



---

Universidad de Valladolid



PROGRAMA DE DOCTORADO EN INVESTIGACIÓN BIOMÉDICA

TESIS DOCTORAL:

**TRANSCRIPCIÓN Y *SPLICING* DE *BRCA2* Y SU  
RELACIÓN CON LA SUSCEPTIBILIDAD A  
CÁNCER DE MAMA Y OVARIO HEREDITARIO**

Presentada por María Eugenia Fraile Bethencourt para  
optar al grado de  
Doctora por la Universidad de Valladolid

Dirigida por:

Dr. Eladio Andrés Velasco Sampedro





## AUTORIZACIÓN DEL DIRECTOR DE TESIS

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

Dr. Eladio Andrés Velasco Sampedro, con D.N.I. 50163233A, Investigador principal del Consejo Superior de Investigaciones Científicas (CSIC).

Centro: Instituto de Biología y Genética Molecular.

Dirección a efecto de notificaciones: Calle Sanz y Forés 3, Laboratorio B8, 47003 Valladolid.

e-mail: eavelsam@ibgm.uva.es

como director de la Tesis Doctoral titulada: "Transcripción y *splicing* de *BRCA2* y su relación con la susceptibilidad a cáncer de mama y ovario hereditario", realizada por D<sup>a</sup> María Eugenia Fraile Bethencourt, alumna del Programa de Doctorado en Investigación Biomédica, **autoriza** su presentación, considerando que el trabajo se ha realizado bajo su dirección y reúne los requisitos de originalidad y contenido para que sea presentado y pueda optar al Grado de Doctor por la Universidad de Valladolid.

Valladolid, 11 de marzo de 2019

El director de la Tesis,

Fdo.: Dr. Eladio A. Velasco Sampedro

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



En este párrafo se agradece a los organismos e instituciones implicadas en la financiación de este trabajo. En primer lugar, el Instituto de Salud Carlos III, Ministerio de Economía y Competitividad (proyecto ID PI13/01749; 2014-2016). Asimismo, a la Consejería de Educación de la Junta de Castilla y León (proyecto ID CSI090U14). Para la realización de esta tesis doctoral, la doctoranda, María Eugenia Fraile Bethencourt, ha contado con la financiación de la beca predoctoral de la Universidad de Valladolid cofinanciada por el Banco Santander (2015-2019), la beca EMBO *Short-Fellowship* (ASTF 517-2016), la beca ERASMUS+ y la ayuda para estancias breves en el desarrollo de tesis doctorales de la Universidad de Valladolid 2018.



**A MIS PADRES,**

**UNA CARRERA DE FONDO.**



## AGRADECIMIENTOS

---

Hace casi diez años, la primera semana de septiembre de 2009, mi madre y yo viajamos a Salamanca cargadas de maletas, comenzaba mi vida universitaria. Los siguientes cuatro años fueron duros a la vez que divertidos. Salamanca me dio el Grado en Biotecnología y el amor por la ciencia, pero también seis amigas de las de verdad: Guada, Rocío, Lara, Elena, Andrea y Laura; que aún hoy siguen escuchándome, aconsejándome y haciéndome reír.

En octubre de 2013 llegué a Valladolid, al IBGM, un centro que se convertiría en mi casa los siguientes seis años. En este punto me gustaría agradecer al Dr. Ángel Gato Casado por creer en mi y haber estado apoyándome durante todo este tiempo.

Con el grado en una mano y el máster en la otra, entré en el B8, laboratorio de *splicing* y cáncer. Empezaban los casi cinco años de doctorado. Siempre estaré agradecida a Eladio por haberme abierto las puertas para hacer ciencia y por descubrir mi pasión por la ingeniería genética. Entre minigenes y variantes de *splicing* pasaron los años y, gracias a la ayuda de Eladio, Cristina, Bea y Alberto, completé mi tesis. Más tarde se incorporaron al equipo Lara y Elena, a las que quiero agradecer su ayuda durante esta última etapa. Me gustaría dar las gracias también al Dr. Emanuele Buratti por haberme dado la oportunidad de trabajar con él en Italia, a la Dra. Elisa Goina por haberme enseñado tantísimo y a Ursa, Francesca, Sara y Cristiana por todos los ratos que pasamos juntas. Además, en Italia tuve la suerte de conocer a dos de mis mejores amigos, André y Cristian, gracias a los dos por todo.

Los años en Valladolid y la tesis doctoral afianzaron lazos difíciles de romper, personas increíbles que han estado ahí para lo bueno y lo malo, y con las que he compartido mucho. Gracias Iván por ser un ejemplo de trabajo duro, esfuerzo y constancia, por tu lealtad y por allanarme el camino todos los días. Gracias Alba, Nagore, Lucia, Sergio, Miguel, Rafa, Alberto, Patricia y Laura por hacerme sentir como en casa y por todos los momentos que hemos pasado juntos.

Entre las personas que he conocido durante el transcurso de estos años, quiero destacar y dar las gracias a la Dra. Elia Aguado y a la Dra. Cristina Espinosa, dos científicas que han

influido en mí y a las que me gustaría parecerme algún día. Gracias a las dos por ayudarme a traspasar fronteras.

Finalmente, gracias a mi familia. A mis abuelos, por haber sido y ser ejemplos que seguir. Gracias yaya por tus ojos llenos de orgullo cuando me miran. Gracias a mi hermana, Cristina, por estar siempre ahí, hacerme reír y tener un consejo para cada ocasión. Y gracias a mis padres, a quien va dedicada esta tesis, por su apoyo incondicional, su paciencia y por entrenarme para superar carreras de obstáculos, de velocidad y, sobre todo, de fondo.

# ÍNDICE



# ÍNDICE

---

|  |           |
|--|-----------|
| <b>ABREVIATURAS.....</b>   | <b>17</b> |
| <b>RESUMEN.....</b>  | <b>25</b> |
| <b>INTRODUCCIÓN .....</b>  | <b>29</b> |
| CÁNCER DE MAMA Y OVARIO HEREDITARIO (CMOH).....                      | 31        |
| <i>A. HISTORIA DEL CMOH.....</i>                                     | <i>31</i> |
| <i>B. CARACTERÍSTICAS DEL CMOH.....</i>                              | <i>32</i> |
| <i>C. CONSEJO GENÉTICO.....</i>                                      | <i>33</i> |
| BREAST CANCER TYPE 2 ( <i>BRCA2</i> ).....                           | 37        |
| <i>A. BRCA2 Y EL CMOH. ....</i>                                      | <i>37</i> |
| <i>B. ESTRUCTURA Y FUNCIÓN DE BRCA2. ....</i>                        | <i>39</i> |
| <i>C. VARIANTES DE SIGNIFICADO CLÍNICO DESCONOCIDO EN BRCA2.....</i> | <i>41</i> |
| TRANSCRIPCIÓN.....   | 45        |
| <i>A. INICIO DE LA TRANSCRIPCIÓN. ....</i>                           | <i>45</i> |
| <i>B. PROMOTOR DE BRCA2.....</i>                                     | <i>47</i> |
| <i>C. ALTERACIONES EN EL PROMOTOR Y ENFERMEDAD. ....</i>             | <i>47</i> |
| PROCESAMIENTO DEL MRNA O <i>SPLICING</i> .....                       | 51        |
| <i>A. EL ESPLICEOSOMA.....</i>                                       | <i>52</i> |
| <i>B. REGULACIÓN DEL SPLICING .....</i>                              | <i>53</i> |
| <i>C. SPLICING ALTERNATIVO.....</i>                                  | <i>55</i> |
| <i>D. SPLICING Y ENFERMEDAD.....</i>                                 | <i>57</i> |
| <b>JUSTIFICACIÓN .....</b>   | <b>63</b> |
| <b>OBJETIVOS.....</b>  | <b>67</b> |
| <b>ARTÍCULOS .....</b>   | <b>71</b> |
| ARTÍCULO 1 .....   | 73        |

|  |            |
|--|------------|
| ARTÍCULO 2 .....   | 111        |
| ARTÍCULO 3 .....   | 127        |
| ARTÍCULO 4 .....   | 143        |
| <b>RESULTADOS .....</b>  | <b>179</b> |
| ANÁLISIS DE LAS VARIANTES DE <i>SPLICING</i> DE LOS EXONES 14 Y 15 DE <i>BRCA2</i> .....                             | 181        |
| <i>A. MINIGEN MGBR2_14-20 Y ESTUDIO DE LAS SECUENCIAS REGULADORAS EN LOS EXONES 14 Y 15. ....</i>                    | <i>181</i> |
| <i>B. RECOPIACIÓN DE VARIANTES Y ANÁLISIS BIOINFORMÁTICO.....</i>  | <i>185</i> |
| <i>C. ENSAYO FUNCIONAL DE VARIANTES DE SPLICING EN LOS EXONES 14 Y 15 CON EL MINIGEN MGBR2_14-20.....</i>            | <i>190</i> |
| Variantes que alteran el reconocimiento del sitio aceptor.....   | 190        |
| Variantes que alteran el reconocimiento del sitio donador.....   | 192        |
| Variantes que alteran secuencias reguladoras de <i>splicing</i> (SREs) .....   | 193        |
| Variantes que crean nuevos sitios de <i>splicing</i> .....   | 194        |
| <i>D. ANÁLISIS DE LOS TRANSCRITOS .....</i>  | <i>194</i> |
| <i>E. INTERPRETACIÓN CLÍNICA DE LAS VARIANTES DE SPLICING EN LOS EXONES 14 Y 15 DE BRCA2.....</i>                    | <i>197</i> |
| <b>DISCUSIÓN .....</b>   | <b>201</b> |
| RELACIÓN ENTRE EL PROMOTOR DE <i>BRCA2</i> Y EL DESARROLLO DEL CMOH .....  | 204        |
| <i>A. REGIONES REGULADORAS DEL PROMOTOR DE BRCA2.....</i>  | <i>204</i> |
| <i>B. EFECTO DE LAS VARIANTES DE DNA EN EL PROMOTOR DE BRCA2 .....</i>   | <i>206</i> |
| PROCESAMIENTO DEL MRNA DE <i>BRCA2</i> Y SU RELACIÓN CON EL CMOH .....   | 211        |
| <i>A. MINIGENES COMO HERRAMIENTAS PARA EL ESTUDIO DE VARIANTES DE SPLICING. ....</i>                                 | <i>212</i> |
| <i>B. SECUENCIAS REGULADORAS Y FACTORES DE SPLICING IMPLICADOS EN EL RECONOCIMIENTO DE LOS EXONES DE BRCA2. ....</i> | <i>214</i> |
| <i>C. VARIANTES EN BRCA2: EFECTO EN EL SPLICING Y SIGNIFICADO CLÍNICO .....</i>                                      | <i>220</i> |
| PRESENTE Y FUTURO EN EL DIAGNÓSTICO DE VUS, ENSAYOS FUNCIONALES Y TRATAMIENTOS. ...                                  | 224        |

|                             |            |
|-----------------------------|------------|
| <b>CONCLUSIONES.....</b>    | <b>233</b> |
| <b>BIBLIOGRAFÍA .....</b>   | <b>237</b> |
| <b>ANEXOS.....</b>          | <b>253</b> |
| FIGURA SUPLEMENTARIA 1..... | 255        |
| FIGURA SUPLEMENTARIA 2..... | 257        |
| TABLA ANEXA 1.....          | 259        |
| TABLA ANEXA 2.....          | 261        |
| TABLA ANEXA 3.....          | 263        |
| TABLA ANEXA 4.....          | 267        |
| TABLA ANEXA 5.....          | 271        |



# **ABREVIATURAS**



## ABREVIATURAS

---

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

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

**aa:** aminoácido.

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

**APC:** *Adenomatous Polyposis Coli gene*.

**AS:** *Alternative Splicing*.

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

**ATP:** *Adenosine Triphosphate*.

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

**BIC:** *The Breast Cancer Information Core*.

**BRC:** Dominio de ocho repeticiones de BRCA2.

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

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

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

**BTK:** *Bruton Tyrosine Kinase gene*.

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

**CDK1:** *Cyclin-Dependent Kinase 1*.

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

**CHD7:** *Chromodomain-Helicase-DNA-binding protein 7 gene*.

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

**CIMBA:** *Consortium of Investigators of Modifiers of BRCA1/2*.

**CK:** Citokeratinas.

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

**COL5A1:** *Collagen Type V Alpha 1 gene*.

**Da / kDa:** Dalton / kiloDalton

**DBD:** *DNA Binding Domain.*

**dbSNP:** Base de datos de SNP.

**DDR:** *DNA Damage Response.*

**DMD:** *Dystrophin gene.*

**DNA:** *Deoxyribonucleic Acid.*

**DPE:** *Downstream Promoter Element.*

**DSB:** *Double-Strand Break.*

**DSS1:** *Deleted in Split-hand/Split-foot Syndrome protein.*

**EKLF:** *Erythroid Kruppel-Like Factor.*

**EMSA:** *Electrophoretic Mobility Shift Assay.*

**ENCODE:** *Encyclopedia of DNA Elements.*

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

**ER:** *Estrogen Receptor.*

**erbB-2 (HER2/neu):** Receptor tirosina-proteína quinasa erbB-2.

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

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

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

**ExSpeU1:** *Exonic Specific U1 snRNA.*

**FA:** *Fanconi Anemia.*

**FANCD1:** *Fanconi Anemia group D1 protein.*

**FANCD2:** *Fanconi Anemia group D2 protein.*

**FDA:** *Food and Drug Administration.*

**GRB7:** *Growth factor Receptor Bound protein 7 gene.*

**GTFs:** *General Transcription Factors.*

**HD:** *Helical Domain.*

**HGMD:** *The Human Gene Mutation Database.*

**HHB:** *Human Hemoglobin subunit Beta gene.*

**hnRNP:** Ribonucleoproteínas heterogéneas nucleares.

**RH:** Recombinación Homóloga.

**IARC:** *International Agency for Research on Cancer.*

**ICLs:** *Interstrand Crosslinks.*

**INR:** *Iniciator Element.*

**InSIGHT:** *International Society for Gastrointestinal Hereditary Tumors.*

**ISE:** *Intronic Splicing Enhancer.*

**ISS:** *Intronic Splicing Silencer.*

**Kb:** Kilobase.

**LDLR:** *Low density lipoprotein receptor.*

**LKB1-STK11:** *Serine/Threonine Kinase 11 gene.*

**mESC:** *mice Embryonic Stem Cells.*

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

**MMR:** *Mismatch Repair.*

**MRN:** complejo proteico formado por MRE11, RAD50 y nibrina (NBN).

**mRNA:** RNA mensajero.

**MRE11:** *Double-strand break repair protein MRE11.*

**MSH2:** *MutS Homolog 2 gene.*

**MSH6:** *MutS Homolog 6 gene.*

**NBN:** *Nibrin gene.*

**NCI:** *National Cancer Institute.*

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

**NFYA:** *Nuclear Transcription Factor Y Subunit Alpha.*

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

**NLS:** *Nuclear Localization Sequence.*

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

**NMDA:** N-Metil-D-Aspartato.

**nt:** nucleótido.

**OB:** *Oligonucleotide Binding domain.*

**ORF:** *Open Reading Frame.*

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

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

**PAX5:** *Paired Box 5.*

**pb:** pares de bases.

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

**PgR:** *Progesterone Receptor.*

**PIC:** *Preinitiation Complex.*

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

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

**PTM:** *Pre-Trans-splicing Molecule.*

**qPCR:** *quantitative Polymerase Chain Reaction.*

**RAD50:** *RAD50 Recombinase.*

**RAD51:** *RAD51 Recombinase.*

**RB1:** *Retinoblastoma 1 gene.*

**RMN:** Resonancia Magnética Nuclear.

**RNA:** Ácido Ribonucleico.

**RNApol:** RNA polimerasa.

**RPB:** Subunidad de la RNA polimerasa.

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

**RRMH:** *RNA Recognition Motif Homologous.*

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

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

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

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

**SMaRT:** *Spliceosome-Mediated RNA Trans-Splicing.*

**SMN:** *Survival Motor Neuron gene.*

**SNP:** *Single Nucleotide Polymorphism.*

**snRNP:** Ribonucleoproteína nuclear pequeña.

**SREs:** *Splicing Regulatory Elements.*

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

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

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

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

**TBP:** *TATA binding protein.*

**TD:** *Tower Domain.*

**TERT:** *Telomerase reverse transcriptase gene.*

**FT:** Factor de Transcripción.

**CMTN:** Cáncer de Mama Triple Negativo.

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

**TSS:** *Transcription Start Site.*

**UMD:** *The Universal Mutation Database.*

**UTR:** *Untranslated Region.*

**VHL:** *Von Hippel-Lindau tumor suppressor gene.*

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

**wt:** *wild type.*



# **RESUMEN**



## RESUMEN

---

**Introducción:** El Cáncer de Mama y Ovario Hereditario (CMOH) es una enfermedad con herencia autosómica dominante caracterizada por la aparición temprana del tumor. Actualmente, aunque se han asociado 26 genes a la aparición de la enfermedad, sólo dos de ellos están considerados como genes principales de cáncer de mama: *BRCA1* y *BRCA2*. El rastreo de los genes *BRCA* en pacientes de alto riesgo ha dado lugar a la aparición de miles de variantes, muchas de ellas clasificadas como variantes de significado clínico desconocido (VUS). Éstas suelen ser variantes *missense*, sinónimas, deleciones o inserciones *in-frame* y variantes en regiones no codificantes como los intrones o las UTRs (*Untranslated Region*). En este contexto, los ensayos funcionales juegan un papel esencial para conocer el impacto de este tipo de variantes y su asociación con la enfermedad. El objetivo de esta tesis doctoral se centra en el estudio de la transcripción y el *splicing* de *BRCA2*. Se pretende estudiar los mecanismos reguladores de la expresión génica y evaluar las variantes que puedan alterar dichos procesos y estar asociadas a la susceptibilidad a CMOH.

**Materiales y métodos:** Las variantes, recogidas de las bases de las principales bases de datos (BIC, UMD, ClinVar y Ensembl) y encontradas en pacientes del sistema de prevención de cáncer de Castilla y León (España), fueron analizadas *in silico*. Se seleccionaron aquellas variantes que potencialmente alteraban los mecanismos de transcripción o de *splicing*. El estudio de la transcripción se realizó a través de un vector reportero con el gen de luciferasa bajo el promotor de *BRCA2*. Los resultados fueron analizados mediante *Dual-Luciferase Assays* y RT-PCR semi-cuantitativa. Para estudiar el *splicing*, se construyeron dos minigenes (MGBR2\_2-9 y MGBR2\_14-20) basados en el vector pSAD (Patente P201231427, CSIC). Las variantes fueron introducidas mediante mutagénesis dirigida y ensayadas en células MCF-7. El estudio de regiones reguladoras se realizó a través de microdeleciones. Los factores implicados en el *splicing* se estudiaron mediante ensayos de interacción RNA-proteína (*pull-down*) y ensayos de pérdida de función (siRNA). Los transcritos generados fueron analizados mediante gel de agarosa, secuenciación SANGER y electroforesis capilar.

**Resultados:** Las secuencias reguladoras de la transcripción se estudiaron a través de 13 microdeleciones a lo largo del promotor de *BRCA2*. Los resultados mostraron que seis disminuían significativamente la transcripción y tres la aumentaban. Una vez mapeado el promotor, se ensayaron 15 variantes que podrían alterar elementos reguladores, tres de ellas encontradas en pacientes del sistema de prevención de Cáncer de Castilla y León (rs3092989\_A, rs206118\_G y rs563971900\_T). Diez variantes alteraron significativamente la expresión génica, de las cuales, tres provocaron la subexpresión del gen (rs551887850\_G, rs570548398\_T y rs55880202\_T). Cabe destacar que ocho de las 10 variantes reguladoras mapeaban en el *cluster* 121 de hipersensibilidad a DNAsa, una región crítica para el inicio de la transcripción. Por otro lado, el estudio del *splicing* reveló que los exones 3 a 8, 15, 17 y 18 de *BRCA2* requieren de elementos reguladores de *splicing* (SREs) para su correcto reconocimiento. Además, los resultados indican que los factores SRSF3 y SRSF2 juegan un papel importante en el *splicing* de los exones 3 y 18, respectivamente. En total, se analizaron mediante minigenes 200 variantes repartidas a lo largo de los exones 2-9 y 14-18 de *BRCA2*. La mayoría de las variantes produjeron alteraciones en el *splicing* (104), entre ellas 49 eran exónicas (*missense*, sinónimas, *nonsense* y *frameshift*). Aproximadamente el 70% de las variantes espliceogénicas (73/104) provocaron graves alteraciones (>60% de transcritos aberrantes). En total, 74 variantes fueron clasificadas como patogénicas o probablemente patogénicas según los criterios propuestos por ACMG (*American College of Medical Genetic and Genomics*), 30 de las cuales se encontraban previamente catalogadas como VUS.

**Conclusiones:** La transcripción y el *splicing* son mecanismos fundamentales en la expresión génica que pueden verse alterados debido a cambios en la secuencia de DNA. El promotor de *BRCA2* es sensible a cambios de nucleótido y su rastreo debería incluirse en los estudios genéticos de pacientes de CMOH. Por otro lado, cualquier variación en la secuencia podría alterar el *splicing*. En este contexto, los estudios con minigenes proporcionan datos fiables para evaluar el impacto de las variantes cuando no se dispone de RNA de pacientes.

# **INTRODUCCIÓN**



# INTRODUCCIÓN

---

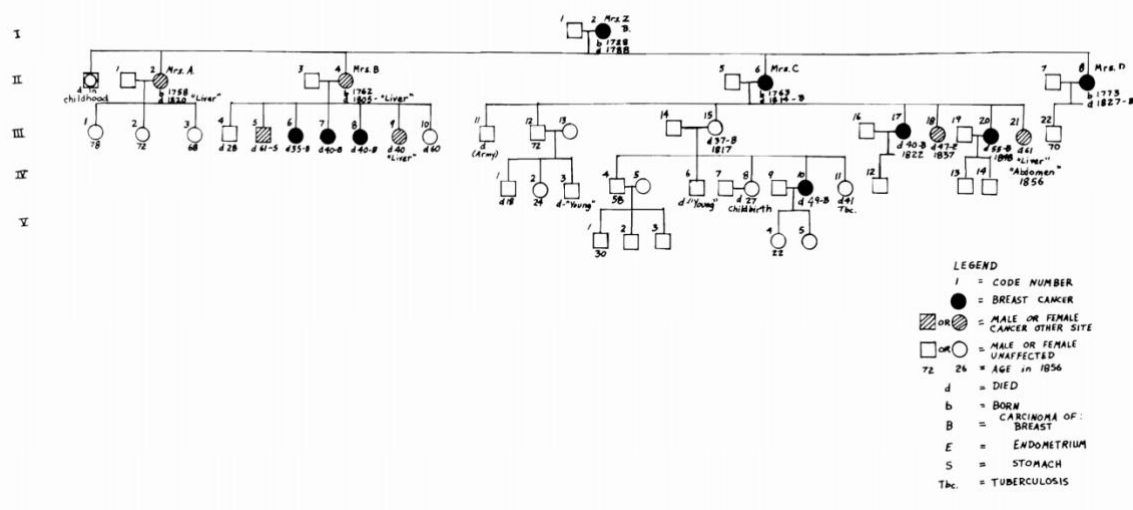
## CÁNCER DE MAMA Y OVARIO HEREDITARIO (CMOH)

El cáncer es el resultado de la proliferación descontrolada de un grupo de células que adquieren la capacidad de migrar e invadir nuevos tejidos. De esta manera se forma el tumor, una masa heterogénea de células, de las cuales, algunas, serán capaces de viajar y colonizar otros órganos dando lugar a lo que se conoce como metástasis. El cáncer se produce debido a la acumulación de mutaciones, generalmente originadas de manera esporádica. Sin embargo, ocasionalmente, estas mutaciones se dan en células germinales, lo que provoca formas hereditarias de cáncer como el síndrome de Cáncer de Mama y Ovario Hereditario (CMOH). Se estima que entre un 5-10% de los casos de cáncer de mama son de carácter hereditario (Allender and Arias, 2008).

### A. HISTORIA DEL CMOH

El primero en reportar un caso familiar de cáncer de mama fue el cirujano francés Paul Broca en 1866 (Broca, 1866). Broca estudió cinco generaciones de la familia de su esposa, la cual sufrió de cáncer de mama a una edad temprana. En su estudio, observó que nueve de las 24 mujeres de la familia habían sufrido la enfermedad y sugirió una susceptibilidad a este tipo de cáncer en la familia (Figura 1). Más tarde, en 1971, Lynch y Krush describieron por primera vez la asociación entre el cáncer de mama y el cáncer de ovario (Lynch and Krush, 1971). En 1990, Hall y colaboradores asociaron el cáncer de mama temprano con una alteración en el cromosoma 17q21 (Hall et al., 1990), hecho que corroboraron Narod y colaboradores un año después cuando encontraron el mismo locus 17q12-q23 relacionado con el cáncer de mama y ovario hereditario (Narod et al., 1991). Este gen, conocido como *BRCA1* (Breast Cancer type 1, OMIM #113705), se clonó por primera vez en 1994 (Miki et al., 1994). Sólo un año más tarde, se descubrió el gen *BRCA2* (Breast Cancer type 2, OMIM #600185) en el cromosoma 13q12-q13 (Wooster et al., 1995). El descubrimiento de los genes *BRCA1* y *BRCA2* facilitó la comprensión y el diagnóstico del CMOH. Ambos genes están implicados en el mantenimiento de la estabilidad del genoma mediante mecanismos de reparación del DNA (Boulton, 2006). Hoy sabemos que el CMOH es un síndrome

heterogéneo con el que se han asociado hasta 26 genes de susceptibilidad (Nielsen et al., 2016). Sin embargo, los genes *BRCA1* y *BRCA2* siguen siendo los únicos considerados como principales genes de susceptibilidad a CMOH. Las mutaciones germinales en los *BRCA* explican en torno a un 20% de los casos de CMOH, cifra que aumenta ligeramente al añadir los nuevos genes de moderada y baja penetrancia (Lynch et al., 2013). Como explicación al resto de los casos, se ha planteado la existencia un tercer gen (*BRCA3*), pero aún no se ha identificado. Otra hipótesis sugerida es la del modelo poligénico en el que variantes en genes de media y baja susceptibilidad actuarían de forma sinérgica para conferir alto riesgo (Alonso Sánchez M. et al., 2006).



