, Joo J, Lee ES, Kang HS, Lee S, Lee DO, Lim MC, Kong SY. 2020. Exon splicing analysis of intronic variants in multigene cancer panel testing for hereditary breast/ovarian cancer. *Cancer Sci* 111:3912–3925.

Sanoguera-Miralles L, Valenzuela-Palomo A, Bueno-Martínez E, Llovet P, Díez-Gómez B, Caloca MJ, Pérez-Segura P, Fraile-Bethencourt E, Colmena M, Carvalho S, Allen J, Easton DF, et al. 2020. Comprehensive functional characterization and clinical interpretation of 20 splice-site variants of the rad51c gene. *Cancers (Basel)* 12:1–21.

Sanz DJ, Acedo A, Infante M, Durán M, Pérez-Cabornero L, Esteban-Cardenosa E, Lastra E, Pagani F, Miner C, Velasco EA. 2010. A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin Cancer Res* 16:1957–67.

Scotti MM, Swanson MS. 2016. RNA mis-splicing in disease. *Nat Rev Genet* 17:19–32.

Sebestyén E, Singh B, Miñana B, Pagès A, Mateo F, Pujana MA, Valcárcel J, Eyrales E. 2016. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res* 26:732–744.

Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, Smith PG, Buonamici S, Yu L, Caesar-Johnson SJ, Demchok JA, Felau I, et al. 2018. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep* 23:282–296.e4.

Shepard PJ, Hertel KJ. 2009. The SR protein family. *Genome Biol* 2009 1010 10:1–9.

Shimelis H, LaDuca H, Hu C, Hart SN, Na J, Thomas A, Akinhanmi M, Moore RM, Brauch H, Cox A, Eccles DM, Ewart-Toland A, et al. 2018. Triple-negative breast cancer risk genes identified by multigene hereditary cancer panel

testing. *J Natl Cancer Inst* 110:855–862.

Siegfried Z, Karni R. 2018. The role of alternative splicing in cancer drug resistance. *Curr Opin Genet Dev* 48:16–21.

Slater EP, Langer P, Niemczyk E, Strauch K, Butler J, Habbe N, Neoptolemos JP, Greenhalf W, Bartsch DK. 2010. PALB2 mutations in European familial pancreatic cancer families. *Clin Genet* 78:490–494.

Smith CWJ, Valcárcel J. 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* 25:381–388.

Song F, Li M, Liu G, Swapna GVT, Daigham NS, Xia B, Montelione GT, Bunting SF. 2018. Antiparallel Coiled-Coil Interactions Mediate the Homodimerization of the DNA Damage-Repair Protein PALB2. *Biochemistry* 57:6581–6591.

Spurdle AB, Healey S, Devereau A, Hogervorst FBL, Monteiro ANA, Nathanson KL, Radice P, Stoppa-Lyonnet D, Tavtigian S, Wappenschmidt B, Couch FJ, Goldgar DE. 2012. ENIGMA-evidence-based network for the interpretation of germline mutant alleles: An international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum Mutat* 33:2–7.

Stadler ZK, Salo-Mullen E, Sabbaghian N, Simon JA, Zhang L, Olson SH, Kurtz R, Offit K, Foulkes WD, Robson ME, Tischkowitz M. 2011. Germline PALB2 mutation analysis in breast-pancreas cancer families. *J Med Genet* 48:523–525.

Steitz JA, Dreyfuss G, Krainer AR, Lamond AI, Matera AG, Padgett RA. 2008. Where in the cell is the minor spliceosome? *Proc Natl Acad Sci U S A* 105:8485–8486.

Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. 2016. Aberrant RNA splicing in cancer; Expression changes and driver mutations of splicing factor genes. *Oncogene* 35:2413–2427.

Sy SMH, Huen MSY, Chen J. 2009a. MRG15 is a novel PALB2-interacting factor involved in homologous recombination. *J Biol Chem* 284:21127–21131.

Sy SMH, Huen MSY, Chen J. 2009b. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proc Natl Acad Sci* 106:7155–7160.

Tavtigian S V, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, Biesecker LG. 2018. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med* 20:1054–1060.

Thanaraj TA, Clark F. 2001. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Res* 29:2581–2593.

Tischkowitz M, Balmaña J, Foulkes WD, James P, Ngeow J, Schmutzler R, Voian N, Wick MJ, Stewart DR, Pal T. 2021. Management of individuals with germline variants in PALB2: a clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* 23:1416–1423.