**Figura 1.** Representación del árbol genealógico de la familia de Mr. Z. recogido en el libro *Traité de tumeurs* de Broca (1866) (Broca, 1866; Krush, 1979).

## B. CARACTERÍSTICAS DEL CMOH

El CMOH se define como un síndrome en el cual el riesgo de sufrir cáncer de mama (especialmente antes de los 50 años) y/o cáncer de ovario es más alto de lo normal (*National Cancer Institute*, NCI). Las familias con CMOH presentan un patrón de herencia dominante, caracterizado por la temprana edad de aparición del cáncer de mama y la presencia de casos de cáncer de ovario, cáncer de mama bilateral y cáncer de mama masculino. Cabe destacar

que el CMOH no es el único síndrome hereditario asociado a cáncer de mama hereditario. Existen otros síndromes que predisponen a la aparición de cáncer de mama con herencia autosómica dominante, entre los cuales destacan: el síndrome de Li-Fraumeni, asociado a mutaciones en el gen *TP53* (Li et al., 1988; Malkin et al., 1990); el síndrome de Cowden, asociado al gen *PTEN* (Liaw et al., 1997; Lloyd, 1963; Nelen et al., 1997); y el síndrome de Peutz-Jeghers, asociado al gen *LKB1-STK11* (Bruwer et al., 1954; Hemminki et al., 1998; Jenne et al., 1998).

El diagnóstico clínico en España para el CMOH se basa en los criterios definidos por la Sociedad Española de Oncología Médica (SEOM):

- Cáncer de mama temprano (<40 años).
- Co-ocurrencia de cáncer de mama y ovario.
- Dos o más casos de cáncer de mama, uno de los cuales bilateral o antes de los 50 años.
- Cáncer de mama bilateral o de aparición temprana (<50 años) y un caso de cáncer de ovario en un familiar de primer o segundo grado.
- Tres casos de cáncer de mama y/u ovario (al menos un caso de cáncer de ovario) en familiares de primer o segundo grado.
- Dos casos de cáncer de ovario en familiares de primer o segundo grado.
- Un caso de cáncer de mama en varón y un familiar de primer o segundo grado con cáncer de mama u ovario.

## **C. CONSEJO GENÉTICO**

La asociación de los genes *BRCA* al CMOH facilitó el diagnóstico molecular de la enfermedad. Sin embargo, sólo aproximadamente el 20% de los casos de CMOH se deben a alteraciones en estos dos genes. En los últimos años, con la secuenciación de nueva generación (NGS), se ha comprobado la existencia de una alta heterogeneidad genética, de hecho, se han propuesto, además de *BRCA1* y *BRCA2*, otros 24 genes asociados al desarrollo de la enfermedad. A nivel molecular, la mayoría de estos genes están relacionados con los *BRCA* en el mantenimiento de la integridad genómica, entre ellos destacan: *ATM*, *PALB2*,

*BRIP1*, *CDH1* y *CHEK2* (Nielsen et al., 2016). En un estudio reciente se rastrearon 26 genes en 2.134 individuos con alto riesgo de cáncer de mama familiar y sin variantes patogénicas en los *BRCA*. Se encontró que un 8,2% de los pacientes portaban variantes patogénicas, de las cuales, aproximadamente la mitad se encontraban en *ATM*, *PALB2*, *CHEK2* y *PT53* (Slavin et al., 2017).

La aparición de nuevos genes implicados en el CMOH, así como el desarrollo de la NGS, ha abierto la puerta a nuevas formas de rastreo genético como la secuenciación de genoma completo, los exomas o los paneles multigénicos (Rizzo and Buck, 2012). Estos últimos son los más utilizados para el diagnóstico de cáncer hereditario ya que, al seleccionar sólo los genes asociados a la enfermedad, es posible abaratar costes, acortar tiempos y obtener resultados más precisos. En el año 2017, se publicó un artículo en el que 35.409 mujeres (de las cuales un 93,2% (32.993) cumplían los criterios de CMOH) se sometieron a un rastreo genético usando un panel de 25 genes relacionados con el cáncer hereditario (*Myriad Genetics Laboratory*). Los principales genes asociados con el cáncer de mama hereditario eran: *BRCA1*, *BRCA2*, *ATM*, *BARD1*, *BRIP1*, *CDH1*, *CHEK2*, *NBN*, *PALB2*, *STK11* y *TP53*. Los resultados indicaron que 9,6% de las mujeres que cumplían los criterios de CMOH portaban una variante patogénica, mientras que, en las no cumplían los criterios, se encontraron variantes patogénicas en un 5,9%. Es importante destacar que, entre todas las mujeres analizadas, el 36,7% portaba al menos una variante de significado clínico desconocido (VUS). De las variantes patogénicas encontradas, aproximadamente la mitad estaban en los genes *BRCA* (*BRCA1*: 24,0% y *BRCA2*: 24,4%); otros genes en los que se encontraron variantes patogénicas en un porcentaje relevante fueron: *CHEK2* (11,7%), *ATM* (9,7%) y *PALB2* (9,3%). Este estudio pone de manifiesto la utilidad de los paneles en la práctica clínica, pues más de la mitad de las variantes patogénicas fueron encontradas en genes distintos de los *BRCA* (Buys et al., 2017).

Las familias portadoras de mutaciones patogénicas asociadas a CMOH pueden beneficiarse de programa de prevención basado en: i) el cribado: autoexamen mensual, examen clínico semestral, mamografía, ecografías y RMN anuales, y ecografías transvaginales semestrales; ii) la quimioprevención en portadoras enfermas con tamoxifeno

para prevenir la aparición de un tumor contralateral (Fisher et al., 2005); iii) el estilo de vida (peso, actividad física y alimentación); iv) la cirugía preventiva: mastectomía bilateral profiláctica y salpingo-ooforectomía bilateral profiláctica (Kurian et al., 2010). Tanto la mastectomía bilateral como la salpingo-ooforectomía han demostrado una disminución del riesgo de padecer cáncer de mama u ovario, respectivamente, hasta en un 90-95% (Mau and Untch, 2017).

Los tumores de mama asociados a CMOH parecen presentar un pronóstico diferente a los tumores esporádicos. En el año 2000, se propuso una clasificación de los tumores basada en sus perfiles de expresión génica según la cual podemos clasificar los tumores de mama en: luminal, ERBB2, mama normal y *basal-like* (Perou et al., 2000). Los tumores de tipo luminal se caracterizan por una alta expresión de los genes típicamente expresados por las células luminales de la mama, como el gen ER (receptor de estrógenos) y las citokeratinas (CK) 8/18. El subtipo ERBB2, también conocido como HER2 (*Human Epidermal Growth Factor*), se caracteriza por la sobreexpresión de este protooncogen, así como de genes asociados a él, como el factor de crecimiento GRB7 (*Growth factor Receptor Bound protein 7*). Los tumores definidos como mama normal expresan los genes típicos del epitelio basal y de las células adiposas. Por último, los tumores *basal-like* están caracterizados por una elevada expresión de las CK 5/6 y 17 y por la ausencia tanto de ER como de genes asociados a él. Los tumores *basal-like* a menudo son llamados triple negativo (CMTN). Cabe destacar que el término *basal-like* viene dado por un patrón de expresión génica definido mediante microarrays, mientras que el término CMTN se refiere a un resultado negativo de la tinción inmunohistoquímica para ER, PgR (receptor de progesterona) y erbB-2 (HER2/neu). Aunque a menudo son intercambiables, estos términos no son equivalentes, ya que cada uno hace referencia a características diferentes del tumor. En los casos esporádicos, aproximadamente el 70% son ER positivos y el 15% sobreexpresan ERBB2, sólo una minoría son CMTN (Larsen et al., 2014). Sin embargo, en los tumores asociados a mutaciones patogénicas en *BRCA1*, el fenotipo CMTN es el más común. En el estudio de Buys y colaboradores (2017), comentado anteriormente, encontraron que el 7,2% de las mujeres con CMTN portaban al menos una variante patogénica en *BRCA1*, mientras que en las mujeres con otros subtipos sólo se

encontraban en el 1,5% (Buys et al., 2017). En portadoras de mutaciones en *BRCA2* el fenotipo es más similar a los tumores esporádicos, aunque el porcentaje de CMTN aumenta levemente, el más común es el luminal con sobreexpresión de ER y PgR (Bosch et al., 2010).

Los CMTN son los tumores de peor pronóstico. Por una parte, la ausencia de ER y PgR impide el tratamiento hormonal, y por otra, la ausencia de HER2/neu hace inefectiva la terapia con anticuerpos monoclonales. Actualmente, la terapia con inhibidores de la poliadenosin difosfato-ribosa polimerasa (PARP) está dando resultados esperanzadores en CMTN con mutaciones en *BRCA*. La PARP se activa como respuesta al daño en el DNA para iniciar mecanismos de reparación. El tratamiento anti-PARP se basa en el incremento de la inestabilidad genómica para provocar la muerte celular. La terapia consiste en inhibir los mecanismos compensatorios de respuesta a daño en el DNA (DDR, *DNA Damage Response*), que se activan en células deficientes de recombinación homóloga (RH), como son las que presentan variantes patogénicas en los genes *BRCA*. Así, la inhibición de mecanismos DDR en células naturalmente deficientes de RH desencadena una inestabilidad genómica letal para la célula (Curtin, 2014; Pilié et al., 2019). La FDA (*Food and Drug Administration*) aprobó por primera vez en 2014 el uso de olaparib para el tratamiento de cáncer de ovario refractario en estado avanzado con mutaciones en *BRCA1/2*. Desde entonces, se han aprobado tres inhibidores-PARP: rucaparib (2016), niraparib (2017) y talazoparib (2018) (<https://www.fda.gov/>). Este último, comercializado con el nombre de Talzenna (Pfizer, Inc.), fue aprobado en octubre de 2018 para el tratamiento de cáncer de mama avanzado o metastásico con mutaciones germinales patogénicas en *BRCA1/2* y HER2-negativo.

## BREAST CANCER TYPE 2 (*BRCA2*)

### A. *BRCA2* Y EL CMOH.

*BRCA2* es un gen de alta susceptibilidad a CMHO, pues variantes patogénicas confieren un elevado riesgo de padecer la enfermedad. Las mutaciones en *BRCA2* se asocian con un riesgo de cáncer de mama en mujeres de hasta el 84% a la edad de 70 años (Ford et al., 1998), mientras que en hombres el riesgo aumenta hasta el 6,8% a los 70 años (Tai et al., 2007). El riesgo de cáncer de ovario aumenta hasta el 22% a los 70 años (Chen et al., 2006). Sin embargo, la penetrancia puede variar dependiendo del tipo de mutación, del contexto genómico y de factores exógenos como el estilo de vida. Además de cáncer de mama y ovario, diferentes estudios demuestran que mutaciones en *BRCA2* confieren un aumento significativo en el riesgo de padecer otros tipos de cáncer como cáncer de páncreas, cáncer de próstata y melanoma (Petrucelli et al., 2016).

En 1997 Kinzler y Vogelstein dividieron los genes asociados a cáncer en dos tipos: los genes "*gatekeeper*" y los genes "*caretaker*". El primer término hace referencia a los genes que regulan la muerte celular, controlando así el crecimiento del tumor. Los genes *gatekeeper* son específicos de tejido, por lo tanto, su inactivación da lugar a tumores en órganos o tejidos concretos. Otra característica de estos genes es la necesidad de inactivar ambas copias de gen para desencadenar el proceso tumoral. Algunos ejemplos de estos genes son: *RB1* (retinoblastoma), *VHL* (cáncer de riñón) y *APC* (cáncer de colon polipósico). Por otro lado, los genes *caretaker* promueven el desarrollo tumoral de una manera indirecta, pues están relacionados con el mantenimiento de la integridad del genoma. Así, la inactivación de estos genes lleva a un incremento en el número de mutaciones de otros genes. Kinzler y Vogelstein propusieron los genes *BRCA* como ejemplo de genes *caretaker* (Kinzler and Vogelstein, 1997).

Existe otra clasificación que divide los genes asociados al desarrollo del cáncer en protooncogenes y genes supresores de tumores. Los protooncogenes son ubicuos y necesarios para el mantenimiento celular, están asociados con proliferación, desarrollo y supervivencia. Las alteraciones en estos genes provocan la desregulación de la ruta en la

que están implicados, modificando así procesos fundamentales del metabolismo celular. Las alteraciones que suelen sufrir son de ganancia de función, incrementan la tasa de crecimiento celular promoviendo la aparición del tumor. Una característica fundamental de los protooncogenes es que la alteración de uno de los alelos es suficiente para desencadenar el proceso, hecho que les confiere una herencia autosómica dominante. Un protooncogen con una mutación de ganancia de función se denomina oncogen. Un ejemplo de protooncogen es *ERBB2* (o *HER2*) que codifica un receptor del factor de crecimiento epidérmico sobreexpresado en el cáncer de mama tipo ERBB2. Por otro lado, los genes supresores de tumores se diferencian de los anteriores principalmente porque el tumor aparece debido a la inactivación del gen (pérdida de función). Normalmente, estos genes juegan un papel fundamental en rutas de reparación de DNA o en apoptosis y están implicados en el control de la proliferación celular, frenando este proceso cuando las condiciones no son adecuadas. En el caso de los genes supresores de tumores es necesaria la inactivación de ambos alelos para desencadenar el desarrollo tumoral. Por esta razón, son los genes supresores de tumores los que están asociados a cánceres hereditarios. Se hereda un alelo mutado y durante la vida, se produce, de manera esporádica, una mutación en el otro alelo. Este mecanismo, conocido como “two-hits” (Knudson and Jr., 1971), explica lo que ocurre con *BRCA2* y el CMOH. *BRCA2* es un gen supresor de tumores implicado en la reparación de roturas de doble hebra del DNA mediante RH. La inactivación de uno de los alelos de *BRCA2* no presenta síntomas, sin embargo, cuando tiene lugar el segundo “hit”, o inactivación del otro alelo se pierde la funcionalidad de la proteína.

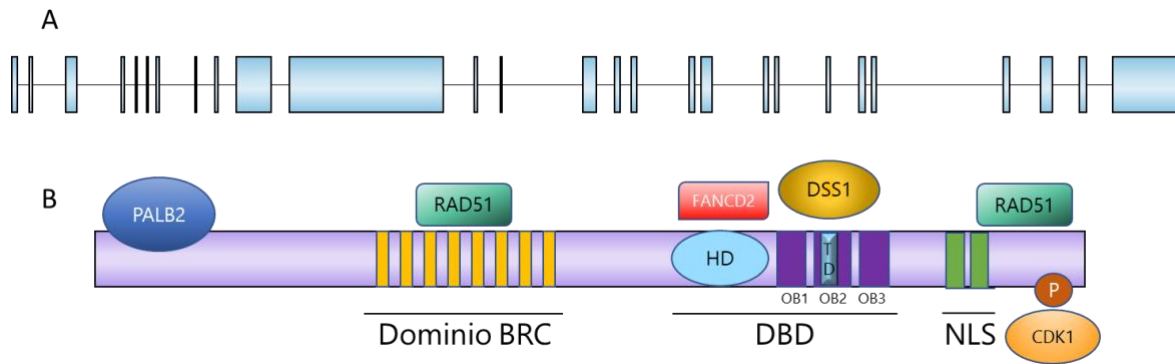
Existen estudios que apuntan a que las células con los dos alelos de *BRCA2* inactivos mueren por apoptosis, siendo necesaria la pérdida de las proteínas de control del ciclo celular p53 y WAF1 (también conocida como p21) para que la célula progrese y se desarrolle el tumor (Narod and Foulkes, 2004). Por otro lado, aunque las mutaciones patogénicas en heterocigosis en *BRCA2* son asintomáticas, el CMOH se considera una enfermedad con herencia autosómica dominante, lo cual es debido a la alta probabilidad de padecer la enfermedad cuando se hereda uno de los alelos mutados. La inactivación completa del gen debida a mutaciones germinales, provoca anemia de Fanconi (FA), una enfermedad rara con

herencia autosómica recesiva, caracterizada por la hipersensibilidad a *Interstrand Crosslinks* (ICLs), que cursa con desarrollo de tumores y fallo progresivo de la médula ósea (Howlett et al., 2002).

## **B. ESTRUCTURA Y FUNCIÓN DE *BRCA2*.**

El gen Breast Cancer Type 2, conocido como *BRCA2*, es un gen de 85.993 pb, situado en el brazo largo del cromosoma 13 (13q.13.1). *BRCA2* está compuesto por 27 exones que dan lugar a un mRNA de 11.386 nt (NM\_000059.3) (Figura 2 A). *BRCA2* codifica una proteína de 3.418 aa, llamada "Breast cancer type 2 susceptibility protein" o BRCA2 (384 kDa), implicada en la reparación de roturas de doble hebra (DSB) mediante RH (Yang et al., 2002).

El gen *BRCA2* fue identificado en 1995 (Wooster et al., 1995). Un año más tarde, se publicó la secuencia codificante, la distribución exónica del gen y el patrón de expresión (Tavtigian et al., 1996). Se observó que, de los 27 exones de *BRCA2*, se traducen 26, pues la pauta abierta de lectura (ORF) y el codón de iniciación de metionina se encuentran en el segundo exón, a 228 nt. Además, detectaron los niveles de máxima expresión en mama y en timo. Se han encontrado homólogos de la proteína BRCA2 humana en organismos eucariotas desde animales hasta hongos, pero no en levaduras. Tampoco se han encontrado homólogos en bacteria o Archea (Lo et al., 2003). BRCA2 consta de varios dominios funcionales (Figura 2 B): a) El dominio de transactivación en el extremo N-terminal, donde se sitúa la región de unión a PALB2, (*partner and localizer of BRCA2*), que une físicamente BRCA1 y BRCA2 (Xia et al., 2006a); b) El dominio BRC, codificado principalmente por el exón 11, que consta de ocho repeticiones a las que se une la recombinasa RAD51 (Pellegrini et al., 2002); c) El dominio de unión a DNA (DBD), el más conservado, está compuesto por un dominio helicoidal (HD) asociado a FANCD2 (Wang et al., 2004), tres dominios de unión a oligonucleótidos de cadena sencilla (OB) y un dominio protuberante llamado Torre (TD) (Guidugli et al., 2014). Al DBD se une la proteína DSS1 (*Deleted in Split-hand/Split-foot syndrome*), que juega un papel fundamental en la estabilidad de BRCA2 (Li et al., 2006; Yang et al., 2002). Además, en el extremo C-terminal encontramos la secuencia de localización nuclear (NLS) y otra zona de unión a RAD51 susceptible de fosforilación por CDK1 (*Cyclin-Dependent Kinase 1*) (Esashi et al., 2005).



**Figura 2.** Representación del gen y de la proteína BRCA2 humana, respectivamente. A) Representación del gen *BRCA2*. Los 27 exones que conforman *BRCA2* aparecen como rectángulos azules, separados por las regiones intrónicas. B) Representación de la proteína BRCA2, los principales dominios y las proteínas de unión (adaptación de Roy, Chun, & Powell, 2012).

La función principal de BRCA2 es el reclutamiento de RAD51 durante la RH. Cuando se produce una DSB, la célula debe detectarla inmediatamente y resolverla. La hebra de DNA expuesta, de cadena sencilla, es reconocida por el complejo MRN, formado por MRE11, RAD50 y nibrina (NBN). Este complejo activa ATM, una quinasa encargada de promover la señalización para la reparación del daño, que termina en el reclutamiento de BRCA2 hacia la DSB (Nielsen et al., 2016). BRCA2, a su vez, recluta a RAD51, que se une a una de las cadenas terminales del DNA y forma filamentos nucleo-proteicos encargados de buscar la secuencia homóloga en la cromátida hermana. Se produce entonces un intercambio de cadenas de DNA para iniciar la RH (Shahid et al., 2014). La RH es un mecanismo de reparación de DSB asociadas a replicación sin errores que se produce durante las fases S y G2 del ciclo celular y contribuye al mantenimiento de la integridad del genoma. Además, BRCA2 participa en la reparación de ICLs a través de la ruta de FA. Las ICLs son daños en el DNA que se producen normalmente debido a agentes quimioterapéuticos. Estos daños impiden la separación de la doble hebra, inhibiendo así la replicación y la transcripción. BRCA2, conocida también como FANCD1, es uno de los de 15 factores implicados en esta ruta (Deans and West, 2011). BRCA2 y PALB2 juegan también un papel importante durante la división celular. Ambas proteínas son necesarias para un correcto funcionamiento del *checkpoint* G2 antes de entrar en mitosis (Menzel et al., 2011). Recientemente, se ha demostrado el papel de BRCA2 en la

estabilidad de las horquillas de replicación reversa. Cuando el DNA sufre un daño, la replicación se ve forzada a retroceder en lo que se conoce como horquillas de replicación reversa, un proceso que permite la reparación del daño sin frenar la replicación. BRCA2 y RAD51 juegan un papel fundamental en la estabilidad de estas horquillas protegiendo las hebras de cadena sencilla de la degradación por nucleasas (Quinet et al., 2017).

### C. VARIANTES DE SIGNIFICADO CLÍNICO DESCONOCIDO EN *BRCA2*.

Llamamos variantes a los cambios en la secuencia de DNA de un determinado gen. En *BRCA2*, según ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/?term=brca2%5Bgene%5D>), se han encontrado unas 10.000 variantes diferentes distribuidas a lo largo de todo el gen (marzo 2019). Existen otras bases de datos, como *The Breast Cancer Information Core* (BIC), en la que aparecen registradas 2.000 variantes diferentes de *BRCA2* (última modificación 8 de mayo de 2017), o *The Universal Mutation Database* (UMD) (Beroud et al., 2016), en la que aparecen recogidas 3.454 variantes de *BRCA2* (última actualización 16 de junio de 2017). Aproximadamente, el 70% son variaciones de nucleótido simple (*Single Nucleotide Polymorphisms*, SNPs) o pequeñas deleciones e inserciones. Aunque, también se han encontrado grandes deleciones y duplicaciones y otros reordenamientos genómicos.

Teniendo en cuenta el efecto que provocan en la proteína, las variantes exónicas se clasifican en:

- *Frameshift*: deleción o inserción que altera la ORF. Normalmente generan en un codón de *stop* prematuro (PTC) truncando la proteína.
- *In-frame*: deleción o inserción que no altera la ORF (múltiplo de 3nt).
- *Missense*: cambio de nucleótido que genera un cambio de aminoácido.
- *Nonsense*: cambio de nucleótido que generan un PTC.
- Sinónimas: cambio de un nucleótido que no produce cambio de aminoácido.

Existen otros tipos de variantes como las intrónicas, que se sitúan en las partes no codificantes del gen, o las de *splicing*, que alteran el procesamiento del mRNA. También se han encontrado variantes en las regiones no traducidas al principio y al final del gen (UTRs). Las variantes que afectan al extremo 5' UTR se conocen como variantes de promotor, y

pueden alterar la expresión del gen. Según ClinVar, la mayoría de las variantes patogénicas en *BRCA2* son variantes *frameshift* (~65%), seguidas de variaciones *nonsense* (~20%).

Atendiendo a su significado clínico, entre las 10.000 variantes descritas en ClinVar, aparecen clasificadas como variantes benignas o probablemente benignas aproximadamente un 30%; como variantes patogénicas o probablemente patogénicas otro 30%; y como variantes de significado clínico desconocido (VUS) un 40%. Por otro lado, según la UMD, de las 3.454 variantes, el 63% son VUS. Las VUS representan un desafío en la práctica clínica, pues no se dispone de la información necesaria para predecir su papel en el desarrollo de la enfermedad. Éstas suelen ser variantes *missense* o sinónimas, variantes en regiones no codificantes o no traducidas e inserciones o deleciones *in-frame*. El principal problema reside en que el desconocimiento del riesgo asociado a la variante es equivalente a un diagnóstico negativo, de manera que el cálculo de riesgo y la toma de decisiones se basan en la historia familiar (Radice et al., 2011). Así, esta aproximación podría estar privando a portadores de mutaciones en los *BRCA* de las medidas profilácticas y/o los tratamientos específicos cuando no existe una historia familiar.

Actualmente, no existe un sistema universal de clasificación de variantes de *BRCA1/2* (Eccles et al., 2015). Entre las distintas aproximaciones que se utilizan para clasificar las variantes encontramos análisis poblacionales, estudios de cosegregación, historial familiar y/o personal, análisis del tumor, análisis de conservación evolutiva, ensayos funcionales y modelos multifactoriales, que incluyen todos los parámetros anteriores. En el año 2008, la Agencia Internacional para el estudio del Cáncer (IARC) presentó un sistema para la clasificación de variantes en genes de alta susceptibilidad a cáncer de mama y cáncer de colon, basado en el modelo de probabilidad multifactorial (Goldgar et al., 2004). Este método multifactorial integra datos independientes para obtener una probabilidad de patogenicidad. Según la IARC, atendiendo a esta probabilidad, las variantes se pueden clasificar en cinco clases (Plon et al., 2008):

- Clase 1: la probabilidad de patogenicidad es  $<0.001$ . El manejo clínico es equivalente a la ausencia de mutación.

- Clase 2: la probabilidad de patogenicidad es 0.001-0.049. El manejo clínico es equivalente a una variante de clase 1.
- Clase 3: la probabilidad de patogenicidad es 0.05-0.949. El manejo clínico depende de la historia familiar y otros factores de riesgo.
- Clase 4: la probabilidad de patogenicidad es 0.95-0.99. El manejo clínico es la puesta en marcha de los sistemas de prevención.
- Clase 5: la probabilidad de patogenicidad es >0.99. El manejo clínico es la puesta en marcha de los sistemas de prevención.

En el año 2009, se creó el consorcio ENIGMA (*Evidence-based Network for the Interpretation of Germline Mutant Alleles*) con el objetivo evaluar el significado clínico de las variantes en los genes de susceptibilidad a cáncer de mama hereditario (Spurdle et al., 2012). El criterio de clasificación propuesto por ENIGMA pretende sentar las bases para unificar el sistema, proporcionando reglas concisas que se pueden encontrar en su página web (<https://enigmaconsortium.org/>). Otra asociación importante es ACMG (*American College of Medical Genetics and Genomics*), cuyo criterio de clasificación de variantes está también basado en cinco clases (Richards et al., 2015): patogénicas, probablemente patogénicas, benignas, probablemente benignas y VUS. Este tipo de consorcios internacionales juegan un papel fundamental en la mejora de la estimación del riesgo asociado a las variantes y permiten establecer criterios universales para la clasificación de variantes y la práctica clínica. Otras asociaciones como CIMBA (*Consortium of Investigators of Modifiers of BRCA1/2*), establecida en el año 2005 y formada por más de 70 grupos de investigación en todo el mundo (<http://cimba.ccge.medschl.cam.ac.uk/>), estudian las variaciones a gran escala. En su último estudio, presentan resultados del análisis de 18.435 familias portadoras de variantes en *BRCA1* y 11.351 portadoras en *BRCA2* (Rebbeck et al., 2018). La actuación coordinada de los estudios a gran escala con el establecimiento de criterios de selección rigurosos proporciona datos fiables para afrontar el desafío clínico que supone la clasificación de las VUS.



## TRANSCRIPCIÓN

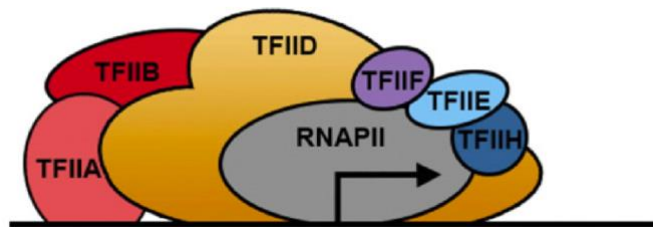
La transcripción es el primer paso de la expresión génica. Se trata del proceso por el cual la información contenida en el DNA es copiada en una molécula de RNA. Las enzimas encargadas de la transformación se conocen como RNA polimerasas (RNAPol). Entre los muchos tipos de RNA producidos, el RNA mensajero (mRNA) es el que dará lugar a la proteína en un proceso conocido como traducción. Sin embargo, antes de la traducción, en organismos eucariotas es necesario un procesamiento previo, conocido como *splicing*. El *splicing*, o procesamiento del mRNA, consiste en escindir las secuencias no codificantes (intrones) y unir secuencialmente las regiones codificantes (exones). Tanto el inicio de la transcripción por la RNAPol como el *splicing* son procesos cruciales de la expresión génica.

Desde principio de los 60s, cuando simultáneamente cuatro laboratorios descubrieron la RNAPol (Huang et al., 1960; Hurwitz et al., 1961; Stevens, 1960; Weiss and Gladstone, 1959), no han parado los esfuerzos por entender la regulación de la expresión génica. En 1977, Philip A. Sharp y Richard J. Roberts descubrieron un mecanismo que complicaría aún más el entendimiento de dicho proceso: el *splicing* (Berget and Sharp; Chow et al., 1977); hallazgo que les valió a ambos el Nobel de Fisiología o Medicina en 1993. En 2006, Roger D. Kornberg recibió el premio Nobel de Química por sus estudios sobre la estructura tridimensional del RNAPol II, encargada de producir el mRNA en organismos eucariotas mediante el reconocimiento de secuencias promotoras. Todo esto demuestra la importancia que sigue teniendo el estudio de la expresión génica aún a día de hoy.

### A. INICIO DE LA TRANSCRIPCIÓN.

Se conoce como promotor a la secuencia de DNA que determina donde comienza la transcripción. Dado que la transcripción se realiza en dirección 5' a 3', la región donde se encuentra el promotor es la región 5' del gen. Aquí, encontramos la secuencia 5' UTR, la cual hace referencia a la región 5' del gen que se transcribe, pero no se traduce. El promotor juega un papel fundamental en la regulación del inicio de la transcripción, pues contiene secuencias específicas de unión a RNAPol y a factores de transcripción (FTs). Además, es donde está definido el sitio de inicio de la transcripción (TSS), es decir, el primer nucleótido

que será convertido en RNA. En torno al TSS se encuentra el núcleo del promotor, la secuencia mínima necesaria para iniciar la transcripción (Butler and Kadonaga, 2002). El núcleo del promotor es una secuencia de aproximadamente 90 nucleótidos, 40 antes del TSS (-40) y 50 después del TSS (+50) (de Vooght et al., 2009), en la que se encuentran los sitios de unión para los factores de transcripción generales (GTFs, *General Transcription Factors*) y la RNAPol II. La RNAPol II es una enzima de 12 subunidades (RPB1-12), capaz de producir moléculas de mRNA con la ayuda de los GTFs: TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, TFIIH. Para el inicio de la transcripción es necesaria una estrecha colaboración entre la RNAPol II y los GTFs. Según el modelo clásico, en primer lugar, se forma el complejo pre-iniciación (PIC) (Figura 3), que comienza cuando TFIIIF recluta a RNAPol II y la dirige hasta el núcleo de promotor. Allí se encontrarán ya unidos TFIIB y TBP (TATA *binding-protein*). TBP forma parte de TFIID, está conservada en eucariotas y es la encargada de reconocer el núcleo del promotor. A este complejo, se unen posteriormente TFIIIF y TFIIH para terminar de formar PIC. El siguiente paso es la apertura de la doble hebra de DNA en un proceso dependiente de ATP, llevado a cabo principalmente por TFIIH. TFIIA funciona como factor auxiliar para la formación de PIC, no se trata de un factor esencial, pero ayuda a estabilizar el complejo TBP-DNA (Sainsbury et al., 2015). Si bien es cierto que TBP es uno de los factores principales en el inicio de la transcripción, se sabe que sólo el 24% de los promotores humanos contienen secuencias TATA (Yang et al., 2007). El mecanismo por el cual comienza la transcripción en promotores sin secuencia TATA es complejo, se ha visto que TBP puede ser sustituida por otros factores específicos.



**Figura 3.** Complejo de pre-iniciación de la transcripción (PIC). Aparecen representada la RNAPol II y los GTFs (Danino et al., 2015).

## **B. PROMOTOR DE *BRCA2*.**

En *BRCA2*, el TSS se encuentra 981 pb antes del primer codón traducido. En 1999, L. Davis *et al.* publicaron un trabajo en el cual encontraron, a través de microdeleciones, la región del promotor de *BRCA2* donde se concentraba la mayor parte de la actividad transcripcional. Así, caracterizaron el núcleo del promotor de *BRCA2* desde la posición -66 hasta la +129, tomando como referencia el TSS. Observaron que no había secuencias TATA, definiendo el promotor de *BRCA2* como promotor carente de caja TATA (*TATA-less promoter*). Además, encontraron factores de transcripción que se unían a esta región como E-box, ETs/E2F y SP1 (Davis et al., 1999). Como previamente se ha comentado, el mecanismo de inicio de la transcripción para promotores TATA está bien caracterizado. Sin embargo, la mayoría de los promotores humanos, como el de *BRCA2*, no contienen elementos TATA, aunque muchos sí requieren TBP y TFIID para un correcto funcionamiento. Los promotores carentes de caja TATA han sido mucho menos estudiados debido a su variedad, pero se ha visto que existen factores de transcripción *downstream* del núcleo del promotor (DPE, *Downstream Promoter Element*) y otros elementos como el iniciador (INR) que promueven la formación del complejo PIC y el inicio de la transcripción (Seizl et al., 2011; Yang et al., 2007). Además, existen *enhancers*, elementos reguladores que se encuentran a distancia del promotor y funcionan favoreciendo la transcripción a través de la formación de lazos. Un ejemplo de *enhancer* es el exón 3 de *BRCA2*, que funciona favoreciendo el inicio de la transcripción (Milner et al., 1997).

## **C. ALTERACIONES EN EL PROMOTOR Y ENFERMEDAD.**

El inicio de la transcripción es el primer paso para la expresión génica. El reclutamiento de la RNAPol a través de los GTFs debe realizarse en un tiempo y espacio adecuados para que la transcripción se produzca correctamente. Además, existen factores específicos y secuencias en *cis* que se unen a la región promotora y actúan coordinadamente para llevar a cabo la transcripción. La alteración de este proceso puede desembocar en un aumento o una disminución en la expresión de un determinado gen, y consecuentemente en la aparición de una enfermedad. En este contexto, el estudio de las variantes de DNA a nivel de promotor cobra vital importancia, pues pueden alterar la unión de los factores

mencionados anteriormente y provocar la desregulación de la transcripción (de Vooght et al., 2009). De hecho, se estima que aproximadamente el 1% de las variantes causantes de enfermedad ocurren en la región promotora (Cooper, 2002). En la mayoría de los casos, las mutaciones en el promotor alteran secuencias de unión a factores de transcripción como TBP, GATA-1 o SP1. Algunos ejemplos de enfermedades causadas por alteraciones a nivel de promotor son: la hipercolesterolemia familiar, en la que han identificado al menos 36 variantes en la región 5'UTR del gen *LDLR* (*Low Density Lipoprotein Receptor*), muchas de las cuales alteran la unión de factores de transcripción clave como SP1 (Khamis et al., 2015); la  $\beta$ -talasemia causada por alteraciones en la caja TATA, la caja CACCC o la secuencia de unión a EKLF (*Erythroid Kruppel-Like Factor*) en el promotor del gen *HHB* (*Human Hemoglobin subunit Beta*) (Hardison et al., 2002); o la hemofilia B causada por alteraciones en el promotor del gen *F9* que codifica el factor de coagulación IX (Crossley and Brownlee, 1990).

En distintos tipos de cáncer encontramos también varios ejemplos de alternaciones a nivel de promotor. Como se ha explicado anteriormente, las variantes de promotor pueden afectar a la unión de FTs, cuyas consecuencias pueden traducirse en un aumento o disminución de la expresión del gen. Así, la generación de sitios de unión a FTs activadores o la disrupción de sitios de unión a FTs represores, provocará la sobreexpresión del gen. Mientras que, la generación de sitios de unión a FTs represores o la disrupción de sitios de unión a FTs activadores, provocará la subexpresión del gen. Esto cobra vital importancia en el contexto del cáncer, pues la sobreexpresión de un protooncogen o la subexpresión de un gen supresor de tumores podría estar asociada al desarrollo de la enfermedad.

El ejemplo de variantes de promotor que provocan la sobreexpresión de un protooncogen lo encontramos en el gen *TERT*, cuyas alteraciones en el promotor han sido asociadas con el desarrollo de melanoma hereditario y esporádico. *TERT* codifica una subunidad de la telomerasa que se encuentra reprimida en células somáticas. Recientemente se ha encontrado una variante en el promotor de *TERT* que genera sitios de unión a FTs activadores, provocando un aumento de la expresión del gen (Horn et al., 2013).

En cuanto a variantes de promotor que disminuyen la expresión de genes supresores de tumores, el ejemplo más claro lo encontramos en las variantes de promotor de los genes MMR (*missmatch repair*) y su asociación con el síndrome de Lynch. El síndrome de Lynch es una enfermedad con herencia autosómica dominante caracterizada por un aumento de la susceptibilidad a cáncer de colon y endometrio. Los pacientes portan una mutación de pérdida de función en los genes MMR en heterocigocis y, como en *BRCA2*, es necesario una segunda mutación somática para la inactivación de la proteína (*two-hits*). Los dos genes de alta susceptibilidad son *MLH1* y *MSH2*, los cuales abarcan hasta el 88% de las variantes encontradas. Según la base de datos de InSIGHT (*International Society for Gastrointestinal Hereditary Tumors*), hay diez variantes registradas en el promotor en *MLH1* y seis variantes en el de *MSH2*, la mayoría clasificadas como clase 1, 2 o 3 (benignas, probablemente benignas o VUS). Por ejemplo, la variante c.-42C>T en *MLH1* disminuye la expresión del gen a un 37% en células HeLa (Green et al., 2003), pero se considera VUS debido a la falta de otros estudios. La única variante de promotor clasificada como clase 4 en la base de datos de InSIGHT es la variante c.-78\_-77del de *MSH2*. Esta variante provoca la pérdida de un sitio de unión a FT activador, reduciendo su expresión a un 16,8% con respecto al *wild type* (Liu et al., 2016).

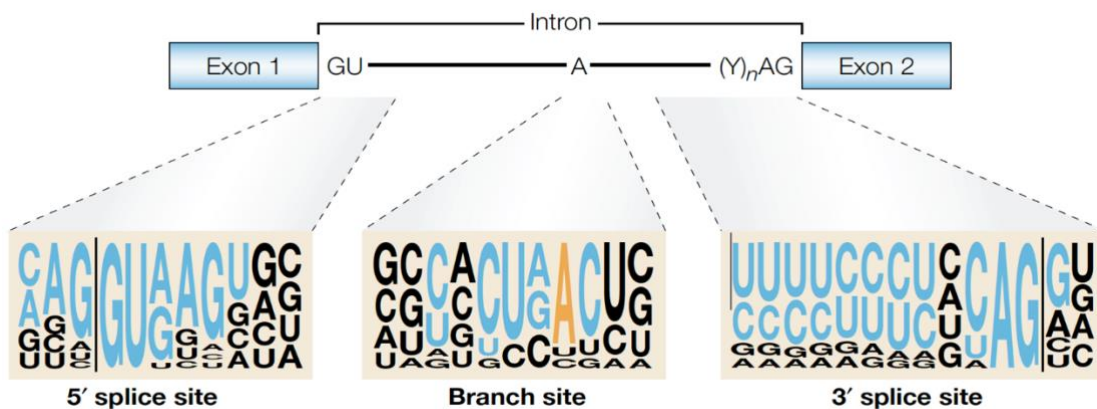
El análisis del promotor es, por tanto, fundamental para la mejora en el diagnóstico de enfermedades genéticas. Sin embargo, existen ciertas dificultades debido principalmente a que las regiones promotoras no están bien definidas y son difíciles de analizar, pues poseen un alto contenido en GC. Además, la presencia de elementos reguladores con secuencias poco conservadas y alejados como los *enhancers*, hace aún más difícil su estudio. Por otro lado, la interpretación clínica de las variantes en el promotor es complicada debido a que, en la mayoría de los casos, producen efectos parciales en la transcripción del gen en cuestión.



## PROCESAMIENTO DEL MRNA O *SPLICING*.

La mayoría de los genes eucariotas están formados por secuencias codificantes (exones), separadas por secuencias no codificantes (intrones). Durante la transcripción de un gen, la secuencia de DNA es convertida en RNA por la RNAPol II. Este RNA se conoce como mRNA precursor (pre-mRNA), pues los intrones deben ser escindidos y los exones unidos secuencialmente para obtener un mRNA maduro que pueda ser traducido (Gilbert, 1978). Este mecanismo es conocido como procesamiento del mRNA o *splicing*.

El *splicing* es un proceso altamente regulado en el que exones e intrones deben ser reconocidos de manera precisa. Existen secuencias intrónicas conservadas que determinan los límites exón-intrón, estas secuencias básicas de *splicing* son: el sitio donador o sitio 5' de *splicing* (5' ss), el punto de ramificación, el tracto de polipirimidina y el sitio aceptor o sitio 3' de *splicing* (3' ss) (Figura 4) (Cartegni et al., 2002). El intrón representado en la Figura 4 corresponde a un intrón U2-dependiente. Los intrones procesados por el espliceosoma menor, U12-dependientes, difieren principalmente en la conservación de los sitios de *splicing*: AT-5'ss y AC-3'ss (Padgett, 2012).



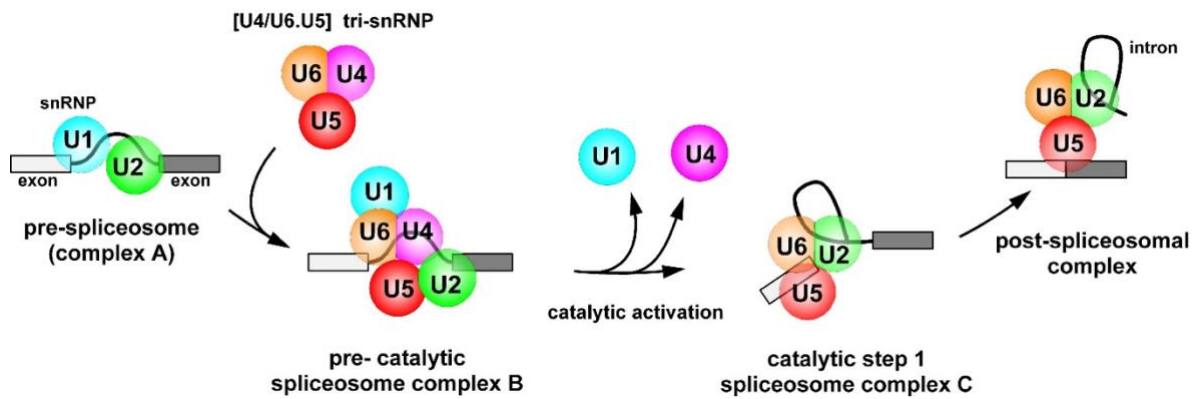
**Figura 4.** Secuencias clásicas de *splicing*: el sitio donador "GU" o 5' *splice site* (5'ss), el punto de ramificación "A" (*branch point*) y el sitio aceptor "AG" o 3' *splice site* (3'ss) con el tracto de polipirimidinas. Debajo, aparece representada la conservación evolutiva de los nucleótidos que conforman estos motivos (Cartegni et al., 2002).

## A. EL ESPLICEOSOMA

El *splicing* consiste en dos reacciones de transesterificación que se realizan en el núcleo gracias a la ayuda de un macrocomplejo molecular dinámico llamado espliceosoma. El espliceosoma fue visto por primera vez por en 1985 a través del microscopio electrónico (Osheim et al., 1985).

En las células eucariotas existen dos tipos de espliceosoma: el espliceosoma mayor (U2-dependiente); y el espliceosoma menor (U12-dependiente). El espliceosoma mayor es el más frecuente y está formado por cinco ribonucleoproteínas pequeñas (snRNPs) ricas en uridina: U1, U2, U4, U5 y U6; y más de 200 proteínas asociadas (Hegele et al., 2012). El espliceosoma menor es el encargado de escindir un determinado tipo de intrones muy poco frecuentes, conocidos como intrones U12, que representan aproximadamente el 0,1% de los intrones en organismos superiores (Levine and Durbin, 2001).

Para llevar a cabo el *splicing*, el espliceosoma debe ensamblarse y formar el complejo catalítico (C) (Figura 5). El ensamblaje comienza con el reclutamiento de las snRNPs U1 y U2 a los sitios de *splicing* 5'ss y 3'ss respectivamente, para formar el complejo A. El siguiente paso es la unión de la tri-snRNP U4/U6.U5 para formar el complejo B, que se activará cuando U1 y U4 sean liberados. El complejo B necesita un aporte energético en forma de ATP para ser catalíticamente activo, cuando esto ocurre pasa a llamarse C y comienza la catálisis de las reacciones de transesterificación. Durante la primera reacción de transesterificación se produce el corte de la secuencia de DNA debido a un ataque nucleofílico de la adenina del punto de ramificación al 5' ss. En la segunda reacción tiene lugar la unión de los extremos exónicos cuando el grupo OH libre del 5' ss se une al 3' ss. Finalmente se produce la disociación del complejo y la liberación de las snRNPs (Will and Lührmann, 2011).

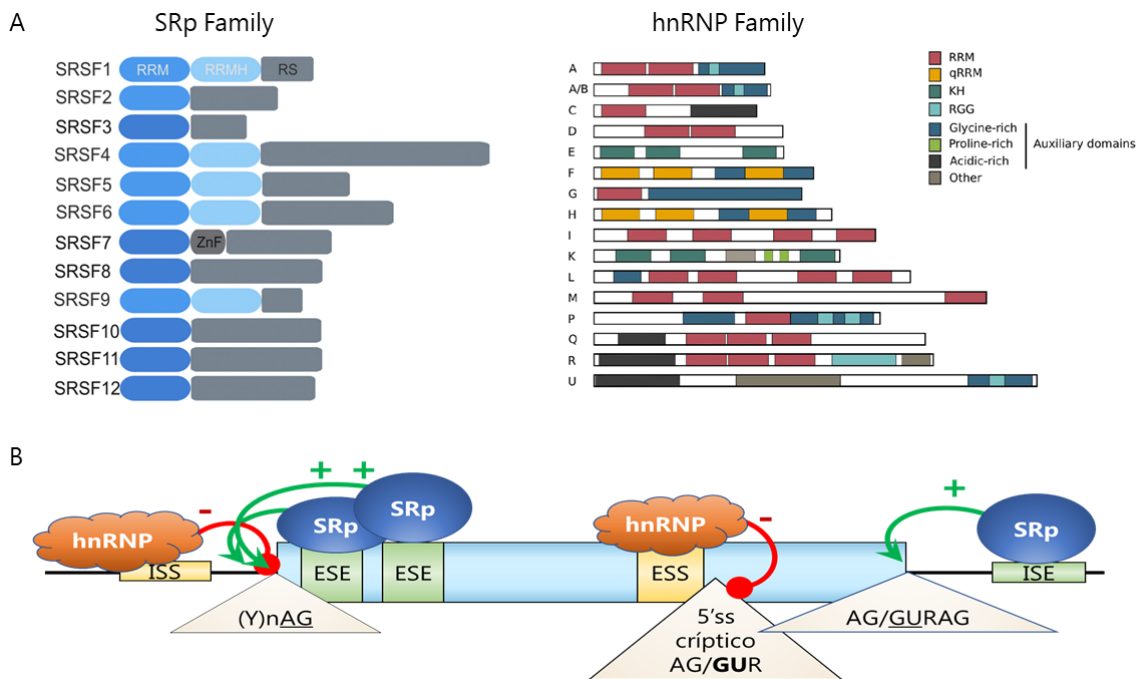


**Figura 5.** Ensamblaje del spliceosoma y reacción de splicing (Ilustración de Reinhard Lührmann – Max Panck Institute for Biophysical Chemistry).

## B. REGULACIÓN DEL *SPLICING*

Además de los sitios clásicos de *splicing*, se pueden encontrar elementos auxiliares en *cis* que actúan como reguladores del *splicing* (*Splicing Regulatory Elements*, SREs). Los SREs pueden actuar como *enhancers*, favoreciendo la inclusión de los exones, o como silenciadores, impidiendo el reconocimiento de los exones. Además, pueden encontrarse en regiones exónicas o intrónicas. Así diferenciamos entre *Exonic Splicing Enhancers* (ESEs) e *Intronic Splicing Enhancer* (ISEs) para elementos exónicos o intrónicos, respectivamente, que favorecen el reconocimiento del exón; y *Exonic Splicing Silencers* (ESSs) e *Intronic Splicing Silencers* (ISSs) para elementos exónicos o intrónicos, respectivamente, que impiden el reconocimiento del exón. A los SREs se unen factores de *splicing* activadores o represores que intervienen en la elección del sitio de *splicing* por el espliceosoma (De Conti et al., 2012). Los factores de *splicing* activadores más comunes son las proteínas SR (*Serin/Arginine Rich proteins*, SRp). La familia SRp se compone de 12 proteínas, actualmente conocidas como SRSF (1-12), caracterizadas por la presencia de uno o dos dominios de reconocimiento de RNA (RRM) o un dominio homólogo a este (RRMH) en el extremo N-terminal y un dominio RS rico en serina/arginina en el C-terminal (Figura 6 A). Estas proteínas participan en el ensamble del espliceosoma y son necesarias para reconocer uniones exón-intrón (Änkö, 2014). Por otro lado, los factores de *splicing* represores más comunes son las ribonucleoproteínas heterogéneas nucleares o hnRNPs. Las hnRNPs son una familia de proteínas capaces de unir RNA, implicadas en todo el metabolismo de ácidos nucleicos. La

función de las hnRNPs varía principalmente dependiendo de su localización (Geuens et al., 2016). Las hnRNPs constan de una estructura modular con al menos un dominio RRM asociado a dominios auxiliares (Figura 6 A). Dentro de la familia, las proteínas hnRNPA, hnRNPA/B y hnRNPH son las mejor caracterizadas como silenciadores de *splicing* (Cartegni et al., 2002). Durante el *splicing*, las hnRNPs son capaces de unirse a ESSs o ISSs y actuar como antagonistas del reconocimiento de sitios de *splicing*. En concreto, hnRNPA1 y hnRNPA2/B1 han sido caracterizadas como represores del reconocimiento exónico en algunos genes tan importantes como el gen de la  $\beta$ -globina, el gen del receptor de NMDA (N-metil-D-aspartato) o *BRCA1* (Han et al., 2010). Las secuencias clásicas de *splicing* en combinación con *enhancers* y silenciadores, es lo que finalmente determina el reconocimiento de los exones (Figura 6 B).

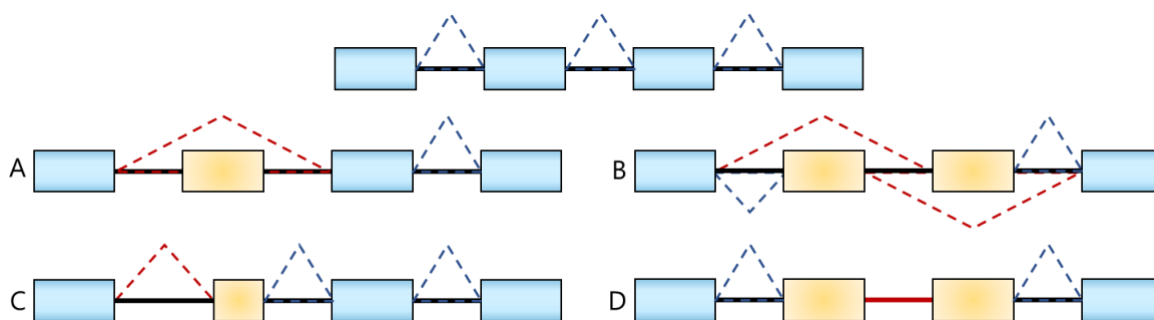


**Figura 6.** Factores de *splicing* y regulación. A) Representación de la familia SRp (Änkö, 2014) y de la familia hnRNPs (Geuens et al., 2016). B) Representación de la regulación de *splicing* por los SREs y las proteínas que se unen a ellos. Las flechas, verdes y rojas, indican la potenciación o inhibición de los sitios de *splicing*, respectivamente.