Tischkowitz M, Xia B. 2010. PALB2/FANCN: Recombining cancer and fanconi anemia. *Cancer Res* 70:7353–7359.

Urbanski LM, Leclair N, Anczuków O. 2018. Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip Rev RNA* 9:e1476.

Valdivia HH. 2007. One gene, many proteins: Alternative splicing of the

ryanodine receptor gene adds novel functions to an already complex channel protein. *Circ Res* 100:761–763.

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, et al. 2021. Splicing predictions, minigene analyses and ACMG/AMP clinical classification of 42 germline PALB2 splice-site variants. *J Pathol* 27:200–206.

Vallée MP, Sera TL Di, Nix DA, Paquette AM, Parsons MT, Bell R, Hoffman A, Hogervorst FBL, Goldgar DE, Spurdle AB, Tavtigian S V. 2016. Adding In Silico Assessment of Potential Splice Aberration to the Integrated Evaluation of BRCA Gene Unclassified Variants. *Hum Mutat* 37:627–639.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, et al. 2001. Celera\_genoma. *Science* (80- ) 291:1–49.

Verlander P, Kaporis A, Liu Q, Zhang Q, Seligsohn U, Auerbach A. 1995. Carrier frequency of the IVS4 + 4 A--&T mutation of the Fanconi anemia gene FAC in the Ashkenazi Jewish population. *Blood* 86:4034–4038.

Wahl MC, Will CL, Lührmann R. 2009. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* 136:701–718.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456:470–6.

Wang GS, Cooper TA. 2007. Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8:749–761.

Wang Z, Burge CB. 2008. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* 14:802–813.

Weatheritt RJ, Sterne-Weiler T, Blencowe BJ. 2016. The ribosome-engaged landscape of alternative splicing. *Nat Struct Mol Biol* 23:1117–1123.

Wilkinson ME, Charenton C, Nagai K. 2020. RNA Splicing by the Spliceosome. *Annu Rev Biochem* 89:359–388.

Will CL, Luhrmann R. 2011. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* 3:a003707–a003707.

Willis IM, Moir RD. 2018. Signaling to and from the RNA Polymerase III Transcription and Processing Machinery. *Annu Rev Biochem* 87:75–100.

Wiltshire T, Ducey M, Foo TK, Hu C, Lee KY, Belur Nagaraj A, Rodrigue A, Gomes TT, Simard J, Monteiro ANA, Xia B, Carvalho MA, et al. 2020. Functional characterization of 84 PALB2 variants of uncertain significance. *Genet Med* 22:622–632.

Winkler C, Eggert C, Gradl D, Meister G, Giegerich M, Wedlich D, Laggenbauer B, Fischer U. 2005. Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. *Genes Dev* 19:2320–2330.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G, Barfoot R, Hamoudi R, et al. 1995. Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378:789–792.

Wu Q, Krainer AR. 1997. Splicing of a divergent subclass of AT-AC introns requires the major spliceosomal snRNAs. *RNA* 3:586–601.

Wu Q, Krainer AR. 1999. AT-AC Pre-mRNA Splicing Mechanisms and

Conservation of Minor Introns in Voltage-Gated Ion Channel Genes. *Mol Cell Biol* 19:3225–3236.

Xia B, Dorsman JC, Ameziane N, Vries Y De, Rooimans MA, Sheng Q, Pals G, Errami A, Gluckman E, Llera J, Wang W, Livingston DM, et al. 2007. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. *Nat Genet* 39:159–161.

Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N, Liu X, Jasin M, Couch FJ, Livingston DM. 2006. Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2. *Mol Cell* 22:719–729.

Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, Dunning AM, Redman J, Scarth J, Plaskocinska I, Luccarini C, Shah M, et al. 2020. Cancer risks associated with germline PALB2 pathogenic variants: An international study of 524 families. *J Clin Oncol* 38:674–685.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11:377–394.

Zhang F, Fan Q, Ren K, Andreassen PR. 2009a. PALB2 Functionally Connects the Breast Cancer Susceptibility Proteins BRCA1 and BRCA2. *Mol Cancer Res* 7:1110–1118.

Zhang F, Ma J, Wu J, Ye L, Cai H, Xia B, Yu X. 2009b. PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response. *Curr Biol* 19:524–529.