## C. *SPLICING* ALTERNATIVO

Debido a la degeneración de las secuencias de *splicing*, a la gran variedad de proteínas que componen el espliceosoma y a los diferentes SREs, el procesamiento del mRNA no sucede siempre de igual manera. Así, se obtienen transcritos de diferente composición mediante lo que se conoce como *splicing* alternativo (AS). La decisión de incluir o excluir parte del gen, se conoce como evento. Los eventos que tienen lugar durante el AS generan distintas moléculas de mRNA a partir de un mismo gen, estas nuevas moléculas se conocen como isoformas. En humanos, más del 95% de los genes poseen AS, lo cual es una de las razones que explican la gran variabilidad proteómica. Además, el AS está implicado en la diferenciación y el desarrollo celular, pues los diferentes eventos se encuentran influenciados por factores de *splicing* específicos de tejido, de momento de desarrollo o de respuesta a estímulo (Lee and Rio, 2015).

Los patrones de AS son muy variados (Figura 7). La pérdida de un exón o exón *skipping* (Figura 7 A) es el evento más frecuente en mamíferos y tiene lugar cuando el espliceosoma no reconoce un sitio de *splicing* y se salta uno de los exones. Otro patrón típico de AS es el mutuamente excluyente (*mutually-exclusive*) (Figura 7 B), que se da cuando se incluye sólo uno u otro exón, pero nunca los dos en el mismo transcrito. Este mecanismo sucede a menudo en el primer o en el último exón, de manera que encontramos isoformas con TSS alternativos o con sitios alternativos de poliadenilación, respectivamente. Otro mecanismo común del AS es el uso de sitios crípticos o sitios alternativos de *splicing*, que ocurre cuando el espliceosoma reconoce preferencialmente estos sitios en lugar de los sitios canónicos de *splicing* (Figura 7 C). Por último, la retención intrónica, que, aunque es extraña en condiciones fisiológica en organismos superiores, se puede observar en el AS de intrones cortos de plantas y de eucariotas unicelulares (Wagner and Berglund, 2014). Cabe destacar que la retención intrónica ha sido observada como mecanismo de inactivación de genes supresores de tumores en células tumorales (Jung et al., 2015) (Figura 7 D).



**Figura 7.** Patrones de *splicing* alternativo. A) Pérdida del exón o *exon skipping*. B) *Splicing* mutuamente excluyente. C) Uso de sitio de *splicing* alternativo. D) Retención intrónica.

El AS, a diferencia del tamaño del genoma, está directamente relacionado con la complejidad del organismo. Así, cuanto mayor sea el grado de desarrollo del organismo, mayor será la tasa de AS. Además, las especies del mismo reino comparten el patrón principal de AS, por ejemplo mientras en el reino animal predomina el exón *skipping*, en hongos y protistas es típica la retención intrónica (McGuire et al., 2008).

*BRCA2*, como la mayoría de los genes humanos, posee AS. En 2016, siete grupos pertenecientes al consorcio ENIGMA publicaron el primer catálogo de isoformas alternativas de *BRCA2* (Fackenthal et al., 2016). Encontraron 24 isoformas naturales de *BRCA2* en diferentes tipos celulares, siendo cuatro de ellas predominantes:

- $\Delta 3$ : pérdida del exón 3.
- $\Delta 6q,7$ : uso de un donador alternativo del exón 6 y pérdida del exón 7.
- $\Delta 12$ : pérdida del exón 12.
- $\Delta 17,18$ : pérdida de los exones 17 y 18.

Es importante destacar que no todas las isoformas de *splicing* son funcionales, de hecho, se estima que un 2% de los transcritos de un gen son producto de errores en el *splicing* (Pickrell et al., 2010). Así, el AS constituye un mecanismo que, aparte de aportar diversidad proteica, es capaz de regular la expresión de los niveles de proteína, pues los transcritos portadores de PTCs serán degradados a través de la ruta *Nonsense Mediated Decay* (NMD) (Kervestin and Jacobson, 2012).

## D. *SPLICING* Y ENFERMEDAD

El *splicing* es un mecanismo fundamental de la expresión génica, de manera que errores en el proceso pueden desembocar en el desarrollo de enfermedades. El correcto reconocimiento de las secuencias de *splicing* es un proceso muy regulado, pues se trata de secuencias cortas y degeneradas que compiten con otras similares. En ocasiones, los sitios crípticos de *splicing*, o sitios no canónicos, encajan mejor con las secuencias de reconocimiento del espliceosoma, sin embargo, el reconocimiento de uno u otros depende de la presencia de secuencias *cis* a las que se unen factores activadores o represores. Así, se hace necesaria la intervención cientos de ribonucleoproteínas y proteínas de unión a RNA que varían en función del espacio y el tiempo. Los cambios en la secuencia de DNA debido a mutaciones pueden provocar alteraciones en los elementos en *cis*, en las proteínas reguladoras del *splicing* o en las moléculas que componen espliceosoma, alterando el correcto procesamiento del mRNA y estando implicados en el desarrollo de enfermedades. Por ejemplo, en síndromes mielodisplásicos se han encontrado diferentes alteraciones de *splicing* debido a mutaciones los factores SF3B1, SRSF2 y U2AF1 (Pellagatti et al., 2018). De hecho, más del 50% de los pacientes con este tipo de síndromes portan mutaciones heterocigotas somáticas en estos genes. El efecto de estas mutaciones en células de la médula ósea, se traduce en un *splicing* aberrante que altera la hematopoyesis y lleva al desarrollo de la enfermedad (Chabot and Shkreta, 2016). Actualmente, se conocen unas 370 enfermedades causadas como consecuencia de un *splicing* aberrante, lo cual confirma la desregulación del *splicing* como mecanismo común asociado al desarrollo de enfermedad (Wang et al., 2012).

En general, se conoce como variantes de *splicing* aquellas que producen alteraciones en *cis*, es decir, variaciones en el DNA que pueden dar lugar a modificaciones de secuencias de *splicing*, tanto básicas como SREs, alterando el reconocimiento de los límites exón/intrón y generando como consecuencia transcritos aberrantes. Se estima que, aproximadamente un 15% de las mutaciones patogénicas reportadas están alterando una secuencia básica de *splicing*, y que un 20% de las mutaciones *missense* se sitúan en SREs (Manning and Cooper, 2017). Además, se debe tener en cuenta que estas estimaciones no incluyen las mutaciones

intrónicas que puedan alterar ISEs o ISSs, ni aquellas que generan sitios *de novo*. Por ejemplo, en 2011, se reportó un caso de agammaglobulinemia ligada al cromosoma X producida por una mutación intrónica en el gen *BTK* que generaba un nuevo 5' ss. El transcrito resultante incorporaba los primeros 107 nt del intrón 3, alterando la pauta de lectura de la proteína. El cambio de nucleótido (T>G) generaba un sitio donador GC, presente solo en el 1% de los intrones humanos. Este tipo de donadores son reconocidos gracias a la ayuda de SREs y factores auxiliares de *splicing* (Kralovicova et al., 2011).

Las consecuencias de las mutaciones de *splicing* son inserciones o deleciones que alteran el mRNA. Cuando esos cambios son *in-frame*, es decir, no alteran la ORF, se producirá una proteína más corta. Sin embargo, lo más común es que se produzcan inserciones o deleciones *frameshift*, es decir, que alteran la ORF. Normalmente, este tipo de modificaciones desembocan en un PTC y el transcrito es digerido a través del NMD. Dependiendo de la cantidad de transcrito aberrante, la degradación del mRNA a través del NMD producirá una reducción parcial o total de los niveles de proteína en la célula.

Existen diferentes tipos de mutaciones de *splicing* en función de donde se produzcan las variaciones en la secuencia de DNA. Las variaciones en los sitios canónicos de *splicing*, afectan a regiones que definen la frontera exón/intrón. El sitio donador o 5' ss (GAG/GUAAGU) y el sitio aceptor o 3' ss (NYAG/G) son secuencias conservadas que deben ser reconocidas por el espliceosoma, de manera que un cambio puede alterar su identificación. Las variantes más frecuentes afectan a las posiciones +1 y +2 en el donador y a las posiciones -1 y -2 en el aceptor (Anna and Monika, 2018). Como consecuencia de este tipo de mutaciones, el evento más común es la pérdida del exón, es decir, el espliceosoma no reconoce el exón y obtenemos un transcrito carente de él. Sin embargo, en ocasiones existen aceptores o donadores crípticos que son reconocidos cuando los canónicos desaparecen. En este caso se producen deleciones o inserciones parciales que dependen del lugar donde se sitúe el sitio críptico de *splicing*. El efecto de las variantes en los sitios canónicos de *splicing* puede ser muy variable en función de la conservación de las secuencias, la cantidad de SREs o la estructura secundaria del mRNA. Todos estos factores hacen posible un reconocimiento parcial, que tendría como consecuencia la producción de

diferentes isoformas. Por ejemplo, en el gen de la distrofina (*DMD*) la mutación IVS25+1G>A provoca la pérdida del exón 25, una delección *in frame* asociada a la distrofia muscular de Becker. En el mismo gen, la variante IVS45+1G>A, asociada a una forma más grave de la enfermedad conocida como distrofia muscular de Duchenne, produce varias isoformas *frameshift* debido al reconocimiento de sitios crípticos de *splicing* (Habara et al., 2008). Otro ejemplo, estudiado por nuestro grupo en colaboración con el Instituto de Investigación Biocruces en Baracaldo (España), es la variante c.5565+1G>T, que afecta al donador del exón 28 del gen *CHD7*, asociado al síndrome de CHARGE (Villate et al., 2018). El estudio con minigenes mostró que la variante provocaba el uso de un donador críptico, resultando en un transcrito con una inserción de 63 nt en el intrón 28. Curiosamente, a pesar de ser una inserción *in-frame*, se generaba un PTC después de ocho nuevos aminoácidos, lo que provocaba una proteína truncada.

Por otro lado, el sitio aceptor puede verse afectado por variantes que alteran el punto de ramificación o el tracto de polipirimidinas. Estas mutaciones, aunque no afectan directamente al sitio 3' de *splicing*, evitan su reconocimiento y, normalmente, desembocan en la pérdida del exón en cuestión. Debido a la degeneración de estas secuencias, las mutaciones de punto de ramificación y de tracto de polipirimidinas son raras y muy difíciles de detectar. En 1998 se detectó por primera vez una mutación de punto de ramificación, la variante IVS32-25T>G, a sólo 2 nt de la adenina altamente conservada que funciona como punto de ramificación en el exón 33 del gen *COL5A1*. Esta mutación provoca el *splicing* aberrante del aceptor del exón 33 y está asociada al síndrome de Ehlers-Danlos (Burrows et al., 1998).

Otro tipo de mutaciones de *splicing* son las variantes intrónicas profundas, localizadas entre exones y, normalmente, en intrones largos. Estas mutaciones suelen generar nuevos sitios de *splicing* o fortalecer sitios crípticos que serán reconocidos por el espliceosoma, provocando la inclusión de un fragmento del intrón (o pseudoexón) en el mRNA (Vaz-Drago et al., 2017). Un ejemplo es la mutación mencionada anteriormente en el gen *BTK*. Las mutaciones intrónicas no son demasiado frecuentes y su patogenicidad depende, en la mayoría de los casos, del mantenimiento o no de la ORF. Aun así, es necesario tenerlas en

cuenta, ya que normalmente se centra la atención en las secuencias exónicas e intrónicas cercanas a la frontera exón/intrón. Los proyectos NGS de secuenciación completa de genes asociados a enfermedad facilitarán la detección de este tipo de variantes y darán una estimación real de su impacto en la etiopatogenia de las enfermedades hereditarias.

Como se ha indicado previamente, existen también variantes exónicas que alteran el *splicing*. Estas mutaciones pueden generar nuevos sitios de *splicing*, activar sitios crípticos o alterar motivos de unión a factores reguladores. Los dos primeros tipos de variantes provocan la pérdida de regiones exónicas debido a la utilización de sitios no canónicos de *splicing* en el exón. El tercer tipo, las que alteran secuencias reguladoras, habitualmente dan lugar a *skipping* del exón debido a la creación de ESSs o a la pérdida de ESEs. Puesto que se trata de variantes exónicas, éstas suelen pasar desapercibidas como variantes sinónimas, *missense* o nonsense. Sin embargo, como se ha explicado anteriormente, se estima que hasta un 20% de las variantes *missense* se sitúan en SREs. De hecho, 18 de las 26 variantes exónicas de *splicing* encontradas en el gen *NF1* alteran secuencias de unión a factores de *splicing*, especialmente ESEs (Wimmer et al., 2007). Normalmente, las variantes de *enhancer* suelen tener un efecto parcial, es decir, generan un transcrito de longitud adecuada que porta la mutación exónica, y uno o varios transcritos con *splicing* aberrante. Por tanto, la patogenicidad de estas variantes dependerá de la cantidad de transcrito aberrante, del mantenimiento de la ORF y de la conservación de las secuencias alteradas.

Según HGMD (*The Human Gene Mutation Database*), aproximadamente un 9% de las variantes reportadas son variantes de *splicing* (agosto, 2018). Sin embargo, esta cifra es probablemente una subestimación, pues la mayoría de las variantes de *splicing*, por ser exónicas, no se encuentran catalogadas como tal. En las bases de datos internacionales, encontramos como variantes de *splicing* aquellas que alteran los sitios aceptor o donador y aquellas que generan nuevos sitios de *splicing*. Las variantes reguladoras que fortalecen o activan sitios crípticos, así como aquellas que crean ESSs o eliminan ESEs, normalmente pasan desapercibidas. Todas estas alteraciones de *splicing* pueden llevar al desarrollo de enfermedades raras de alta penetrancia con herencia Mendeliana. Sin embargo, lo más frecuente es que tengan impacto en enfermedades complejas más comunes, y que se

encuentren diferencias entre individuos en términos de riesgo, penetrancia y repuesta terapéutica. Por ejemplo, en el Cáncer, las alteraciones de *splicing* son una característica común en genes asociados a migración, regulación del crecimiento celular, respuesta a estímulos o apoptosis (Wang and Cooper, 2007). Determinadas mutaciones pueden alterar el patrón de *splicing*, alterando la proteína y estando implicadas en el desarrollo de una enfermedad. En concreto, en *BRCA2* se han encontrado variantes asociadas al riesgo de padecer CMOH que alteran los sitios canónicos de *splicing* en todos los exones, excepto en el 4, el 10, el 11, 12 y el 20 (Fackenthal et al., 2016).

En este trabajo se pretende elucidar las secuencias reguladoras de la transcripción y el *splicing* del gen *BRCA2*. Además, se pretende ensayar y clasificar aquellas variantes candidatas de alterar dichos mecanismos. Se tratará de evaluar las variantes encontradas en pacientes incluidas en el sistema de prevención de cáncer de Castilla y León (España) y las reportadas en las principales bases de datos (ClinVar, BIC, UMD y Ensembl) en *BRCA2* mediante ensayos funcionales en células MCF-7, una línea celular humana de adenocarcinoma de mama. Para estudiar las variantes situadas en el promotor se realizará un constructo basado en el vector pGL4.10, con el gen de luciferasa bajo el promotor de *BRCA2*, al que llamaremos pGL4.10-prBR2. Para ensayar las variantes de *splicing*, se construirán los minigenes MGBR2\_2-9 y MGBR2\_14-20 sobre el plásmido reportero pSAD (Patente P201231427-CSIC). Así, durante la realización de la tesis doctoral se diseñarán y construirán vectores estables para el estudio de transcripción y *splicing*, se llevarán a cabo técnicas de cultivo celular y se aplicarán métodos precisos para analizar los transcritos generados. Además, se realizarán estudios de interacción RNA-proteína en el *International Centre for Genetic Engineering and Biotechnology* en Trieste (Italia), con el fin de conocer las proteínas involucradas en la regulación del *splicing* de *BRCA2*.



# **JUSTIFICACIÓN**



## JUSTIFICACIÓN

---

Desde el descubrimiento de los genes *BRCA* no han parado los esfuerzos para mejorar el diagnóstico molecular del CMOH. Actualmente, debido a la democratización de las nuevas técnicas de secuenciación masiva, cada día aumenta el número de variantes reportadas en estos genes. Se estima que la frecuencia de variantes en los genes *BRCA* se sitúa en torno a 1/250 mujeres (Narod and Foulkes, 2004). Las variantes patogénicas en *BRCA1/2* incrementan significativamente el riesgo de padecer cáncer de mama y/u ovario hasta en un 87% y un 63%, respectivamente, a la edad de 70 años (Petrucelli et al., 2016). Sin embargo, uno de los mayores problemas a la hora del diagnóstico de CMOH es la falta de información sobre el papel funcional de las variantes encontradas. Así, una gran parte de las variantes de *BRCA* están clasificadas como VUS. La clasificación de variantes no es sencilla y, normalmente, tienden a catalogarse en función del efecto que producirían en la proteína. Sin embargo, muchas de ellas se encuentran en regiones no codificantes como las regiones UTR y los intrones, o bien provocan alteraciones de secuencias consenso en elementos reguladores del gen. De hecho, aproximadamente el 90% de las variantes asociadas a enfermedad están localizadas fuera de las regiones codificantes (Hindorff et al., 2009), lo que sugiere la gran importancia de los mecanismos reguladores en el metabolismo de los ácidos nucleicos.

Estudios previos han demostrado que las variantes en el promotor pueden alterar los niveles de proteína y estar implicadas en el desarrollo de enfermedades (de Vooght et al., 2009). Sin embargo, a pesar de la importancia de la región promotora, ésta no está normalmente incluida en el rastreo rutinario de los genes *BRCA* en pacientes de alto riesgo a padecer CMOH. Las secuencias de unión a factores de transcripción son elementos esenciales en el promotor de un gen, de manera que variaciones en dichas secuencias podrían suponer la desregulación de la tasa de transcripción. De hecho, se han detectado variantes tanto en el promotor de *BRCA1* (c.-287C>T, TSS) como en *BRCA2* (c.-296C>T, TSS) que alteran secuencias de unión a los factores de transcripción NFYA y PAX5, respectivamente (Burke et al., 2018).

Por otro lado, el *splicing* de *BRCA2* parece ser un proceso altamente regulado, pues se han detectado 24 isoformas fisiológicas distintas (Fackenthal et al., 2016). Además, nuestro grupo y otros, han identificado previamente variantes en *BRCA2* que afectan al procesamiento del mRNA o *splicing* (Acedo et al., 2012; Rebbeck et al., 2018; Sanz et al., 2010; Spurdle et al., 2008). Esto nos hace pensar en la presencia de elementos conservados en *cis* que actúen regulando el *splicing* y que puedan verse alterados debido a cambios en la secuencia de DNA.

Por tanto, con el objetivo de ampliar el conocimiento sobre la regulación del gen *BRCA2* y de mejorar en el diagnóstico precoz del CMOH, en este trabajo se pretende elucidar las secuencias y los mecanismos implicados en la transcripción y el *splicing* de este gen, así como la caracterización funcional y clínica de las variantes que puedan afectar a estos procesos.

# OBJETIVOS



## OBJETIVOS

---

El objetivo central de esta tesis es la caracterización de secuencias funcionales implicadas en la regulación de la transcripción y el *splicing* de *BRCA2*, así como clasificación de variantes genéticas que puedan estar desregulando dichos mecanismos.

Objetivos específicos a nivel de transcripción:

1. Construcción y validación de un vector reportero estable con el promotor de *BRCA2*.
2. Mapeo de las secuencias reguladoras del promotor de *BRCA2*.
3. Estudio funcional y clasificación clínica de variantes en el promotor de *BRCA2*.

Objetivos específicos a nivel de *splicing*:

1. Construcción y validación de minigenes estables con los exones 2-9 (MGBR2\_2-9) y 14-20 (MGBR2\_14-20) de *BRCA2*.
2. Análisis de las regiones reguladoras de los exones 2-9 y 14-18 del gen *BRCA2*.
3. Análisis bioinformático y selección de variantes candidatas de alterar el *splicing* en los exones 2-9 y 14-18 de *BRCA2*.
4. Estudio funcional y clasificación clínica de variantes de *splicing* en *BRCA2*.



# **ARTÍCULOS**



# ARTÍCULOS

---

## ARTÍCULO 1

Eugenia Fraile-Bethencourt, Beatriz Díaz-Gómez, Valeria Velásquez-Zapata, Alberto Acedo, David J. Sanz, Eladio A. Velasco.

**Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of *BRCA2* exons 17 and 18.**

PLOS Genetics 2017; 13: e1006691 <https://doi.org/10.1371/journal.pgen.1006691>

*Impact factor 2016: 6.100*

**Introducción:** Mediante el rastreo de los genes de cáncer de mama *BRCA1* y *BRCA2* se han identificado un gran número de variantes de significado clínico desconocido (VUS) cuya interpretación funcional y clínica supone un reto para la Genética Médica. Asimismo, cada vez son más las evidencias que demuestran la implicación de determinadas variantes en la alteración del procesamiento del mRNA. Nuestro objetivo se centra en investigar el impacto en el *splicing* de las variantes encontradas en los exones 17 y 18 de *BRCA2* y evaluar su papel en el Cáncer de Mama y Ovario Hereditario, así como identificar regiones críticas con elementos reguladores (*enhancers de splicing*) que puedan constituir *hotspots* de variantes con efecto en el *splicing*.

**Materiales y métodos:** Se construyó un minigen basado en el vector pSAD (Patente P201231427, CSIC) que incluía los exones del 14 al 20 de *BRCA2* (MGBR2\_ex14-20). La presencia de elementos reguladores se analizó a través de microdeleciones solapantes en el comienzo y final de los exones 17 y 18. Se realizó un análisis bioinformático con programas de predicción de *splicing*. Se seleccionaron 52 variantes candidatas que fueron introducidas en el minigen mediante mutagénesis dirigida y ensayadas por triplicado en células MCF-7.

**Resultados:** El minigen *wild type* produjo un transcrito estable con el tamaño (1.806 nt) y la estructura (V1-[*BRCA2*\_exones\_14-20]-V2) esperados. El mapeo funcional mediante microdeleciones reveló secuencias esenciales para el reconocimiento exónico en el extremo

3' del exón 17 (c.7944\_7973) y en el extremo 5' del exón 18 (c.7979\_7988, c.7999\_8013). De las 52 variantes seleccionadas, 30 provocaban un *splicing* anómalo en el minigen, encontrándose más de 16 isoformas distintas, de las cuales, la pérdida del exón o *exon skipping* era el evento más común. Los motivos de *splicing* afectados por las variantes fueron muy variados, encontramos 15 variantes que alteraban sitios de *splicing*, tres que generaban sitios alternativos de *novos*, tres que afectaban al tracto de polipirimidinas y nueve que modificaban *enhancers* y/o silenciadores. De acuerdo a las reglas propuestas por ACMG (*American College of Medical Genetic and Genomics*), 20 variantes podrían ser clasificadas como patogénicas (c.7806-2A>G, c.7806-1G>A, c.7806-1G>T, c.7806-1\_7806-2dup, c.7976+1G>A, c.7977-3\_7978del, c.7977-2A>T, c.7977-1G>T, c.7977-1G>C, c.8009C>A, c.8331+1G>T y c.8331+2T>C) o probablemente patogénicas (c.7806-9T>G, c.7976G>C, c.7976G>A, c.7977-7C>G, c.7985C>G, c.8023A>G, c.8035G>T y c.8331G>A), alcanzando el 30,8% de todas las variantes patogénicas o probablemente patogénicas reportadas en UMD (*The Universal Mutation Database*) en los exones 17 y 18 de *BRCA2*. Otras ocho variantes (c.7975A>G, c.7977-6T>G, c.7988A>T, c.7992T>A, c.8007A>G, c.8009C>T, c.8009C>G y c.8072C>T) inducían anomalías parciales en el *splicing* manteniendo altos ratios de transcrito canónico (>70%), por lo que se mantuvo su clasificación como VUS.

**Conclusiones:** El *splicing* aberrante es especialmente prevalente en los exones 17 y 18 de *BRCA2* debido a la presencia de elementos reguladores involucrados en el reconocimiento de los exones. Los ensayos funcionales mediante minigenes representan una valiosa estrategia para la caracterización inicial de los efectos en el *splicing* y la subsecuente interpretación clínica de las variantes genéticas. Sin embargo, siempre y cuando sea posible, estos resultados deberían ser contrastados con el RNA de los pacientes portadores.

RESEARCH ARTICLE

# Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of *BRCA2* exons 17 and 18

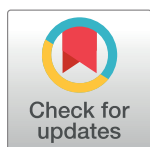
Eugenia Fraile-Bethencourt, Beatriz Díez-Gómez, Valeria Velásquez-Zapata, Alberto Acedo<sup>¶a</sup>, David J. Sanz<sup>¶b</sup>, Eladio A. Velasco\*

Splicing and genetic susceptibility to cancer, Instituto de Biología y Genética Molecular (CSIC-UVa), Valladolid, Spain

<sup>¶a</sup> Current address: AC-GEN Reading Life SL, Valladolid, Spain

<sup>¶b</sup> Current address: Department of Physiology, University College Cork, Cork, Ireland

\* [eavelsam@ibgm.uva.es](mailto:eavelsam@ibgm.uva.es)



## OPEN ACCESS

**Citation:** 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(3): e1006691. <https://doi.org/10.1371/journal.pgen.1006691>

**Editor:** Charis Eng, Cleveland Clinic Genomic Medicine Institute, UNITED STATES

**Received:** July 25, 2016

**Accepted:** March 14, 2017

**Published:** March 24, 2017

**Copyright:** © 2017 Fraile-Bethencourt et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files except the following files: Insert sequence of minigene (*BRCA2* exons 14 to 20); Complete List of DNA variants from such exons and flanking intronic sequences; Bioinformatic analysis of variants with Human Splicing Finder; Fragment analysis by capillary array electrophoresis; Sequencing traces of RT-PCR products, which can be found at <https://figshare.com/s/b4f226871198dc53db83>.

## Abstract

Mutation screening of the breast cancer genes *BRCA1* and *BRCA2* identifies a large fraction of variants of uncertain clinical significance (VUS) whose functional and clinical interpretations pose a challenge for genomic medicine. Likewise, an increasing amount of evidence indicates that genetic variants can have deleterious effects on pre-mRNA splicing. Our goal was to investigate the impact on splicing of a set of reported variants of *BRCA2* exons 17 and 18 to assess their role in hereditary breast cancer and to identify critical regulatory elements that may constitute hotspots for spliceogenic variants. A splicing reporter minigene with *BRCA2* exons 14 to 20 (MGBR2\_ex14-20) was constructed in the pSAD vector. Fifty-two candidate variants were selected with splicing prediction programs, introduced in MGBR2\_ex14-20 by site-directed mutagenesis and assayed in triplicate in MCF-7 cells. Wild type MGBR2\_ex14-20 produced a stable transcript of the expected size (1,806 nucleotides) and structure (V1-[*BRCA2*\_exons\_14–20]–V2). Functional mapping by microdeletions revealed essential sequences for exon recognition on the 3' end of exon 17 (c.7944-7973) and the 5' end of exon 18 (c.7979-7988, c.7999-8013). Thirty out of the 52 selected variants induced anomalous splicing in minigene assays with >16 different aberrant transcripts, where exon skipping was the most common event. A wide range of splicing motifs were affected including the canonical splice sites (15 variants), novel alternative sites (3 variants), the polypyrimidine tract (3 variants) and enhancers/silencers (9 variants). According to the guidelines of the American College of Medical Genetics and Genomics (ACMG), 20 variants could be classified as pathogenic (c.7806-2A>G, c.7806-1G>A, c.7806-1G>T, c.7806-1\_7806-2dup, c.7976+1G>A, c.7977-3\_7978del, c.7977-2A>T, c.7977-1G>T, c.7977-1G>C, c.8009C>A, c.8331+1G>T and c.8331+2T>C) or likely pathogenic (c.7806-9T>G, c.7976G>C, c.7976G>A, c.7977-7C>G, c.7985C>G, c.8023A>G, c.8035G>T and c.8331G>A), accounting for 30.8% of all pathogenic/likely pathogenic variants of exons 17–18 at the BRCA Share database. The remaining 8 variants (c.7975A>G, c.7977-6T>G, c.7988A>T, c.7992T>A, c.8007A>G, c.8009C>T, c.8009C>G, and c.8072C>T) induced

**Funding:** EAV's lab was supported by grants from the Spanish Ministry of Economy and Competitiveness, Plan Nacional de I+D+I 2013-2016, ISCIII (Fis: PI13/01749) co-funded by FEDER from Regional Development European Funds (European Union), and grant CSI090U14 from the Consejería de Educación (ORDEN EDU/122/2014), Junta de Castilla y León. EFB was supported by a predoctoral fellowship from the University of Valladolid and Banco de Santander (2015-2019). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

partial splicing anomalies with important ratios of the full-length transcript ( $\geq 70\%$ ), so that they remained classified as VUS. Aberrant splicing is therefore especially prevalent in *BRCA2* exons 17 and 18 due to the presence of active ESEs involved in exon recognition. Splicing functional assays with minigenes are a valuable strategy for the initial characterization of the splicing outcomes and the subsequent clinical interpretation of variants of any disease-gene, although these results should be checked, whenever possible, against patient RNA.

## Author summary

A significant proportion of disease-causing mutations of inherited disorders impair splicing. Massive sequencing projects of genetic diseases generate thousands of sequence variations that require functional and clinical interpretations. We have shown that splicing reporter minigenes of the breast cancer genes *BRCA1* and *BRCA2* are useful tools to functionally test DNA variants. In this work, we have constructed a 7-exon *BRCA2* minigene (exons 14 to 20) where we mapped critical splicing regulatory sequences and tested 52 selected variants of exons 17 and 18 detected in breast cancer patients. We finely located three DNA segments on both exons that presumably contain splicing enhancer sequences. We observed that a total of 30 variants of any type disrupted the splicing patterns and, given the severity of their outcomes, we classified 20 of them as pathogenic or likely pathogenic. We also showed that a wide range of splicing elements were affected including canonical and novel 5' and 3' splice sites, the polypyrimidine tract and enhancer and silencer sequences. We concluded that splicing aberrations are frequent in Hereditary Breast and Ovarian Cancer and that minigenes are valuable tools to functionally classify DNA variants of any human disease gene under the splicing viewpoint.

## Introduction

Germline pathogenic variants in the tumor suppressor genes *BRCA1* (MIM# 113705) and *BRCA2* (MIM# 600185) are associated with increased risk of breast and ovarian cancer [1,2], and account for about 16% of the familial risk for breast cancer [3]. More than 25 breast cancer susceptibility genes have been identified so far, most of which play a role in the DNA repair pathway linked to *BRCA1* and *BRCA2* [4]. Additionally, a vast number of SNPs have been associated with breast cancer risk [5,6], increasing the complexity of the genetic landscape of Hereditary Breast/Ovarian Cancer (HBOC). Moreover, according to the BRCA Share Database (<http://www.umd.be/BRCA2/>; last accessed date, April 2016), more than 2700 different sequence variations have been reported at the *BRCA2* gene, ~30% of which are causal. A large proportion of the recorded pathogenic variants truncate the *BRCA2* protein (nonsense and frameshift). However, up to 20% of *BRCA1/2* tests report variants of uncertain clinical significance (VUS) [7]. These pose a challenge in genetic counselling as VUS-carrier families are usually considered as negative (undetermined) so they cannot benefit from prevention protocols [8].

In fact, other factors must be involved in the pathogenesis of genetic disorders since gene expression is regulated by a wide range of cis-regulatory sequences that control it, as for example, transcription initiation (promoter) [9], pre-mRNA splicing [10] or post-transcriptional regulation and mRNA stability (3'UTR) [11]. It is therefore expected that point mutations in

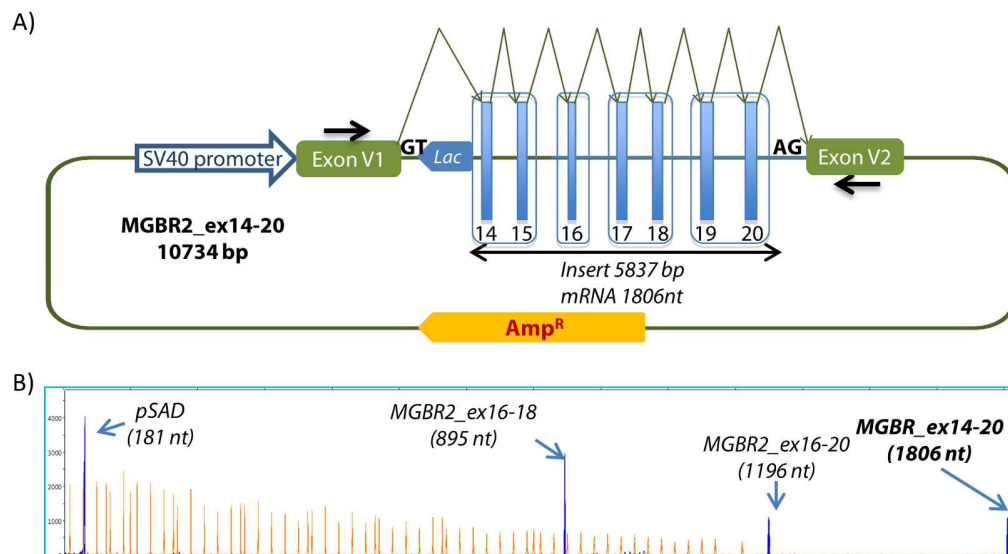
those motifs can be correlated with gene expression alterations and disease. Interestingly, nearly 90% of disease-associated SNPs are placed outside protein-coding regions (45% intronic, 43% intergenic), suggesting a relevant role of the non-coding sequence variations [12]. Splicing is a central process of gene expression whereby introns are excised and exons are joined sequentially. It has been calculated that >90% of mammalian genes undergo alternative splicing which is controlled by a dense array of diverse cis-acting elements and splicing factors [10,13]. According to GENCODE (V.24, <http://www.genencodegenes.org/stats/current.html>), the average number of protein coding transcripts per gene is ~4. In this regard, it has been recently reported the existence of 24 naturally occurring alternative splicing events of the *BRCA2* gene [14]. Alternative splicing not only allows transcriptome and proteome diversity, but it also regulates important processes such as embryonic development or cell differentiation. Exon recognition requires specific signals at the 5' and 3' splice sites, the polypyrimidine tract, the branch point, and supplementary sequences referred to as Exonic Splicing Enhancers (ESE) and Silencers (ESS) [15]. Interestingly, an unexpectedly large fraction of variants can actually disrupt pre-mRNA processing [16,17]. Remarkably, aberrant splicing is common in cancer so that it can be considered a hallmark of this disease [18]. We previously estimated that a significant proportion of likely pathogenic *BRCA1/2* variants (33.9%) from 14 exons would impair splicing [19].

The most suitable method to identify splicing aberrations is based on the study of patient RNA of the affected tissue although this sort of sample is not always available [20,21]. Nevertheless, direct RNA analysis of several disease genes, including *NF1* (MIM# 613113) and *BRCA1/2*, has proven to be a highly sensitive method to identify splicing anomalies [22,23]. Minigene-based technologies have become alternative approaches to primarily test whether a specific DNA variant affects splicing, especially when patient samples cannot be collected. In a recent report we presented the new splicing reporter plasmid pSAD (Patent P201231427-CSIC) that constituted the backbone of the largest *BRCA2* minigene ever reported with 9 exons (19 to 27) [24]. This construct allowed the analysis of 40 variants spread throughout these exons and flanking introns, demonstrating its capability for the successful identification and characterization of *spliceogenic* variants. Moreover, up to date, we have examined the impact on splicing of 112 different DNA variants by minigene assays, 51 of which induced aberrant splicing patterns [19,24,25].

The intronic GT dinucleotide in positions +1 and +2 is the most conserved element of the donor splice signal. However, in a small fraction of the donor sites (<1%), GT is replaced by GC that are rather located in alternatively spliced introns [26,27]. Recognition of a GC donor site at the *BTB* gene was associated with splicing enhancers for SR proteins 9G8, Tra2 $\beta$  and SC35 [27]. The *BRCA2* intron 17 has also a 5' GC motif and there have been identified minor natural alternative splicing isoforms  $\Delta 18$  (exon 18 skipping) and  $\Delta 17,18$  (exons 17+18 skipping) [14,28]. To study the regulatory mechanisms of both exons and how DNA variants affect this process we constructed a large minigene with exons 14 to 20 in the pSAD plasmid (MGBR2\_ex14-20) to keep the genomic context. Splicing regulatory elements were searched by functional microdeletion mapping. Finally, 52 variants detected in HBOC patients were selected and assayed in the minigene MGBR2\_ex14-20.

## Results

The minigene MGBR2\_ex14-20 (10,734 bp) was built as indicated in materials and methods. The insert (5,837 bp) corresponds to a genomic region of 16,762 bp of the *BRCA2* gene (Fig 1A). To validate it, we transfected the final (14–20) and the intermediate constructs (17–18, 16–18 and 16–20) into MCF-7 cells, we isolated the RNA and performed the RT-PCR with



**Fig 1. Structure and functional analysis of the minigene MGBR2\_ex14-20.** A) Structure of the minigene MGBR2\_ex14-20 (slashes indicate shortened introns): [IVS14 (328 pb)–EX14 (428 pb)–IVS14 (1139 pb)–EX15 (182 pb)–IVS15 (358 pb) // IVS15 (333 pb)–EX16 (188 pb)–IVS16 (234 pb) // IVS16 (181 pb)–EX17 (171 pb)–IVS17 (485 pb)–EX18 (355 pb)–IVS18 (314 pb) // IVS18 (235 pb)–EX19 (156 pb)–IVS19 (398 pb)–EX20 (145 pb)–IVS20 (207 pb)]. Boxes highlight the 4 different cloning steps of this minigene. The expected splicing reactions in eukaryotic cells are indicated by arrows. B) Splicing functional analysis of the empty vector pSAD and the final (14–20) and intermediate (16–18, 16–20) constructs in MCF-7 cells. The four electropherograms were overlaid. cDNAs were amplified with vector exon specific primers SD6-PSPL3\_RTFW and RTpSAD-RV (arrows within vector exons V1 and V2 above). Full-length transcripts are shown as blue peaks and the Genescan Liz-1200 size standard is shown as orange/faint peaks. Fragment sizes (bp) and relative fluorescent units are indicated on the x- and y-axes, respectively.

<https://doi.org/10.1371/journal.pgen.1006691.g001>

specific primers at the vector-exons. Fig 1B shows the transcripts generated by the pSAD vector, two intermediate (16–18 and 16–20) and the final constructs of this minigene. MGBR2\_ex14-20 produced a full-length transcript of the expected size (1,806 nucleotides—nt-) and structure (V1-*BRCA2* exons 14 to 20-V2) without any anomalies, so this minigene was ready for regulatory studies and variant analysis. The switch of the GC site of exon 17 to a strong canonical GT site (artificial variant c.7976+2C>T; NNSplice = 1.0) rendered the same canonical transcript without any anomalies (9 independent assays), like the GC counterpart (S1 Fig), that was confirmed by sequencing. Finally, exons 12 and 13 were also cloned in two independent stages but splicing reactions did not produce the expected canonical transcripts so they were ruled out.

## Mapping of exonic and intronic splicing regulatory sequences

The efficient inclusion of the exons in the mature mRNA may require the presence of enhancer sequences and the binding to SR-proteins [29]. The highest density of active ESEs is near splice sites (~50 nt at both exon ends) with a maximum between 10 and 20 nucleotides from the canonical 5' and 3' splice sites of each exon [30]; so DNA variants at these regions have a higher likelihood of disrupting ESEs. Furthermore, previous studies suggested a specific regulation of the donor GC-sites by ESEs [27]. We therefore proceeded to map regulatory sequences involved in exons 17 and 18 processing by functional tests of four exonic 30-nt deletions of each exon. These deletions covered the 5' and 3' 55 nucleotides of each exon excluding the first two and last three nucleotides. Three 30-nt microdeletions, c.7944\_7973del (exon 17),

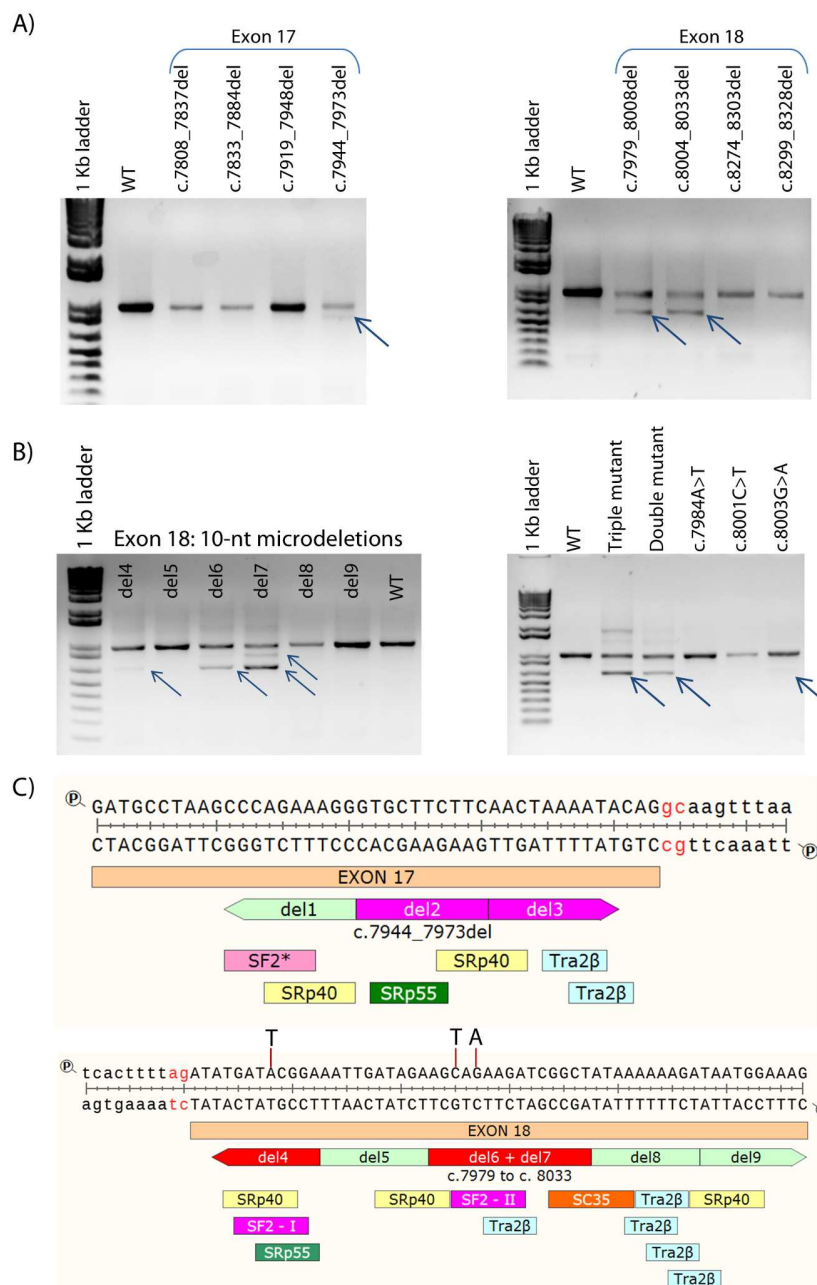
c.7979\_8008del (exon 18) and c.8004\_8033del (exon 18), had impacts on splicing (exon 17 or 18 skipping) (Fig 2A), indicating that these sequences probably contain regulatory motifs guiding exon recognition. According to the ESEfinder algorithm [31], these sequences contain several putative enhancer sequences (see Fig 2C), but only SF2/ASF and SRp40 motifs were present in the three deletions suggesting that these SR proteins might be required for competent exon identification. We then carried out the fine mapping of ESEs with additional internal 10-nt deletions of exons 17 [c.7944\_7953del (del1), c.7954\_7963del (del2) and c.7964\_7973del (del3)] and 18 [c.7979\_7988del (del4), c.7989\_7998del (del5), c.7999\_8008del (del6), c.8004\_8013del (del7; 5-nt overlap between del6 and del7), c.8014\_8023del (del8) and c.8024\_8033del (del9)] (S2 Table; Fig 2B). The del2 and del3 deletions of exon 17 only disrupted splicing weakly (2.2% and 4.6% of aberrant transcripts, respectively; S3 Table) whereas del1 did not at all. Exon 18 skipping was found at del4, del6 and del7. These segments must therefore contain splicing enhancer sequences. Deletions 5, 8 and 9 of exon 18 did not affect splicing and produced the expected transcripts.

According to the bioinformatics predictions of ESEs (Fig 2C), the common feature of the three positive exon 18 deletions was the presence of two putative ESEs for SF2/ASF (one in the overlapping segment of del6-7) at nucleotides c.7981\_7987 (GATACGG) and c.8001\_c.8007 (CAGAAGA). We then proceeded to disrupt both motifs by mutagenesis with the following sequence variations: c.7984A>T, c.8001C>T and c.8003G>A, which target conserved nucleotides of the two SF2 motifs (Fig 2C, S2 Table) [31]. We carried out the assays with the triple mutant, a double mutant c.[8001C>T;8003G>A], which targeted the SF2-II site, and the three independent variants (Fig 2B and 2C). Intriguingly, only the triple and the double mutants remarkably impaired splicing with 59.7% and 44.2% of aberrant transcripts (Fig 2B), respectively, but these effects were not observed either with c.7984A>T or c.8001C>T, which did not apparently affect splicing, whereas c.8003G>A only induced weak exon 18 skipping (8.8%), suggesting a synergistic effect of these variants on splicing. We can conclude that these sequences are required for exon 18 recognition so that any variation in these nucleotides may affect splicing and confer breast cancer risk. In order to investigate the participation of SF2/ASF in this process, we initially performed inhibition experiments with siRNAs of splicing factors SF2/ASF, SC35 and Tra2 $\beta$ . Preliminary data unexpectedly suggested a role for SC35 in exons 17 and 18 definitions as well as small contributions of Tra2 $\beta$  and SF2/ASF (S2 Fig). In fact, a putative SC35 motif (GGCTATAA, c.8010-8017) was located at the spliceogenic del7 deletion (Fig 2C).

We also searched for ESE sequences within intron 17. Firstly, we selected a region of 115 nucleotides (c.7976+231\_7977-141del) where Human Splicing Finder (HSF) had predicted the presence of a notable concentration of high-scored Tra2 $\beta$  sites (S2 Table), being Tra2 $\beta$  one of the SR-proteins involved in GC recognition in the *BTK* gene [27]. The presence of SREs was also checked in the rest of the intron 17 with another three deletions, c.7976+21\_7976+140del, c.7976+136\_7976+240del and c.7977-150\_7977-21del. However, none of them had an impact on splicing suggesting that regulatory elements of the GC site are not located within intron 17.

## Identification of spliceogenic variants

A total of 221 reported DNA variants (BRCA Share, BIC and 1000 Genomes databases, last accessed date: April 2016) were analyzed with NNSPLICE and HSF. We selected fifty-four out of them (24.4%) on basis of these criteria: splice site disruption or modification, creation of alternative splice sites, disruption of an ESE within positive 30-nt microdeletions, creation of silencers (specifically hnRNPA1 sites) (S2 Table). Remarkably, 36 variants had previously been classified as VUS by the BRCA Share and BIC databases. All the selected DNA changes were



**Fig 2. Map of splicing enhancers for recognition of exons 17 and 18.** A) Functional mapping of ESEs of exons 17 (left) and 18 (right) by exonic microdeletions (30-nt above) of the wild type minigene MGBR2\_ex14-20. cDNA was amplified with primers of *BRCA2* exon 16 (RTBR2\_ex16-FW) and vector exon V2 (RTpSAD-RV; amplicon size = 1012 nt) and run in 1.5% agarose gel. Arrows indicate abnormal transcripts. B) Functional analysis of 10-nt microdeletions of exon 18 (left) and single nucleotide substitutions (right) c.7984A>T that disrupts SF2-I, and c.8001C>T and 8003G>A that target SF-II motif (see ESE map below). The triple mutant is the combination c.[7984A>T;8001C>T;8003G>A], and the double one is c.[8001C>T;8003G>A]. C) HSF predictions of putative ESE motifs of exon 17 (above) between cDNA positions 7944 and 7973 and exon 18 (below) between cDNA positions 7979–8008 and 8004–8033. Intronic sequences are in lower case. SF2\* motif of exon 17 is detected by the specific SF2/ASF (IgM-*BRCA1*) algorithm of ESEfinder. Red and pink microdeletions alter splicing. Artificial SF2- single-nucleotide substitutions are indicated above exon 18 sequence.

<https://doi.org/10.1371/journal.pgen.1006691.g002>

introduced into the wild type MGBR2\_ex14-20 construct by site-directed mutagenesis except for two (c.7829dup—exon17- and c.8169\_8172dup—exon18-) that were disregarded because of the recurrent failure of the mutagenesis experiments.

Fifty-two variants (17 of exon 17 and 35 of exon 18) were functionally tested in the splicing reporter minigene MGBR2\_ex14-20. All the transcripts were quantified to evaluate their possible implication in disease pathogenesis (S3 Table). Only variants with  $\geq 5\%$  of anomalous transcripts were considered as positive. Thirty DNA variants (57.7%) impaired splicing (Table 1, S3 Table, Figs 3 and 4, S3 Fig), whilst another three variants (c.7875A>G, c.7985C>T and c.8042C>G) had weak impacts on splicing (4.7%, 3.3% and 2.3%, respectively; S3 Table). All the splicing outcomes were highly reproducible with low intra-variability (standard deviations  $< 1.8\%$  for 27 variants; S3 Table). Spliceogenic variants consisted of 14 intronic and 16 exonic variants that had previously been predicted as 11 missense, 1 nonsense, 1 frameshift and 3 synonymous variants, confirming that any type of genetic variant can potentially disrupt pre-mRNA processing.

According to the prediction software and their location, 15 variants disrupted the canonical 3' and 5' splice sites (c.7806-2A>G, c.7806-1G>A, c.7806-1G>T, c.7806-1\_7806-2dup—previously reported as c.7806insAG-, c.7975A>G, c.7976G>C, c.7976G>A, c.7976+1G>A, c.7977-3\_7978del, c.7977-2A>T, c.7977-1G>T, c.7977-1G>C, c.8331G>A, c.8331+1G>T and c.8331+2T>C), three disrupted the polypyrimidine tract (c.7806-9T>G, c.7977-7C>G and c.7977-6T>G), three created novel active splice sites (c.8023A>G, c.8035G>T and c.8168A>G, but also c.7977-7C>G and c.7806-1\_7806-2dup—see above-), seven affected enhancer or silencer motifs (c.7992T>A, c.8007A>G, c.8009C>A, c.8009C>T, c.8009C>G, c.8072C>T and c.8249\_8250del, all of them in exon 18) and two were presumed to alter ESE/ESS motifs and generate alternative sites (c.7985C>G—weak 3'ss- and c.7988A>T—strong 5'ss-) that actually were not used, so they should be considered as ESE/ESS-variants. Seven of the ESE/ESS variants were placed into the positive ESE-containing microdeletions c.7979\_8008del30 and c.8004\_8033del30, spanning a 25-nt interval of exon 18 (c.7985-8009).