# Anexos

**Tabla S1. Clasificación clínica ACMG/AMP de 42 variantes de *PALB2* detectadas en la cohorte BRIDGES.**

c.HGVS <sup>1</sup>	p.HGVS <sup>1</sup>	p.HGVS NM_024675.3	BRIDGES BC casos (N=60471)	BRIDGES Controles (N=53461)	Variantes
<b>c.47A&gt;G</b>	p.(Lys16Arg)	NP_078951.2:p.Lys16Arg	2	0	chr16_23652432_T_C
<b>c.48G&gt;A</b>	p.(Lys16=)	NP_078951.2:p.Lys16%3D	1	2	chr16_23652431_C_T
<b>c.48+1G&gt;A</b>	p.?	NA	0	0	chr16_23652430_C_T
<b>c.48+2T&gt;C</b>	p.?	NA	2	3	chr16_23652429_A_G
c.48+4C>T	p.?	NA	1	0	chr16_23652427_G_A
c.48+7G>C	p.?	NA	3	0	chr16_23652424_C_G
c.49-2A>T	p.?	NA	1	1	chr16_23649452_T_A
<b>c.108+1G&gt;A</b>	p.?	NA	31	6	chr16_23649390_C_T
c.108+2T>C	p.?	NA	1	0	chr16_23649389_A_G
c.109-6_109-4del	p.?	NA	0	0	chr16_23649276_AGAG_A
<b>c.109-2A&gt;G</b>	p.?	NA	0	1	chr16_23649275_T_C
c.211+1G>A	p.?	NA	4	1	chr16_23649170_C_T

c.211+5del	p.?	NA	0	1	chr16_23649165_AT_A
c.1684+4A>G	p.?	NA	2	0	chr16_23646179_T_C
<b>c.1685-2A&gt;C</b>	p.?	NA	0	0	chr16_23641792_T_G
<b>c.1685-2A&gt;G</b>	p.?	NA	1	0	chr16_23641792_T_C
c.2513A>C	p.(Gln838Pro)	NP_078951.2:p.Gln838Pro	1	0	chr16_23640962_T_G
<b>c.2515-2A&gt;G</b>	p.?	NA	0	0	chr16_23640598_T_C
c.2586+4A>T	p.?	NA	0	1	chr16_23640521_T_A
<b>c.2587-2A&gt;G</b>	p.?	NA	3	2	chr16_23637720_T_C
<b>c.2748+1G&gt;T</b>	p.?	NA	0	0	chr16_23637556_C_A
<b>c.2748+1G&gt;A</b>	p.?	NA	1	0	chr16_23637556_C_T
<b>c.2748+2dup</b>	p.?	NA	1	0	chr16_23637554_T_TA
c.2748+4A>T	p.?	NA	0	1	chr16_23637553_T_A
<b>c.2749-1G&gt;T</b>	p.?	NA	0	1	chr16_23635416_C_A
c.2750T>C	p.(Val917Ala)	NP_078951.2:p.Val917Ala	2	0	chr16_23635414_A_G
c.2834G>C	p.(Arg945Thr)	NP_078951.2:p.Arg945Thr	0	1	chr16_23635330_C_G
<b>c.2834+1G&gt;A</b>	p.?	NA	1	0	chr16_23635329_C_T
<b>c.2834+2T&gt;C</b>	p.?	NA	1	0	chr16_23635328_A_G
c.2834+3A>G	p.?	NA	2	1	chr16_23635327_T_C

c.2834+5G>A	p.?	NA	3	2	chr16_23635325_C_T
c.2834+6T>C	p.?	NA	2	0	chr16_23635324_A_G
c.2996+4A>G	p.?	NM_024675.3:c.2996+4A>G	1	0	chr16_23634286_T_C
<b>c.2997-2del</b>	p.?	--	--	--	Not in the analysis
<b>c.2997-1G&gt;A</b>	p.?	NA	0	0	chr16_23632800_C_T
<b>c.3113G&gt;A</b>	p.(Trp1038*)	NP_078951.2:p.Trp1038Ter	1	0	chr16_23632683_C_T
<b>c.3113+3A&gt;G</b>	p.?	NA	1	0	chr16_23632680_T_C
<b>c.3113+5G&gt;C</b>	p.?	NA	0	0	chr16_23632678_C_G
<b>c.3201+1G&gt;A</b>	p.?	NA	0	1	chr16_23625324_C_T
<b>c.3201+3_+4insTG</b>	p.?	NM_024675.3:c.3201+1_3201+2dup	1	0	chr16_23625322_T_TAC
c.3201+6T>A	p.?	NA	1	0	chr16_23625319_A_T
c.3348C>T	p.(Gly1116=)	NP_078951.2:p.Gly1116%3D	1	0	chr16_23619187_G_A
<b>c.3350+4A&gt;G</b>	p.?	NA	1	0	chr16_23619181_T_C
<b>c.3350+5G&gt;A</b>	p.?	NA	0	1	chr16_23619180_C_T

<sup>1</sup>NM\_024675.4;

**Tabla S2. Clasificación clínica ACMG-AMP de 42 variantes genéticas de *PALB2* detectadas en la cohorte BRIDGES.**

c.HGVS <sup>1</sup>	p.HGVS <sup>1</sup>	ClinVar <sup>2</sup>	clasificación ACMG-AMP <sup>3</sup>	PVS1 <sup>4</sup>	PS3/BS3 <sup>5</sup> (mgPALB2)	PS4 <sup>6</sup>	PM2 <sup>7</sup>	PM3 <sup>8</sup>	PM4 <sup>9</sup>	PP3/BP4 <sup>10</sup> (splicing)	PP3/BP4 <sup>11</sup> (proteína)	BP7 <sup>12</sup>
c.47A>G	p.(Lys16Arg)	VUS	LP		PS3_VS		PM2_P			PP3		
c.48G>A	p.(Lys16=)	VUS/LP/P	LP		PS3_VS		PM2_P		(PM4)	PP3		
c.48+1G>A	p.?	LP	LP	PVS1_M	PS3_VS		PM2_P		(PM4)			
c.48+2T>C	p.?	(-)	LP	PVS1_M	PS3_VS		PM2_P					
c.48+4C>T	p.?	VUS	LB		BS3		PM2_P				BP4	
c.48+7G>C	p.?	B/LB/VUS	LB		BS3						BP4	
c.49-2A>T	p.?	VUS/LP	VUS	PVS1_M	PS3_M		PM2_P		(PM4)			
c.108+1G>A	p.?	LP	P	PVS1	PS3_VS	PS4	PM2_P					
c.108+2T>C	p.?	LP	VUS	N/A	N/A		PM2_P					
c.109-6_109-4del	p.?	(-)	LB		BS3		PM2_P				BP4	
c.109-2A>G	p.?	LP	LP	PVS1	PS3_VS		PM2_P					
c.211+1G>A	p.?	LP	VUS	PVS1_P	PS3_P		PM2_P					
c.211+5del	p.?	VUS	LB		BS3		PM2_P				BP4	
c.1684+4A>G	p.?	(-)	VUS		N/A		PM2_P				BP4	
c.1685-2A>C	p.?	VUS	LP	PVS1	PS3_VS		PM2_P					
c.1685-2A>G	p.?	VUS/LP	LP	PVS1	PS3_VS		PM2_P					
c.2513A>C	p.(Gln838Pro)	(-)	VUS		N/A		PM2_P			PP3	(BP4_M)	
c.2515-2A>G	p.?	(-)	LP	N/A	PS3_VS		PM2_P					
c.2586+4A>T	p.?	VUS	LB		BS3		PM2_P				BP4	
c.2587-2A>G	p.?	(-)	LP	PVS1	PS3_VS		PM2_P					
c.2748+1G>T	p.?	LP/P	LP	PVS1	PS3_VS		PM2_P					
c.2748+2dup	p.?	VUS	LP		PS3_VS		PM2_P			PP3		
c.2748+4A>T	p.?	(-)	LB		BS3		PM2_P			PP3		
c.2749-1G>T	p.?	LP	LP	PVS1	PS3_VS		PM2_P		(PM4)			
c.2750T>C	p.(Val917Ala)	VUS	LB		BS3_M		PM2_P			n/a	(BP4_M)	