## Characterization of aberrant transcripts

Variants of exon 17 and flanking intronic sequences rendered 6 different abnormal transcripts (Fig 5): ex17 skipping, ex17-ins8 (alternative intronic acceptor 8 nt upstream), ex17-del1 (novel acceptor 1 nt downstream), ex17-del20 (alternative acceptor 20 nt downstream), ex17-del69 (alternative acceptor 69 nt downstream) and ex17-insAG (novel acceptor 2 nt upstream), where exon 17 skipping was the most abundant event. Mutations at exon 18 and contiguous sequences induced more than 10 different aberrant transcripts (Fig 5): ex18 skipping, ex18-ins6 (novel intronic acceptor 6 nt upstream), ex18-del191 (alternative acceptor 191 nt downstream), ex18-del309 (new donor 309 nt upstream), ex18-del298 (new donor 298 nt upstream), ex18-del164 (new donor 164 nt upstream), ex18-del157 (use of cryptic donor 157 nt upstream), and rare phenomena such as ex17-del20+ex18 skipping (cryptic acceptor plus skipping), ivs17\_58 nt retention+ex18 skipping (intronic cryptic donor plus skipping), one 878-nt transcript as well as other uncharacterized aberrant transcripts. Exon 18 skipping was the most frequent outcome (19 out of 21 variants induced it). Twelve transcripts would introduce premature termination codons (PTC).

## Discussion

The functional and clinical classifications of DNA variants of breast cancer genes provide essential information for clinical management of patients and asymptomatic carriers. The identification of VUS in patients hampers the genetic counseling of BC families since the result

**Table 1. Splicing outcomes of *BRCA2* exons 17 and 18 variants.**

| DNA variant <sup>1</sup> | Motif <sup>2</sup> | Splicing outcome <sup>3</sup>                                 | RNA effect <sup>4</sup>                                 | Protein Effect <sup>4</sup>                     |
|--------------------------|--------------------|---|---|---|
| c.7806-9T>G              | Pyr                | Ex17 skipping (41.5%); lvs16-ins8 (36.3%); Ex17-del69 (22.2%) | r.[7806_7976del,7805_7806ins7806-8_7806-1,7806_7874del] | p.[A2603_R2659del; R2602Sfs*49; A2603_R2625del] |
| c.7806-2A>G              | [-] 3'SS           | Ex17-del20 (51.8%); Ex17-del69 (28.1%); Ex17 skipping (20.1%) | r.[7806_7825del,7806_7874del,7806_7976del]              | p.[A2603Cfs*8; A2603_R2625del; A2603_R2659del]  |
| c.7806-1G>A              | [-] 3'SS           | Ex17-del1 (100%)  | r.[7806_7807del]  | p.A2603Lfs*45                                   |
| c.7806-1G>T              | [-] 3'SS           | Ex17-del20 (100%)   | r.7806_7825del  | p.A2603Cfs*8                                    |
| c.7806-1_7806-2dup       | [+]3'SS            | Ex17-insAG (92.6%); Ex17 skipping (5.1%); Ex17-del69 (2.3%)   | r.[7805_7806insAG,7806_7976del,7806_7874del]            | p.[A2603Gfs*46; A2603_R2659del; A2603_R2625del] |
| c.7975A>G                | [-] 5'SS           | CT (73.8%); Ex17 skipping (26.2%)                             | r.[7975a>g,7806_7976del]                                | p.[R2659G; A2603_R2659del]                      |
| c.7976G>C                | [-] 5'SS           | Ex17 skipping (100%)  | r.7806_7976del  | p.A2603_R2659del                                |
| c.7976G>A                | [-] 5'SS           | Ex17 skipping (100%)  | r.7806_7976del  | p.A2603_R2659del                                |
| c.7976+1G>A              | [-] 5'SS           | Ex17 skipping (100%)  | r.7806_7976del  | p.A2603_R2659del                                |
| c.7977-7C>G              | [+] 3'SS/<br>Pyr   | Ex18-ins6 (78.4%); exon 18 skipping (21.6%)                   | r.[7976_7977ins6,7977_8331del]                          | p.[Y2658_R2659insSF; Y2660Ffs*43]               |
| c.7977-6T>G              | Pyr                | CT (66.7%); Ex18 skipping (31%); ex18-del191 (2.3%)           | r.[=, 7977_8331del,7977_8167del]                        | p.[=; Y2660Ffs*43; Y2660Wfs*6]                  |
| c.7977-3_7978del         | [-] 3'SS           | Ex18 skipping (90%) ex18-del191 (10%)                         | r.[7977_8331del,7977_8167del]                           | p.[Y2660Ffs*43; Y2660Wfs*6]                     |
| c.7977-2A>T              | [-] 3'SS           | Ex18 skipping (93.3%); ex18-del191 (6.7%)                     | r.[7977_8331del,7977_8167del]                           | p.[Y2660Ffs*43; Y2660Wfs*6]                     |
| c.7977-1G>T              | [-] 3'SS           | Ex18 skipping (91.5%); ex18-del191 (7%); ex18-del236 (1.5%)   | r.[7977_8331del,7977_8167del,7977_8212del]              | p.[Y2660Ffs*43; Y2660Wfs*6;R2659Sfs*26]         |
| c.7977-1G>C              | [-] 3'SS           | Ex18 skipping (89.8%); Ex18-del191 (10.2%)                    | r.[7977_8331del,7977_8167del,?]                         | p.[Y2660Ffs*43; Y2660Wfs*6;?]                   |
| c.7985C>G                | [-]ESE/<br>[+]ESS  | Ex18 skipping (90.2%); Ex18-del191 (5%); others (4.8%)        | r.[7977_8331del,7977_8167del,?]                         | p.[Y2660Ffs*43; Y2660Wfs*6;?]                   |
| c.7988A>T                | [+]5'SS<br>[-]ESE  | CT (84.2%); Ex18 skipping (8.6%) + others (7.2%)              | r.[7988a>u,7977_8331del,?]                              | p.[E2663V;Y2660Ffs*43;?]                        |
| c.7992T>A                | [-]ESE/<br>[+]ESS  | CT (68.6%); ex18 skipping (31.4%)                             | r.[7992u>a,7977_8331del]                                | p.[=; Y2660Ffs*43]                              |
| c.8007A>G                | [-]ESE/<br>[+]ESS  | CT (84.8%); ex18 skipping (15.2%)                             | r.[8007a>g,7977_8331del]                                | p.[=; Y2660Ffs*43]                              |
| c.8009C>A                | [-]ESE/<br>[+]ESS  | Ex18 skipping (91.2%); Ex18-del191 (4.8%); CT (4%)            | r.[7977_8331del,7977_8167del,8009c>a,?]                 | p.[Y2660Ffs*43; Y2660Wfs*6;S2670*,?]            |
| c.8009C>T                | [-]ESE/<br>[+]ESS  | CT (76.6%); ex18 skipping (23.4%)                             | r.[8009c>u,7977_8331del]                                | p.[S2670L;Y2660Ffs*43]                          |
| c.8009C>G                | [-]ESE/<br>[+]ESS  | CT (79.9%); ex18 skipping (20.1%)                             | r.[8009c>g,7977_8331del]                                | p.[S2670W;Y2660Ffs*43]                          |
| c.8023A>G                | [+]5'SS            | Ex18-del309 (93%); other aberrant transcripts (7%)            | r.[8023_8331del,?]                                      | p.[Ile2675_K2777del;?]                          |
| c.8035G>T                | [+]5'SS            | Ex18-del298 (93.6%); 878-nt transcript (4%); CT: 2.4%         | r.[8034_8331del,?]                                      | p.[D2679Ffs*43;?]                               |
| c.8072C>T                | [-]ESE/<br>[±]ESS  | CT (94.9%); ex18 skipping (5.1%)                              | r.[8072c>u,7977_8331del]                                | p.[S2691F;Y2660Ffs*43]                          |
| c.8168A>G                | [+]5'SS            | CT (69.6%); Ex18-del164 (25.9%) /Ex18 skipping (4.5%)         | r.[8168a>g,8168_8331del,7977_8331del]                   | p.[D2723G;G2724Ffs*3; Y2660Ffs*43]              |
| c.8249_8250del           | [-]ESE/<br>[-]ESS  | CT (93.0%); ex18 skipping (7.0%)                              | r.[8249_8250del,7977_8331del]                           | p.[K2750Asnfs*13; Y2660Ffs*43]                  |

(Continued)

Table 1. (Continued)

| DNA variant <sup>1</sup> | Motif <sup>2</sup> | Splicing outcome <sup>3</sup>  | RNA effect <sup>4</sup>  | Protein Effect <sup>4</sup>                    |
|--------------------------|--------------------|--|--|--|
| <b>c.8331G&gt;A</b>      | [-] 5'SS           | Ex18 skipping (52%); <b>CT</b> (40.7%); aberrant transcripts (7.3%)  | r.[8331g>a,7977_8331del,?]   | p. [=; Y2660Ffs*43;?]                          |
| <b>c.8331+1G&gt;T</b>    | [-] 5'SS           | Ex18 skipping (81%); Ex18-del157 (6.4%); ex17-del151+ex18 skipping (6.1%); ivs17 58-nt retention+Ex18 skipping (3.7%); others (2.8%) | r.[7977_8331del,8175_8331del,7826_8331del,7977_8331delins7976+1_7976+58] | p. [Y2660Ffs*43;W2725*;G2609Dfs*4; Y2660Qfs*3] |
| <b>c.8331+2T&gt;C</b>    | [-] 5'SS           | Ex18 skipping (87.1%); ex17-del151+ex18 skipping (12.9%)   | r.[7977_8331del,7826_8331del]  | p.[Y2660Ffs*43;G2609Dfs*4]                     |

<sup>1</sup> Bold type variants indicate proposed causal or likely causal variants attending to the guidelines of the American College of Medical Genetics and Genomics (see also [S4 Table](#));

<sup>2</sup> Affected motifs: 3'SS, 3' splice site; 5'SS: 5' splice site; Pyr, polypyrimidine tract; ESE, Exonic Splicing Enhancer; ESS, Exonic Splicing Silencer; [-] disruption; [+] creation.

<sup>3</sup> The proportion of each transcript is indicated between parentheses; CT: Canonical transcript.

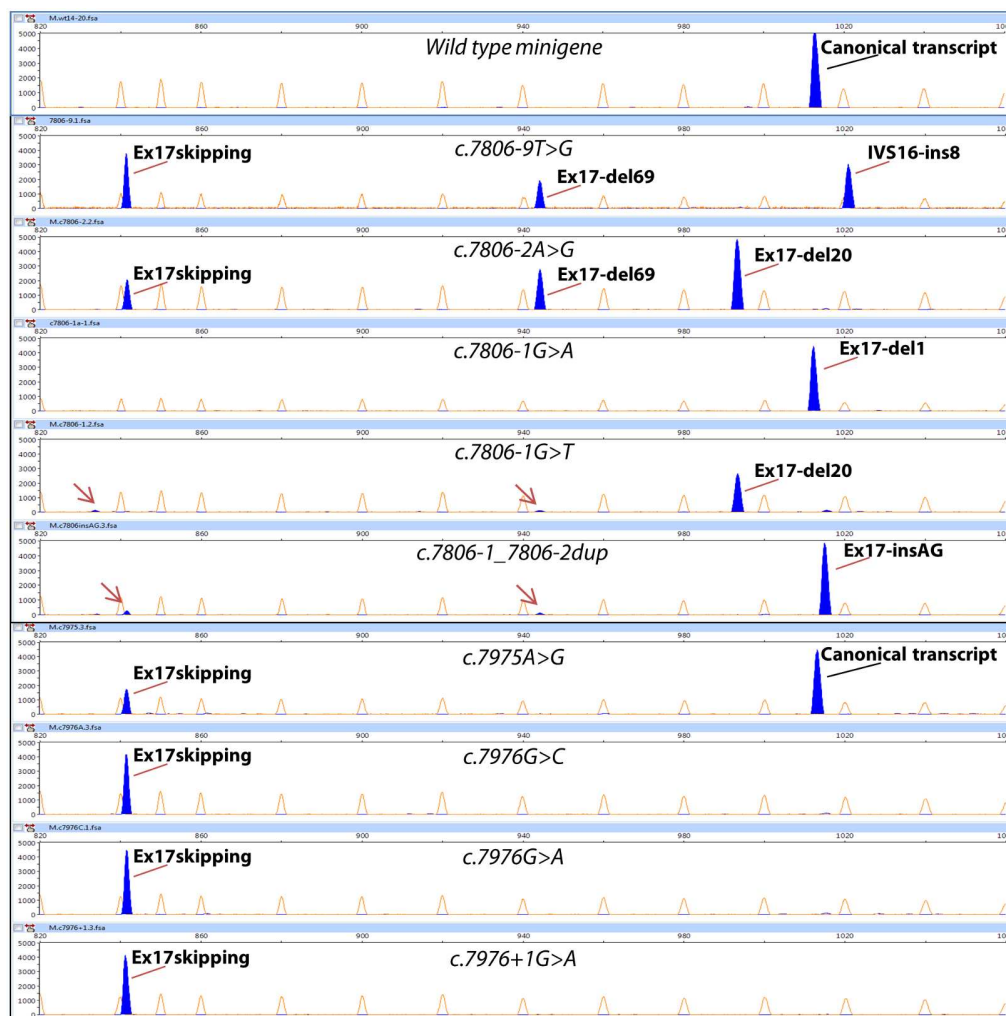
<sup>4</sup> HGVS nomenclature.

<https://doi.org/10.1371/journal.pgen.1006691.t001>

of the genetic test is ambiguous. Pathogenicity of genetic alterations is usually assumed for sequence variants that introduce PTCs, such as frameshift and nonsense mutations, or some missense changes that disrupt protein function. However, many other DNA variants can affect disease risk as other mechanisms of gene expression, such as pre-mRNA processing, are affected. In fact, many germ-line variants of human disease genes have already been associated with aberrantly spliced mRNAs that encode defective proteins [32].

In this work, we have carried out one of the most comprehensive studies of the correlation between aberrant splicing and breast cancer, whereby we evaluated the impact on splicing of 52 DNA variants of *BRCA2* exons 17 and 18 with the stable seven-exon minigene MGBR2\_ex14-20. We have shown that spliceogenic variants are relatively abundant in exons 17 and 18 (30/52 tests), representing ~15% of all reported variants of these exons and flanking intronic sequences. This rate triplicates the frequency of splice sites variants (<10 bp from exon) at the BRCA Share database (131/2818 recorded variants, 4.6%). Remarkably, these 30 variants were found in 157 different records of the BIC and BRCA Share databases. A high prevalence of spliceogenic variants has been reported in several disease genes, such as *NF1*, where up to 44% of patients carry such type of alterations [23,33]. So far we have performed lymphocyte or minigene RNA assays of 188 variants, 93 of which (49.5%) disrupted splicing [19,24,25].

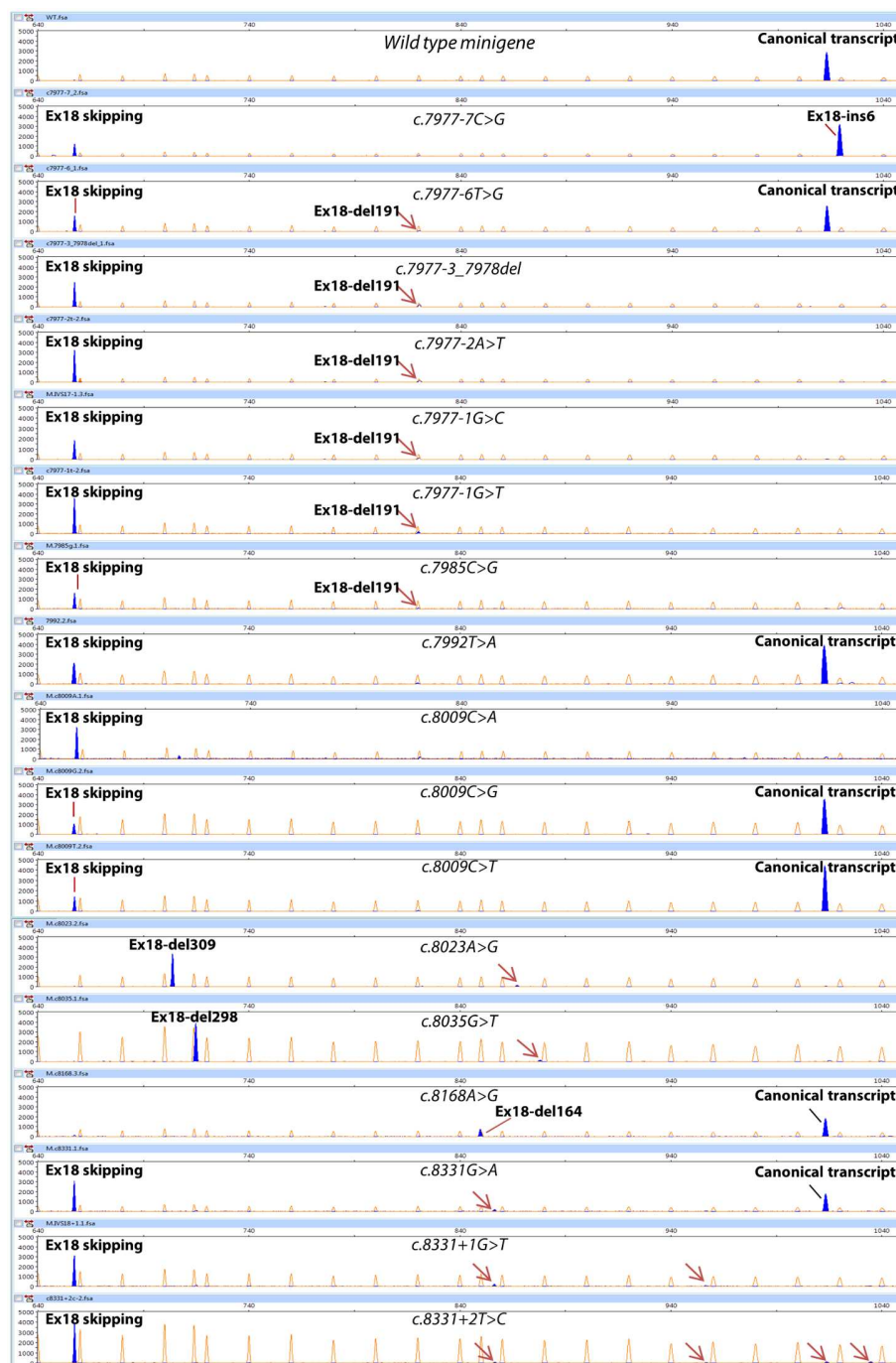
We have once more demonstrated the high capacity, robustness and simplicity of splicing reporter minigenes, with the following advantages: a) observation of a single-mutant allele effect without the interference of the wild type counterpart in patient samples; b) precise quantification of all anomalous transcripts by inhibiting the NMD; c) high capacity of this simple technical approach (Cloning-Mutagenesis-Functional Assay) [19,24,25,28,34]; d) one single minigene-construct allows the analysis of multiple variants from different exons (seven exons in this case); e) high reproducibility of physiological/pathological splicing patterns by virtue of keeping the genomic context of each exon. This was supported by comparing our results with previous studies based on patient RNA data. We found at previous reports that 8 DNA variants under study yielded, all of them, the same or similar results for patient RNA and minigene assay. Thus, DNA variants: c.7806-9T>G (exon 17 skipping+ex17-del69+ivs16-ins8) [35], c.7975A>G (partial exon 17 skipping) [36], c.7976G>A and c.7976G>C (total exon 17 skipping) [37,38], c.7988A>T (partial exon 18 skipping) [38], c.7992A>T (partial exon 18



**Fig 3. Analysis of transcripts induced by DNA variants from exon 17.** cDNA was amplified with primers RTBR2\_ex16-FW (*BRCA2* exon 16) (blue peaks) and FAM-labelled RTpSAD-RV (vector exon V2) and electrophoresed on a DNA sequencer with Genescan LIZ 1200 as size standard (orange/faint peaks). Arrows indicate minor aberrant transcripts. Screenshots of electropherograms visualized with the Peak Scanner software v1.0 are shown. Fragment sizes (bp) and relative fluorescent units are indicated on the x- and y-axes, respectively.

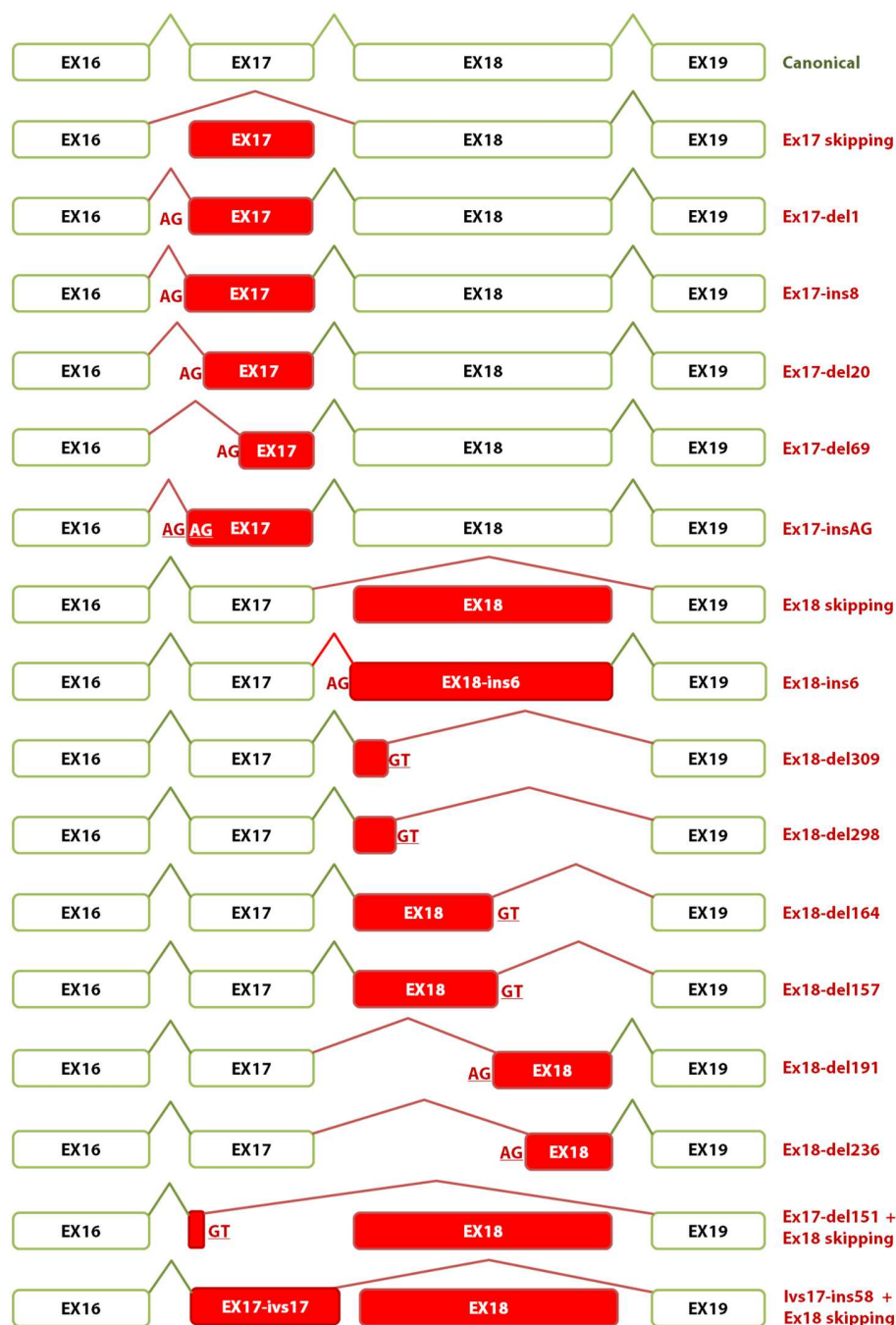
<https://doi.org/10.1371/journal.pgen.1006691.g003>

skipping) [39], c.8023A>G (ex18-del309) [21] and c.8168A>G (partial ex18-del163) [40] displayed the same splicing patterns in patient RNA and minigene MGBR2\_ex14-20, lending further support to the reproducibility of minigene results. Furthermore, a recent splicing study of 30 variants of *MLH1*, *MSH2*, *MSH6*, and *PMS2* genes (Lynch syndrome) showed that outcomes of patient RNA and minigene assays were almost identical [41]. Thus, we can conclude that the minigene strategy is sensitive and specific, so its use is suitable for the initial characterization of the splicing anomalies [21]. Nevertheless, it should be considered only as a supportive test rather than a confirmatory one since minigene results should be confirmed in patient RNA or even at protein levels whenever possible. However, it is also worth mentioning that splicing outcomes from patient RNA may be biased by technical limitations, physiological alternative transcripts, and cell-type specific differences between the patient sample (principally leukocyte RNA) and the affected tissue [42].



**Fig 4. Analysis of transcripts induced by DNA variants from exon 18.** cDNA was amplified with primers RTBR2\_ex16-FW (*BRCA2* exon 16) (blue peaks) and FAM-labelled RTpSAD-RV (exon V2 of minigene) and electrophoresed on a DNA sequencer with Genescan LIZ 1200 as size standard (orange/faint peaks). Arrows indicate minor aberrant transcripts. Screenshots of electropherograms visualized with the Peak Scanner software v1.0 are shown. Fragment sizes (bp) and relative fluorescent units are indicated on the x- and y-axes, respectively. Electropherograms of c.8007A>G, c.8072C>T and c.8249\_8250del are not represented since they show similar patterns (partial exon 18 skipping) to other DNA variants (e.g. c.8009C>G).

<https://doi.org/10.1371/journal.pgen.1006691.g004>



**Fig 5. Schematic representation of splicing events and transcripts generated by wild type and mutant minigenes.** Exons and splicing reactions exons are represented by boxes and broken lines, respectively. Anomalous/skipped exons and aberrant events are shown in red.

<https://doi.org/10.1371/journal.pgen.1006691.g005>

Finally, we have already cloned 43 out of the 50 exons of *BRCA1/2* in the following 6 constructs: *BRCA1* exons 2 to 10, and 12 to 23, *BRCA2* exons 2 to 9, 9 to 10 (<http://www.ibgm.med.uva.es/servicios/servicio-de-splicing-minigenes/>), 14 to 20 (this study) and 19 to 27 [24]. Any variant of such exons can be easily tested with a straightforward protocol in less than two weeks.

## Types of variants and splicing regulatory elements

Splicing is specifically regulated by a dense array of motifs and splicing factors so that an important message of our study is that any nucleotide change has the potential of disrupting this process. In fact, the 30 positive changes comprised 14 intronic and 16 exonic variants including 11 missense, 1 frameshift, 1 nonsense and 3 synonymous predicted changes [43]. Synonymous variations are particularly interesting since they have traditionally been considered as neutral. Many sequence variations affect disease risk, including synonymous variants that, actually, may have unexpected deleterious effects over the splicing and protein translation mechanisms [44,45]. It has been shown that these variants account for 6–8% of all driver mutations in oncogenes, where about half of them impair splicing [46]. Conversely, protein truncating variants (nonsense) are directly classified as deleterious though we have herein shown that the associated-nucleotide changes can affect splicing regulatory elements, so we could observe a “dangerous” unclassifiable splicing effect whenever they induced in-frame deletions of an exon.

With regard to the affected splicing motifs of positive variants (Table 1, S2 Table) and taking into account the bioinformatics and splicing outcomes, we can conclude that 15 variants affected the natural 5' and 3' splice sites, three the polypyrimidine tract, three created novel alternative splice sites and 9 affected ESE/ESS motifs. The NNSplice, MaxEnt and HSF algorithms accurately anticipated the splice site disruptions and the generation of novel active sites, but the splicing outcomes were absolutely unpredictable reinforcing the current need of functional assays. The characterization of the physiological alternative splicing events of *BRCA1* and *BRCA2* [14,47] and improved computer tools will help to estimate the aberrant transcripts that a particular DNA variant may generate. It is also worthy to mention that three variants of the polypyrimidine tract, c.7806-9T>G, c.7977-7C>G and c.7977-6T>G, produced defective splicing. Pyrimidine to Purine changes at this element are critical for exon recognition as we had previously described [25]. However, these modifications are barely identified by the splicing software with slight reductions of the splice site score (S2 Table). For example, NNSplice of c.7806-9T>G calculated a weak decrease of the 3' splice site score of exon 17 from 0.95 to 0.83, yet it was associated with a total splicing disruption.

*In silico* predictions of ESE/ESS motifs, which are constituted by short-degenerate sequences, showed low sensitivity. Nevertheless, there have been recently postulated two *in silico* approaches, ΔtESRseq and ΔHZ<sub>EB</sub>, that accurately detect potential ESE-variants [43]. We have found that 9 out of 28 pre-selected ESE/ESS variants affected splicing, so we have even improved its accuracy with respect to former studies by virtue of the functional mapping by microdeletions that has proven to be an exceptional method to refine ESE-variant selection. This strategy revealed the presence of operating ESEs in intervals c.7944-7973 (exon 17) and c.7979-8008 and c.8004-8033 (exon 18). Remarkably, 7 out of 9 ESE/ESS variants are placed within these intervals of exon 18 confirming the value of preliminary ESE-mapping to choose candidate variants and to fine map regulatory sequences. Interestingly, only the triple (c.7984, c.8001 and c.8003) and double (c.8001 and c.8003) mutants of SF2 sites significantly affected splicing whereas single mutants did not or did only weakly, suggesting a precise and compound control of exon 18 processing, where ESE sequences might act cooperatively. In this regard, while two possible SF2/ASF sites were bioinformatically predicted (c.7981\_7987, GATACGG, and c.8001\_c.8007, CAGAAGA) (Fig 2C), preliminary siRNAs experiments suggested the participation of the splicing factor SC35 in the regulation of exons 17 and 18 (S2 Fig), which is also involved in the regulation of a pathological GC site of the *BTK* gene [27]. Nevertheless, the definite identification of the splicing factors involved in exons 17 and 18 processing should be carried out by further siRNA and pulldown assays [48]. Independently of

the factors involved, these data allowed us to underline three small DNA segments (c.7944-7973, c.7979-7988 and c.7999-8013) where spliceogenic ESE-variants may occur. Given the poor precision of ESE/ESS-prediction software (12.2% of selected variants) [19,24], these data will provide a very valuable information for genetic counselors with a view to selecting specific exonic mutations within those intervals for splicing assays.

Donor-GC sites, such as that of exon 17, have been linked to alternative splicing [26] so that they require the control by factors that promote their efficient selection [27]. Certainly, exons 17 and 18 undergo naturally-occurring alternative splicing producing minor transcripts  $\Delta 18$  and  $\Delta 17+18$  [28], although in our study  $\Delta 18$  was only detected at even lower levels in the wild type minigene (<1%; S3 Table), together with the full-length transcript ( $\geq 99\%$ ). This may probably be due to: i) the genomic context that influences exon recognition [15]; ii) tissue-dependent alternative splicing as we used different host cells (MCF-7 vs. HeLa); and iii) RNA preparation and storage conditions, primer design, PCR conditions, and PCR product detection methodology can introduce small variations in splicing isoform ratios as previously reported [14,42].

## Clinical interpretation of DNA variants

Identification of pathogenic variants with impact on splicing will aid in breast cancer prediction, prevention and surveillance, but the clinical interpretation of the splicing outcomes of candidate variants is a particularly complex task. It is accepted that a variant would be considered likely pathogenic when it causes a majority of aberrant RNA isoforms and generates a stop codon or loss of a known functional domain. The identification of numerous anomalous transcripts of exons 17 and 18 and the production of  $\geq 2$  transcripts by many variants are proofs of this arduous undertaking.

Twelve transcripts introduced a frameshift in the open reading frame and a PTC so they inactivated *BRCA2*: ex17-del1, ex17del20, ex17ins8, ex17insAG, ex18 skipping, ex18-del191, ex18-del298, ex18-del164, ex18-del157, ex18-del236, ex17-del151+ex18 skipping and ivs17\_58-nt retention+ex18 skipping. Conversely, exon 17 skipping, ex17-del69, ex18-ins6 and ex18-del309 kept the reading frame with *a priori* unknown impact on *BRCA2* function. Exon 17 skipping and ex17-del69 led to deletions of 57 and 23 amino acids, respectively, at the essential  $\alpha$ -helical domain of the *BRCA2* protein (amino acids 2479 to 2667). This domain facilitates *BRCA2* binding to single-stranded and double-stranded DNA [49]. Moreover, 30 out of the 57 residues encoded by exon 17 are strictly conserved from sea urchin to human revealing its importance for *BRCA2* activity (IARC *BRCA2* alignment; [http://agvgd.iarc.fr/BRCA2\\_Align.htm](http://agvgd.iarc.fr/BRCA2_Align.htm)). Likewise, it has been shown that the loss of exon 17 inactivates *BRCA2* function [37]. Furthermore, exon 17 variant c.7976G>A, which is associated with total exon 17 skipping, reached odds of causality of >3,000:1 [38]. Consequently, we can infer that the rest of the variants with ex17 skipping as the unique transcript, such as c.7976G>C and c.7976+1G>A, are also likely pathogenic. Moreover, variants c.7806-9T>G and c.7806-2A>G with at least 3 abnormal transcripts, including ex17 skipping, could also be considered as likely pathogenic, given that the other transcripts disrupt the reading frame (Ivs16ins8 or ex17-del20) or leads to in-frame loss of 23 aminoacids (ex17-del69), 13 of which are strictly conserved. The RNA isoform ex18-ins6 would insert new amino acids Ser-Phe between Tyr2658 and Arg2659. Precisely, amino acids from Val2652 to Asp2661 are conserved from sea urchin, and two missense changes at this protein segment, p.Leu2653Pro and p.Arg2659Lys were formerly classified as deleterious [38,50]. Consequently, transcript ex18-ins6 might have a deleterious impact on *BRCA2* function but it is required further protein function studies. Finally, abnormal transcript ex18-del309 was predicted to cause an in-frame deletion of 103 amino acids between codons Ile2675 and Lys2777 of the OB1 (oligonucleotide ssDNA-binding fold) motif at the DNA binding domain of

*BRCA2*, 24 of which are strictly conserved from sea urchin. Variant c.8023A>G, which induced ex18del309, had previously been classified as pathogenic [21] (BIC and UMD databases), so this transcript disrupts *BRCA2* function. Also, c.8331G>A might be an important risk allele as abnormal transcripts almost reach 60%, which is the suggested threshold for severe splicing aberrations [51]. According to the guidelines of the American College of Medical Genetics and Genomics (ACMG), [52] 12 spliceogenic variants were classified as pathogenic (c.7806-2A>G, c.7806-1G>A, c.7806-1G>T, c.7806-1\_7806-2dup, c.7976+1G>A, c.7977-3\_7978del, c.7977-2A>T, c.7977-1G>T, c.7977-1G>C, c.8009C>A, c.8331+1G>T and c.8331+2T>C) and 8 as likely pathogenic (c.7806-9T>G, c.7976G>C, c.7976G>A, c.7977-7C>G, c.7985C>G, c.8023A>G, c.8035G>T and c.8331G>A), under the splicing viewpoint (S4 Table). Remarkably, all of them account for 72 independent records at the mutation databases (S4 Table) and nine of them had been classified as VUS. Reclassification of VUS as deleterious will notably increase the number of HBOC families who may benefit from tailored preventive and prophylactic measures as well as new targeted therapies, such as Poly (ADP-ribose) polymerase (PARP)-inhibitors, for patients with *BRCA1/2* associated cancers [53]. It is also worthy to mention that causal and likely causal splicing variants account for a remarkable 30.8% (20/65) of all predicted pathological variants of exons 17 and 18 at the *BRCA2* Share database (S5 Table), representing the second more frequent type of causal variants after frameshift mutations (44.6%).

On the other hand, two variants with weaker splicing alterations, c.8168A>G/ p.Asp2723Gly (30%) and c.8249\_8250del (7%), were previously classified as likely pathogenic (protein function and truncation, respectively) [38], so their pathogenicity may probably be due to a double mechanism: protein inactivation and splicing disruption, like *BRCA1* c.5123C>A (p.A1708E) [25,54]. Likewise, c.8009C>A was previously classified as causal because of its predicted non-sense change (p.Ser2670X), but it actually induces 96% of aberrant transcripts so it should be reclassified as a spliceogenic variant. Reclassification of missense and protein truncation variants as splicing alterations might also have an effect in their penetrance and expressivity. Taken together, 8 spliceogenic variants remain classified as VUS since relevant proportions of the full-length transcript were detected (c.7975A>G, c.7977-6T>G, c.7988A>T, c.7992T>A, c.8007A>G, c.8009C>T, c.8009C>G and c.8072C>T) (Table 1; S3 and S4 Tables). It is complex to interpret the role of variants with partial splicing anomalies in HBOC under the clinical perspective as they will require more studies to elucidate it. Nevertheless, we can speculate that they represent low BC risk alleles that might interact with other susceptibility and protector alleles to modify the overall BC risk. The incorporation of all these data into a single integrated model of BC risk would improve disease prediction and prevention.

In conclusion, dysregulation of splicing should be considered as a primary mechanism of gene inactivation to be investigated in human disease genes. Spliceogenic variants are comparatively abundant in *BRCA2* exons 17 and 18 because recognition of both exons additionally requires the regulation of specific ESE motifs in exons 17 and 18, whose abolitions drive splicing aberrations. Furthermore, the pSAD-based minigenes are useful tools for molecular diagnostics and genetic counseling of hereditary breast/ovarian cancer or other genetic disorders as well as for the basic research on the splicing process. Hence, RNA assays supply essential information for the clinical interpretation of variants that should be incorporated in the genetic counselling of human hereditary diseases.

## Materials and methods

### Databases and bioinformatics analyses

*BRCA2* variants of breast/ovarian cancer patients were available from the BIC (<https://research.nhgri.nih.gov/projects/bic/Member/index.shtml>) and the BRCA Share databases (last

accessed date 2016/04/01; <http://www.umd.be/BRCA2/>) [55]. Variants of intron 17 were collected from the 1000 Genomes database ([http://browser.1000genomes.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000139618;r=13:32889611-32973805;t=ENST00000380152](http://browser.1000genomes.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000139618;r=13:32889611-32973805;t=ENST00000380152)). Variant descriptions were according to the *BRCA2* GenBank sequence NM000059.1 and the guidelines of the Human Genome Variation Society (HGVS; <http://www.hgvs.org/mutnomen/>).

Mutant and wild type (wt) sequences were analyzed with NNSPLICE ([http://www.fruitfly.org/seq\\_tools/splice.html](http://www.fruitfly.org/seq_tools/splice.html)) [56], and Human Splicing Finder version 3.0 (HSF; <http://www.umd.be/HSF3/>) [57], which includes algorithms for splice sites, silencers and enhancers [31,58–62].

## Minigene construction

MGBR2\_ex14–20 was assembled in four steps by overlapping extension PCR or classical restriction digestion/ligation cloning with three intermediate constructs: MGBR2EX17–18, MGBR2EX16–18, and MGBR2EX16–20. All the inserts were amplified with Phusion High Fidelity polymerase (Thermo Fisher Scientific, Waltham, MA, USA) and primers indicated on S1 Table. Exons 17–18 were subcloned into the pSAD vector by overlapping extension PCR. Then, exon 16 was added by the same technique. Exons 19–20 were inserted between the XhoI and BamHI restriction sites of the 16–18 construct. Finally, exons 14–15 were introduced using the EagI and SacI restriction sites. All clones were functionally checked in MCF-7 cells.

## Site-directed mutagenesis

DNA variants were introduced with the QuikChange Lightning kit (Agilent, Santa Clara, CA). The wt minigene MGBR2\_ex14–20 was used as template to generate 52 BIC/BRCA Share DNA variants as well as seventeen exonic (17 and 18) and four intronic (ivs17) microdeletions (S2 Table). The first two and the last three nucleotides of each exon were always preserved to avoid any disruptions of the canonical acceptor and donor sites, respectively. Deletions were introduced by PCR-mutagenesis with chimeric 50–60mer primers containing 25–30 nucleotides of each end of the deletion.

## Transfection of eukaryotic cells

Approximately  $2 \times 10^5$  MCF7 cells were grown to 90% confluency in 0.5 mL of medium (MEME, 10% Fetal Bovine Serum, 2 mM glutamine, 1% Non-essential amino acids and 1% Penicillin/Streptomycin) in 4-well plates (Nunc, Roskilde, Denmark). Cells were transiently transfected with 1 µg of each minigene and 2 µL of Lipofectamine 2000 or low toxicity Lipofectamine (Life Technologies, Carlsbad, CA). To inhibit nonsense mediated decay (NMD), cells were incubated with cycloheximide (Sigma-Aldrich, St. Louis, MO) 300 µg/mL for 4 hours. RNA was purified with the Genematrix Universal RNA Purification Kit (EURx, Gdansk, Poland) with on-column DNase I digestion to degrade genomic DNA that could interfere in RT-PCR.

## RT-PCR of minigenes

Retrotranscription was carried out with 400 ng of RNA and RevertAid H Minus First Strand cDNA Synthesis Kit (Life Technologies), using gene specific primer RTPSPL3-RV (5'TGAG GAGTGAATTGGTCGAA 3'). Samples were incubated at 42°C for 1 hour, and reactions were inactivated at 70°C for 5 min. Then, 1–2 µL of the resultant cDNA were amplified with SD6-PSPL3\_RTFW (5'-TCACCTGGACAACCTCAAAG-3') or RTBR2\_ex16FW

(5'-TATGGACTGGAAAAGGAATAC-3') and RTpSAD-RV (Patent P201231427, CSIC) (sizes: 1012 and 1806 bp, respectively) using Platinum Taq DNA polymerase (Life Technologies). Samples were denatured at 94°C for 2 min, followed by 35 cycles consisting of 94°C for 30 sec, 59°C for 30 sec, and 72°C (1 min/kb), and a final extension step at 72°C for 5 min. Sequencing reactions were performed either using the kit BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) following the manufacturer's instructions or by the sequencing facility of Macrogen Europe (Amsterdam, The Netherlands). All transcripts from exon 18 microdeletion c.8004\_8013del were subcloned into the pJet1.2 PCR cloning vector (Thermo Fisher Scientific) and sequenced.

In order to quantify all transcripts, semi-quantitative fluorescent PCRs were undertaken at least in triplicate (>234 assays were performed of 52 natural plus 26 artificial mutants; [S2 Table](#)) with primers RTBR2\_ex16-FW and FAM-RTpSAD-RV and Platinum Taq DNA polymerase (Life Technologies) under standard conditions except that 26 cycles were herein applied [[19,63](#)]. FAM-labeled products were run with LIZ-1200 Size Standard at the Macrogen facility and analyzed with the Peak Scanner software V1.0. Only peaks with heights  $\geq 50$  RFU (Relative Fluorescence Units) were taken into account. Peak areas were used to quantify the relative abundance of each transcript that was the average of at least three experiments [[19](#)].

## Supporting information

**S1 Table. Primers used for the construction of the MGBR2\_ex14-20 minigene.**  
(PDF)

**S2 Table. Bioinformatics analysis and mutagenesis primers of selected reported and artificial DNA variants of exons 17 and 18.**  
(DOCX)

**S3 Table. Relative quantification of transcripts induced by DNA variants and microdeletions of *BRCA2* exons 17 and 18 in MCF-7 cells.**  
(PDF)

**S4 Table. Analysis of the 52 assayed variants according to the guidelines of the American College of Medical Genetics and Genomics (ACMG).**  
(PDF)

**S5 Table. Distribution of pathogenic/likely pathogenic variants by type according to the BRCA Share database and this study.**  
(PDF)

**S1 Fig. Capillary electrophoresis of the [FAM]-RT-PCR products generated by the mutant c.7976+2C>T (canonical donor GT site, above) and the wild type (below) minigenes.** The blue peaks denote the full-length transcripts induced by both minigenes and the Genescan 1200 Size Standard (Applied Biosystems) is shown as orange peaks.  
(TIF)

**S2 Fig. Impact of RNA interference-mediated depletion of splicing factors SC35, SF2/ASF and Tra2 $\beta$  on splicing of minigene MGBR2\_ex14-20.** About  $1.5 \times 10^5$  MCF7 cells were subjected to a two-hit transfection with Oligofectamine (Thermo Fisher Scientific) and custom-made small interfering RNA (siRNA) at a final concentration of 0.08  $\mu$ M on day 2, the minigene was transfected with Lipofectamine (Thermo Fisher Scientific) on day 4, and RNA was extracted on day 5. The following siRNAs were used: SC35 (aaccaggucgcgaucgaa), SF2/ASF (acgauugccgcaucacgu), Tra2 $\beta$  (ggaucucgugaaguauuu) and a control luciferase siRNA.

RT-PCR products were run on an agarose gel (1.0%) and stained with Ethidium Bromide. Several aberrant transcripts are visualized, some of which are indicated by arrows.

(TIF)

**S3 Fig. High resolution of capillary electrophoresis of fluorescent RT-PCRs.** Electropherograms of the wild type minigene and mutations c.7806-1G>A and c.7806-1\_7806-2dup were overlaid. Transcripts ex17-del1 and ex17-insAG differed in size by only 1 and 2 nucleotides, respectively, from the canonical transcript despite the large size (1,012 nt) of the RT-PCR products. The Liz-1200 size standard is shown as orange peaks.

(TIF)

## Acknowledgments

We are grateful to Alberto Valenzuela for his technical support and helpful comments and to Susana Gómez-Barrero for the critical reading of the manuscript.

## Author Contributions

**Conceptualization:** EAV EFB.

**Formal analysis:** EFB BDG VVZ EAV.

**Funding acquisition:** EAV.

**Investigation:** EFB EAV BDG DJS AA VVZ.

**Methodology:** EFB BDG DJS AA VVZ EAV.

**Resources:** EAV.

**Supervision:** EAV.

**Visualization:** BDG EAV.

**Writing – original draft:** EAV EFB.

**Writing – review & editing:** EAV EFB BDG DJS AA VVZ.

## References

1. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science*. 1994; 266: 66–71. PMID: [7545954](https://pubmed.ncbi.nlm.nih.gov/7545954/)
2. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the Breast-Cancer Susceptibility Gene *Brca2*. *Nature*. 1995; 378: 789–792. <https://doi.org/10.1038/378789a0> PMID: [8524414](https://pubmed.ncbi.nlm.nih.gov/8524414/)
3. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet*. 2008; 40: 17–22. <https://doi.org/10.1038/ng.2007.53> PMID: [18163131](https://pubmed.ncbi.nlm.nih.gov/18163131/)
4. Nielsen FC, van Overeem Hansen T, Sørensen CS. Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat Rev Cancer*. 2016; 16: 599–612. <https://doi.org/10.1038/nrc.2016.72> PMID: [27515922](https://pubmed.ncbi.nlm.nih.gov/27515922/)
5. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013; 45: 353–61, 361–2. <https://doi.org/10.1038/ng.2563> PMID: [23535729](https://pubmed.ncbi.nlm.nih.gov/23535729/)
6. Osorio A, Milne RL, Kuchenbaecker K, Václavá T, Pita G, Alonso R, et al. DNA glycosylases involved in base excision repair may be associated with cancer risk in *BRCA1* and *BRCA2* mutation carriers. *PLoS Genet*. 2014; 10: e1004256. <https://doi.org/10.1371/journal.pgen.1004256> PMID: [24698998](https://pubmed.ncbi.nlm.nih.gov/24698998/)

7. Eccles EB, Mitchell G, Monteiro ANA, Schmutzler R, Couch FJ, Spurdle AB, et al. BRCA1 and BRCA2 genetic testing-pitfalls and recommendations for managing variants of uncertain clinical significance. *Ann Oncol*. 2015; 26: 2057–2065. <https://doi.org/10.1093/annonc/mdv278> PMID: 26153499
8. Radice P, De Summa S, Caleca L, Tommasi S. Unclassified variants in BRCA genes: guidelines for interpretation. *Ann Oncol*. 2011; 22 Suppl 1: i18–23.
9. De Vooght KMK, Van Wijk R, Van Solinge WW. Management of gene promoter mutations in molecular diagnostics. *Clin Chem*. 2009; 55: 698–708. <https://doi.org/10.1373/clinchem.2008.120931> PMID: 19246615
10. Wang G-S, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet*. 2007; 8: 749–61. <https://doi.org/10.1038/nrg2164> PMID: 17726481
11. Brewster BL, Rossiello F, French JD, Edwards SL, Wong M, Wronski A, et al. Identification of fifteen novel germline variants in the BRCA1 3'UTR reveals a variant in a breast cancer case that introduces a functional miR-103 target site. *Hum Mutat*. 2012; 33: 1665–75. <https://doi.org/10.1002/humu.22159> PMID: 22753153
12. Hindorf LA, Sethupathy P, Jenkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009; 106: 9362–7. <https://doi.org/10.1073/pnas.0903103106> PMID: 19474294
13. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456: 470–6. <https://doi.org/10.1038/nature07509> PMID: 18978772
14. Fackenthal JD, Yoshimatsu T, Zhang B, de Garibay GR, Colombo M, De Vecchi G, et al. Naturally occurring BRCA2 alternative mRNA splicing events in clinically relevant samples. *J Med Genet*. 2016; 53: 548–558. <https://doi.org/10.1136/jmedgenet-2015-103570> PMID: 27060066
15. Buratti E, Baralle M, Baralle FE. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res*. 2006; 34: 3494–510. <https://doi.org/10.1093/nar/gkl498> PMID: 16855287
16. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet*. 1992; 90: 41–54. PMID: 1427786
17. López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett*. 2005; 579: 1900–3. <https://doi.org/10.1016/j.febslet.2005.02.047> PMID: 15792793
18. Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene*. 2016; 35: 2413–27. <https://doi.org/10.1038/onc.2015.318> PMID: 26300000
19. Acedo A, Sanz DJ, Durán M, Infante M, Pérez-Cabornero L, Miner C, et al. Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res*. 2012; 14: R87. <https://doi.org/10.1186/bcr3202> PMID: 22632462
20. Baralle D, Lucassen A, Buratti E. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep*. 2009; 10: 810–816. <https://doi.org/10.1038/embor.2009.170> PMID: 19648957
21. Bonnet C, Krieger S, Vezain M, Rousselin A, Tournier I, Martins A, et al. Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *J Med Genet*. 2008; 45: 438–446. <https://doi.org/10.1136/jmg.2007.056895> PMID: 18424508
22. Byers H, Wallis Y, van Veen EM, Lalloo F, Reay K, Smith P, et al. Sensitivity of BRCA1/2 testing in high-risk breast/ovarian/male breast cancer families: little contribution of comprehensive RNA/NGS panel testing. *Eur J Hum Genet*. 2016; 24: 1591–1597. <https://doi.org/10.1038/ejhg.2016.57> PMID: 27273131
23. Evans DG, Bowers N, Burkitt-Wright E, Miles E, Garg S, Scott-Kitching V, et al. Comprehensive RNA Analysis of the NF1 Gene in Classically Affected NF1 Affected Individuals Meeting NIH Criteria has High Sensitivity and Mutation Negative Testing is Reassuring in Isolated Cases With Pigmentary Features Only. *EBioMedicine*. 2016; 7: 212–20. <https://doi.org/10.1016/j.ebiom.2016.04.005> PMID: 27322474
24. Acedo A, Hernández-Moro C, Curiel-García Á, Díez-Gómez B, Velasco EA. Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat*. 2015; 36: 210–21. <https://doi.org/10.1002/humu.22725> PMID: 25382762
25. Sanz DJ, Acedo A, Infante M, Durán M, Pérez-Cabornero L, Esteban-Cardena E, et al. A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin Cancer Res*. 2010; 16: 1957–67. <https://doi.org/10.1158/1078-0432.CCR-09-2564> PMID: 20215541