<b>c.2834G&gt;C</b>	p.(Arg945Thr)	VUS	VUS		N/A		PM2_P			PP3	(BP4_M)	
<b>c.2834+1G&gt;A</b>	p.?	LP/P	LP	PVS1	PS3_VS		PM2_P					
<b>c.2834+3A&gt;G</b>	p.?	LB/VUS	VUS		N/A		PM2_P			PP3		
<b>c.2834+5G&gt;A</b>	p.?	VUS	VUS		N/A					PP3		
<b>c.2834+6T&gt;C</b>	p.?	(-)	LB		BS3		PM2_P			BP4		
<b>c.2996+4A&gt;G</b>	p.?	(-)	VUS		N/A		PM2_P			PP3		
<b>c.2997-2del</b>	p.?	(-)	LP	PVS1	PS3_VS		PM2_P					
<b>c.2997-1G&gt;A</b>	p.?	VUS/LP	LP	PVS1	PS3_VS		PM2_P					
<b>c.3113G&gt;A</b>	p.(Trp1038*)	LP/P	P		PS3_VS	PS4				PP3		
<b>c.3113+3A&gt;G</b>	p.?	(-)	LP		PS3_VS		PM2_P			BP4		
<b>c.3113+5G&gt;C</b>	p.?	LP	P		PS3_VS		PM2_P	PM3		PP3		
<b>c.3201+1G&gt;A</b>	p.?	(-)	LP	PVS1	PS3_VS		PM2_P					
<b>c.3201+3_+4insTG</b>	p.?	(-)	LP		PS3_VS		PM2_P			PP3		
<b>c.3201+6T&gt;A</b>	p.?	VUS	VUS		N/A		PM2_P			PP3		
<b>c.3348C&gt;T</b>	p.(Gly1116=)	VUS	VUS		N/A		PM2_P			BP4		(~BP7)
<b>c.3350+4A&gt;G</b>	p.?	P/LP	P		PS3_VS		PM2_P	PM3		BP4		
<b>c.3350+5G&gt;A</b>	p.?	VUS/LP	P		PS3_VS		PM2_P	PM3		BP4		

<sup>1</sup>NM\_024675.4. <sup>2</sup>ClinVar último acceso el 14/10/2021. LB (probablemente benigna), VUS (variante de significado incierto), LP (probablemente patológica), P (patológica), (-) no informado. <sup>3</sup> Usamos una aproximación del sistema de puntos ACMG / AMP para combinar todas las evidencias patológicas y benignas. Para cada una de las variantes *PALB2* ensayadas, evaluamos los 16 códigos ACMG / AMP patológicos y 12 benignos. La tabla muestra solo los códigos patológicos y benignos que contribuyeron a la clasificación final. <sup>4</sup> Las fortalezas del código de PVS1 se basan en las predicciones de SpliceAI y una adaptación *PALB2* de las recomendaciones del árbol de decisión de ClinGen-SVI PVS1. Más tarde, usamos un algoritmo ad-hoc (basado en el juicio de expertos) para combinar las evidencias de fuerza general del código PS3 / BS3. (N/A) Los resultados con minigenes no aportaron datos concluyentes. <sup>6</sup> Para cada una de las variantes estudiadas, analizamos los datos de casos y controles de cáncer de mama de BRIDGES. Solo c.108+1G> A alcanzó una evidencia de PS4. Además, buscamos en la literatura científica análisis de casos y controles adicionales. Un estudio apoyó PS4 para c.3113G>A. <sup>7</sup> Asignamos el código PS2 degradado (PS2\_P), a variantes con un recuento ≤1 (en 100.000 alelos globales de gnomADv2.1). <sup>8</sup> Asignamos PM3 a variantes identificadas en trans con una variante patológica en pacientes con anemia de Fanconi. <sup>9</sup> No asignamos PM4 a variantes específicas, sino a transcritos específicos (lecturas del mgPALB2) producidos por esas variantes. Por lo tanto, hemos integrado PM4 en la fuerza del código final de PS3/BS3. Los corchetes (PM4) reflejan este hecho. <sup>10</sup> PP3/BP4 están basados en las predicciones de SpliceAI. <sup>11</sup> No asignamos PP3/BP4 (según las puntuaciones de REVEL) a variantes específicas, sino a transcritos específicos (lecturas de mgPALB2) producidos por esas variantes. Por lo tanto, hemos integrado PP3/BP4 en la fuerza del código PS3/BS3 final. <sup>12</sup> Asignamos BP7 a la única variante sinónima estudiada y de la que no se esperaba que afectase al *splicing*. Sin embargo, el análisis de mgPALB2 demostró que la variante afectaba el *splicing*. En base a eso, BP7 no contribuyó a la clasificación final. Para obtener más detalles consultar la tabla suplementaria S1.