26. Churbanov A, Winters-Hilt S, Koonin E V, Rogozin IB. Accumulation of GC donor splice signals in mammals. *Biol Direct*. 2008; 3: 30. <https://doi.org/10.1186/1745-6150-3-30> PMID: 18613975
27. Kralovicova J, Hwang G, Asplund a. C, Churbanov A, Smith CIE, Vorechovsky I. Compensatory signals associated with the activation of human GC 5' splice sites. *Nucleic Acids Res*. 2011; 39: 7077–7091. <https://doi.org/10.1093/nar/gkr306> PMID: 21609956
28. Ruiz de Garibay G, Acedo A, García-Casado Z, Gutiérrez-Enríquez S, Tosar A, Romero A, et al. Capillary electrophoresis analysis of conventional splicing assays: IARC analytical and clinical classification of 31 *BRCA2* genetic variants. *Hum Mutat*. 2014; 35: 53–7. <https://doi.org/10.1002/humu.22456> PMID: 24123850
29. Graveley BR. Sorting out the complexity of SR protein functions. *RNA*. 2000; 6: 1197–1211. PMID: 10999598
30. Fairbrother WG, Holste D, Burge CB, Sharp PA. Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol*. 2004; 2: E268. <https://doi.org/10.1371/journal.pbio.0020268> PMID: 15340491
31. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*. 2003; 31: 3568–71. PMID: 12824367
32. Singh RK, Cooper TA. Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med*. 2012; 18: 472–82. <https://doi.org/10.1016/j.molmed.2012.06.006> PMID: 22819011
33. Pros E, Gómez C, Martín T, Fábregas P, Serra E, Lázaro C. Nature and mRNA effect of 282 different NF1 point mutations: focus on splicing alterations. *Hum Mutat*. 2008; 29: 173–193.
34. Lara B, Martínez MT, Blanco I, Hernández-Moro C, Velasco E a, Ferrarotti I, et al. Severe alpha-1 antitrypsin deficiency in composite heterozygotes inheriting a new splicing mutation QOMadrid. *Respir Res*. 2014; 15: 125. <https://doi.org/10.1186/s12931-014-0125-y> PMID: 25287719
35. Kwong A, Wong LP, Chan KYK, Ma ESK, Khoo US, Ford JM. Characterization of the pathogenic mechanism of a novel *BRCA2* variant in a Chinese family. *Fam Cancer*. 2008; 7: 125–133. <https://doi.org/10.1007/s10689-007-9155-7> PMID: 17657584
36. Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on *BRCA1* and *BRCA2* variants. *Hum Mutat*. 2012; 33: 1228–38. <https://doi.org/10.1002/humu.22101> PMID: 22505045
37. Wu K, Hinson SR, Ohashi A, Farrugia D, Wendt P, Tavtigian S V, et al. Functional evaluation and cancer risk assessment of *BRCA2* unclassified variants. *Cancer Res*. 2005; 65: 417–26. PMID: 15695382
38. Farrugia DJ, Agarwal MK, Pankratz VS, Deffenbaugh AM, Pruss D, Frye C, et al. Functional assays for classification of *BRCA2* variants of uncertain significance. *Cancer Res*. 2008; 68: 3523–31. <https://doi.org/10.1158/0008-5472.CAN-07-1587> PMID: 18451181
39. Théry JC, Krieger S, Gaildrat P, Révillion F, Buisine M-P, Killian A, et al. Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the *BRCA* genes. *Eur J Hum Genet*. 2011; 19: 1052–8. <https://doi.org/10.1038/ejhg.2011.100> PMID: 21673748
40. Walker LC, Whiley PJ, Couch FJ, Farrugia DJ, Healey S, Eccles DM, et al. Detection of splicing aberrations caused by *BRCA1* and *BRCA2* sequence variants encoding missense substitutions: implications for prediction of pathogenicity. *Hum Mutat*. 2010; 31: E1484–505. <https://doi.org/10.1002/humu.21267> PMID: 20513136
41. van der Klift HM, Jansen AML, van der Steenstraten N, Bik EC, Tops CMJ, Devilee P, et al. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol Genet Genomic Med*. 2015; 3: 327–345. <https://doi.org/10.1002/mgg3.145> PMID: 26247049
42. Whiley PJ, de la Hoya M, Thomassen M, Becker A, Brandão R, Pedersen IS, et al. Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. *Clin Chem*. 2014; 60: 341–52. <https://doi.org/10.1373/clinchem.2013.210658> PMID: 24212087
43. Soukariéh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, et al. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet*. 2016; 12: e1005756. <https://doi.org/10.1371/journal.pgen.1005756> PMID: 26761715
44. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet*. 2011; 12: 683–691. <https://doi.org/10.1038/nrg3051> PMID: 21878961
45. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet*. 2002; 3: 285–98. <https://doi.org/10.1038/nrg775> PMID: 11967553

46. Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell*. 2014; 156: 1324–35. <https://doi.org/10.1016/j.cell.2014.01.051> PMID: 24630730
47. Colombo M, Blok MJ, Whiley P, Santamariña M, Gutiérrez-Enríquez S, Romero A, et al. Comprehensive annotation of splice junctions supports pervasive alternative splicing at the *BRCA1* locus: a report from the ENIGMA consortium. *Hum Mol Genet*. 2014; 23: 3666–80. <https://doi.org/10.1093/hmg/ddu075> PMID: 24569164
48. Goina E, Skoko N, Pagani F. Binding of DAZAP1 and hnRNP A1/A2 to an exonic splicing silencer in a natural *BRCA1* exon 18 mutant. *Mol Cell Biol*. 2008; 28: 3850–60. <https://doi.org/10.1128/MCB.02253-07> PMID: 18391021
49. Roy R, Chun J, Powell SN. *BRCA1* and *BRCA2*: different roles in a common pathway of genome protection. *Nat Rev Cancer*. 2012; 12: 68–78.
50. Guidugli L, Pankratz VS, Singh N, Thompson J, Erding C a., Engel C, et al. A classification model for *BRCA2* DNA binding domain missense variants based on homology-directed repair activity. *Cancer Res*. 2013; 73: 265–275. <https://doi.org/10.1158/0008-5472.CAN-12-2081> PMID: 23108138
51. Vallée MP, Di Sera TL, Nix DA, Paquette AM, Parsons MT, Bell R, et al. Adding In Silico Assessment of Potential Splice Aberration to the Integrated Evaluation of *BRCA* Gene Unclassified Variants. *Hum Mutat*. 2016; 37: 627–639. <https://doi.org/10.1002/humu.22973> PMID: 26913838
52. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. IOP Publishing; 2015; 17: 405–423.
53. Balmaña J, Domchek SM, Tutt A, Garber JE. Stumbling blocks on the path to personalized medicine in breast cancer: the case of PARP inhibitors for *BRCA1/2*-associated cancers. *Cancer Discov*. 2011; 1: 29–34. <https://doi.org/10.1158/2159-8274.CD-11-0048> PMID: 22586318
54. Vallon-Christersson J, Cayan C, Haraldsson K, Loman N, Bergthorsson JT, Brøndum-Nielsen K, et al. Functional analysis of *BRCA1* C-terminal missense mutations identified in breast and ovarian cancer families. *Hum Mol Genet*. 2001; 10: 353–60. PMID: 11157798
55. Caputo S, Benboudjema L, Sinielnikova O, Rouleau E, Bérout C, Lidereau R, et al. Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-*BRCA1/BRCA2* databases. *Nucleic Acids Res*. 2012; 40: 992–1002.
56. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997; 4: 311–23. <https://doi.org/10.1089/cmb.1997.4.311> PMID: 9278062
57. Desmet F-O, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009; 37: e67. <https://doi.org/10.1093/nar/gkp215> PMID: 19339519
58. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004; 11: 377–94. <https://doi.org/10.1089/1066527041410418> PMID: 15285897
59. Fairbrother WG, Yeh R-F, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. *Science*. 2002; 297: 1007–13. <https://doi.org/10.1126/science.1073774> PMID: 12114529
60. Zhang XH-F, Chasin LA. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev*. 2004; 18: 1241–50. <https://doi.org/10.1101/gad.1195304> PMID: 15145827
61. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. *Cell*. 2004; 119: 831–45. <https://doi.org/10.1016/j.cell.2004.11.010> PMID: 15607979
62. Sironi M, Menozzi G, Riva L, Cagliani R, Comi GP, Bresolin N, et al. Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res*. 2004; 32: 1783–91. <https://doi.org/10.1093/nar/gkh341> PMID: 15034146
63. Dallman M.J.; Porter A.C.G. Semi-quantitative PCR for the analysis of gene expression (Chapter). “PCR, a practical approach.” McPherson MJ, Taylor GR, Quirke P, editors. IRL Press at Oxford University Press; 1991.



## Material suplementario

---

**S1 Table.** Primers used for the construction of the MGBR2\_ex14-20 minigene.

**S2 Table.** Bioinformatics analysis and mutagenesis primers of selected reported and artificial DNA variants of exons 17 and 18.

**S3 Table.** Relative quantification of transcripts induced by DNA variants and microdeletions of BRCA2 exons 17 and 18 in MCF-7 cells.

**S4 Table.** Analysis of the 52 assayed variants according to the guidelines of the American College of Medical Genetics and Genomics (ACMG).

**S5 Table.** Distribution of pathogenic/likely pathogenic variants by type according to the BRCA Share database and this study.

**S1 Fig.** Capillary electrophoresis of the [FAM]-RT-PCR products generated by the mutant c.7976+2C>T (canonical donor GT site, above) and the wild type (below) minigenes. The blue peaks denote the full-length transcripts induced by both minigenes and the Genescan 1200 Size Standard (Applied Biosystems) is shown as orange peaks.

**S2 Fig.** Impact of RNA interference-mediated depletion of splicing factors SC35, SF2/ASF and Tra2  $\beta$  on splicing of minigene MGBR2\_ex14-20. About  $1.5 \times 10^5$  MCF7 cells were subjected to a two-hit transfection with Oligofectamine (Thermo Fisher Scientific) and custom-made small interfering RNA (siRNA) at a final concentration of  $0.08 \mu\text{M}$  on day 2, the minigene was transfected with Lipofectamine (Thermo Fisher Scientific) on day 4, and RNA was extracted on day 5. The following siRNAs were used: SC35 (aauccaggucgcgaucgaa), SF2/ASF (acgauugccgcaucuacgu), Tra2  $\beta$  (ggauucuugugaaguauuu) and a control luciferase siRNA. RT-PCR products were run on an agarose gel (1.0%) and stained with Ethidium Bromide. Several aberrant transcripts are visualized, some of which are indicated by arrows.

**S3 Fig.** High resolution of capillary electrophoresis of fluorescent RT-PCRs. Electropherograms of the wild type minigene and mutations c.7806-1G>A and c.7806-1\_7806-2dup were overlaid. Transcripts ex17-del1 and ex17-insAG differed in size by only 1 and 2 nucleotides, respectively, from the canonical transcript despite the large size (1,012 nt) of the RT-PCR products. The Liz-1200 size standard is shown as orange peaks.

**S1 Table.** Primers used for the construction of the MGBR2\_ex14-20 minigene.

| Primer                | 5'-3' Sequence                                    | Size |
|-----------------------|---|------|
| MGBR2_ex14-15-SacIFW  | CACACAGAGCTCTGAAGATGTGAAGGTGAGAGAA                | 2405 |
| MGBR2_ex14-15-EagIRV  | CACACACGGCCGGTTCAGGGCTATCAGTTATTCA                |      |
| MGBR2_ex17-18_Ins-FW  | GCTCTAGAACTAGTGGATCCCCCGGTCAGTATGATACTTTGATACATGT | 1556 |
| MGBR2_ex17-18_Ins-RW  | ATAAGCTTGATATCGAATTCCTGCAGTGGAAATAGGGATCTGATCAA   |      |
| MGBR2_ex16_Ins-FW     | GCTCTAGAACTAGTGGATCCCCCGGCCCTTTTGTTCCTCATCTAAGT   | 805  |
| MGBR2_ex16_Ins-RW     | GACATGTATCAAAGTATCATACTGAATAAATGCCTAAGAAAAATGT    |      |
| MGBR2_ex19-20_XhoI-FW | CACACACTCGAGATAGCATTAAAGAACTTGTAGCA               | 1188 |
| MGBR2_ex19-20-KpnI-RV | CACACAGGTACCATTACAAATGGCTTAGACCTGA                |      |

**S2 Table.** Bioinformatics analysis and mutagenesis primers of selected reported and artificial DNA variants of exons 17 and 18.

| DNA variant <sup>1</sup>         | Type                  | Previous Classification | Bioinformatics <sup>2,3</sup>  |  |  | Primers for Site-directed Mutagenesis (5'- 3')  |
|----------------------------------|-----------------------|-------------------------|--|--|--|---|
|                                  |                       |                         | Acceptor/Donor   | ESEs   | ESSs   |   |
| <b>EXON 17</b>                   |                       |                         |  |  |  |   |
| c.7806-40A>G                     | Intronic              | VUS                     | HSF: [-]branch site 45 nt upstream (67.63→38);   | [+] SRp55 (74.82); [+] PESE  | [-] 2 Sironi ESS; [+] Fas-ESS; [-] hnRNPA1 (66.19)           | GTGGAATTCAGTATCATCCTGTGTGGTTTTATGATAATAT<br>ATATTATCATAAAAACCACACAGGATGATACTGAATTCAAC   |
| c.7806-9T>G                      | Intronic              | VUS                     | NNS: ↓3' ss (0.95→0.83, [+] 3' ss=0.69.<br>MaxEnt: 3' ss:8.33→1.73; [+] 3' ss=6.01                     | -  | -  | ATGATAATATTCTACTTTTAGTTGTTTCAGGGCTCTGTGTGA<br>TCACACAGAGCCCTGAACAACTAAAAGTAGAATATTATCAT |
| c.7806-2A>G                      | Intronic              | Causal                  | [-] 3' ss (NNS: 0.95 →<0.4)  | [-] 9G8 (59.87)  | [+] 3 Sironi ESS; [-] hnRNPA1 (82.86)                        | TCTACTTTTATTTGTTTCGGGGCTCTGTGTGACACT<br>AGTGTCACACAGAGCCCCGAACAAATAAAAGTAGA             |
| c.7806-1G>A                      | Intronic              | Causal                  | [-] 3' SS (MaxEnt: 8.33 → -0.41; NNS: 0.95→<0.4);<br>[+] 3' SS (MaxEnt: -2.51 → 5.54) (1nt downstream) | [-] SC35 (81.13)   | [+] hnRNPA1 (73,81)  | ATTCTACTTTTATTTGTTCAAGGCTCTGTGTGACACTCCAG<br>CTGGAGTGTACACAGAGCCTTGAACAAATAAAAGTAGAAT   |
| c.7806-1G>T                      | Intronic              | Causal                  | [-] 3' SS (NNS: 0.95 →<0.4);   | [-] SC35 (81.13)   | [+] 2 PESS; [-] hnRNPA1 (82,86)                              | CTACTTTTATTTGTTTCATGGCTCTGTGTGACACTC<br>GAGTGTACACAGAGCCATGAACAAATAAAAGTAG              |
| c.7806-1_7806-2dup (c.7806insAG) | Intronic              | Causal (frameshift)     | MaxEnt: 8.33→6.35 (2 nt downstream)<br>NNS (=)   | [+] SC35 (77.75); [+] 2 SF2 (89.92; 90.74); [+] SRp40 (89.70)                                  | [+] Sironi ESS; [+] hnRNPA1 (86.67)                          | TCTACTTTTATTTGTTTCAGAGGGCTCTGTGTGACACTCC<br>GGAGTGTACACAGAGCCCTCTGAACAAATAAAAGTAGA      |
| c.7819A>C                        | Missense p.Thr2607Pro | VUS                     | -  | [-] SF2 (72.31), [-] SC35 (77.63), [-] 2 SRp40 (78.98; 85.27)                                  | [+] 2 Sironi ESS   | TGTTTCAGGGCTCTGTGTGACCTCCAGGTGTGGATCCAAAG<br>CTTTGGATCCACACCTGGAGGGTCACACAGAGCCCTGAACA  |
| c.7875A>G                        | Synonymous p.Arg2625= | VUS                     | [+] Cryptic 3' SS (NNS: 0.54 → 0.81)   | [+] SRp40 (81.74); [-] 2 Rescue-ESE  | [-] hnRNPA1 (69.29); [+] Fas-ESS;                            | TGGGTTTATAATCACTATAGGTGGATCATATGAAACTGGC<br>GCCAGTTTCCATATGATCCACCTATAGTGATTATAAACCCA   |
| c.7940T>C                        | Missense p.Leu2647Pro | VUS                     | -  | [-] SRp55 (78.98); [+] 9G8 (61.28)   | [-] hnRNPA1 (70.95); [+] Sironi ESS                          | GGAATTTGCTAATAGATGCCCAAGCCAGAAAGGGTGCTTC<br>GAAGCACCTTTCTGGGCTTGGGCATCTATTAGCAAATTCC    |
| c.7947A>G                        | Synonymous p.Pro2649= | VUS                     | -  | [-] SF2 (77.23), [-] SRp40 (79.40); [-] 2 Rescue-ESE   | [-] Sironi ESS   | GCTAATAGATGCCTAAGCCCGAAAGGGTGCTTCTTCAACT<br>AGTTGAAGAAGCACCTTTCCGGGCTTAGGCATCTATTAGC    |
| c.7952G>T                        | Missense p.Arg2651Met | VUS                     | -  | [-] SRp40 (79.40); [-] Rescue-ESE  | [-] 2 Sironi ESS; [-] Fas-ESS; [+] PESS; [-] hnRNPA1 (83.81) | TAGATGCCTAAGCCAGAAATGGTGCTTCTTCAACTAAAAT<br>ATTTTAGTTGAAGAAGCACCATTCTGGGCTTAGGCATCTA    |
| c.7971A>G                        | Synonymous p.Lys2657= | VUS                     | -  | [-] 2 Tra2-β (81.02; 85.42)  | [+] hnRNPA1 (70,48)  | AGGGTGCTTCTTCAACTAAAGTACAGGCAAGTTTAAAGCAT<br>ATGCTTTAAACTTGCCTGTACTTTAGTTGAAGAAGCACCT   |
| c.7975A>G                        | Missense p.Arg2659Gly | VUS                     | -  | [-] SRp40 (91.86), [+] SRp55 (75.65); [+] Rescue ESE   | [-] hnRNPA1 (74.76)  | CTTCTTCAACTAAAATACGGGCAAGTTTAAAGCATTAA<br>TAATGCTTTAAACTTGCCCGTATTTTAGTTGAAGAAG         |
| c.7976G>C                        | Missense p.Arg2659Thr | VUS                     | [-] GC donor (MaxEnt: 3.1 → 2.4)   | [+] 3 SF2 (70.85; 83.84; 80.66); [+] Rescue-ESE  | [-] 2 hnRNPA1 (74.76, 66.43)                                 | TTCTTCAACTAAAATACACGCAAGTTTAAAGCATTAC<br>GTAATGCTTTAAACTTGCGTGTATTTTAGTTGAAGAA          |
| c.7976G>A                        | Missense p.Arg2659Lys | VUS                     | [-] GC donor (MaxEnt: 3.1 → 2.32)  | [+] 2 Rescue-ESE   | [-] Sironi ESS; [-] 2 hnRNPA1 (74.76; 66.43)                 | GCTTCTTCAACTAAAATACAAGCAAGTTTAAAGCATTACAT<br>ATGTAATGCTTTAAACTTGCTTGTATTTTAGTTGAAGAAGC  |
| c.7976+1G>A                      | Intronic              | VUS                     | [-] GC donor (MaxEnt: 3.1 → -5.07)   | [-] SRp55 (80.71); [+] 3 SF2 (81.15; 83.58; 74.08); [+] 2 Rescue-ESE; [+] 2 9G8 (61.07; 63.76) | [-] Sironi ESS; [-] hnRNPA1 (74.76)                          | TCTTCAACTAAAATACAGACAAGTTTAAAGCATTACAA<br>TGTAATGCTTTAAACTTGCTGTATTTTAGTTGAAGA          |

|                  |                          |                     |   |   |  |  |
|------------------|--------------------------|---------------------|---|---|--|--|
| c.7976+35C>A     | Intronic                 | Neutral             | -   | [-] SRp55(86,98); [-] SF2 (77,46); [-] SF2 (73,44)                                  | [+] hnRNPA1 (74,76)  | TTACATTACGTAATCATATAAGGCAGTATGGTTAAGGTTTC<br>GAAACCTTAACCATACTGCCTTATATGATTACGTAATGTAA                   |
| <b>EXON 18</b>   |                          |                     |   |   |  |  |
| c.7977-119A>T    | Intronic                 | VUS                 | [+] 3'SS (NNS: <0.4 → 0.82)   | -   | [+] 2 PESS   | TTCTAGCTACAAAATTTTTATTCTCAGTATTTCTTAGAT<br>ATCTAAGAAATACTGAGAATAAAAAATTTGTAGCTAGGAA                      |
| c.7977-7C>G      | Intronic                 | VUS                 | [-] 3'SS (MaxEnt:11.5 →5.7)<br>[+] 3'SS 6-nt upstream (MAxEnt: 1.1 → 9.2) | [+] SC35 (80.15)  | [+] Fas-ESS  | AATATGCATTTTTGTTTTTCAGTTTTAGATATGATACGGAAA<br>TTTCCGTATCATATCTAAACTGAAAACAAAATGCATATT                    |
| c.7977-6T>G      | Intronic                 | VUS                 | 3'SS NNS: 0,98 → 0,90   |   |  | ATATGCATTTTTGTTTTTCAGTTTTAGATATGATACGGAAAT<br>ATTTCCGTATCATATCTAAACGTGAAAACAAAATGCATAT                   |
| c.7977-3_7978del | Intronic/exonic deletion | VUS                 | [-] 3'SS (NNS: 0,98 → <0,4)   | [-] Rescue-ESE  | [-] hnRNPA1 (76.90); [-] 2Fas-ESS                                      | TGCATTTTTGTTTTCACTTTTATGATACGGAAATTGATAGA<br>TCTATCAATTTCCGTATCATAAAGTGAACAAAATGCA                       |
| c.7977-1G>C      | Intronic                 | VUS                 | [-] 3'SS (NNS: 0.98→<0.4)   | [+]SRp55 (85.90); [-] Rescue ESE  | [-] Sironi ESS; [-] 2 Fas-ESS; [+] PESS; [-]hnRNPA1 (76.90)            | CATTTTTGTTTTCACTTTTACATATGATACGGAAATTGATA<br>TATCAATTTCCGTATCATATGTAAAAGTGAACAAAATG                      |
| c.7977-2A>T      | Intronic                 | VUS                 | [-] 3'SS (NNS: 0,98 → <0,4)   |   | [-] hnRNPA1 (76.90); [-] 2Fas-ESS                                      | GCATTTTTGTTTTCACTTTTATGATATGATACGGAAATTGAT<br>ATCAATTTCCGTATCATATCAAAAAGTGAACAAAATGC                     |
| c.7977-1G>T      | Intronic                 | VUS/causal          | [-] 3'SS (NNS: 0,98 → <0,4)   | [+]SRp55 (75.97)  |  | CATTTTTGTTTTCACTTTTATATATGATACGGAAATTGATA<br>TATCAATTTCCGTATCATATATAAAAGTGAACAAAATG                      |
| c.7984dupA       | Frameshift               | Causal (frameshift) | -   | [-] SRp40 (78.98), [-] SF2 (73.09), [-] SRp55 (84.62); [+] 9G8 (75.24)              | -  | TTTTCACTTTTATGATATGATAACGGAAATTGATAGAAGCAGA<br>TCTGCTTCTATCAATTTCCGTATCATATCTAAAAGTGA                    |
| c.7985C>G        | Missense p.Thr2662Arg    | VUS                 | [+] Cryptic 3'SS (NNS: <0.4 → 0.55)                                       | [-] SF2 (73.09), [-] SRp55 (84.62)  | [+] 3 Sironi ESS; [+] 2 Fas-ESS  | TCACTTTATGATATGATAGGGAAATTGATAGAAGCAGAA<br>TTCTGCTTCTATCAATTTCCGTATCATATCTAAAAGTGA                       |
| c.7985C>T        | Missense p.Thr2662Met    | VUS                 | -   | [-] SRp40 (78.98), [-] SF2 (73.09); [+] 4 Rescue-ESEs                               | [+] PESS   | TTCACTTTTATGATATGATATGAAATTGATAGAAGCAGAA<br>TTCTGCTTCTATCAATTTCCATATCATATCTAAAAGTGA                      |
| c.7988A>T        | Missense p.Glu2663Val    | VUS                 | [+] Cryptic 5'SS (NNS: <0.4 → 0.96)                                       | [-] SRp55 (84.62); [-] PESE   | -  | CACTTTATGATATGATACGGTAATTGATAGAAGCAGAAGATC<br>GATCTTCTGCTTCTATCAATTACCGTATCATATCTAAAAGT                  |
| c.7992T>A        | Synonymous p.Ile2664=    | VUS                 | -   | [-] PESE; [+] Tra2-beta [87.60]   | [+] hnRNPA1 (76.90)  | TTAGATATGATACGGAAATAGATAGAAGCAGAAGATCGG<br>CCGATCTTCTGCTTCTATCTATTTCCGTATCATATCTAAA                      |
| c.8007A>G        | Synonymous p.Arg2669=    | VUS                 | -   | [-] 2 Rescue-ESEs; [-] 4 PESE   | [+] 2 hnRNPA1 (70.48; 71.19); [+] 2 Sironi ESS                         | GAAATTGATAGAAGCAGAAGGTCGGCTATAAAAAAGATAAT<br>ATTATCTTTTTTATAGCCGACCTTCTGCTTCTATCAATTT                    |
| c.8009C>A        | Nonsense p.Ser2670X      | Causal (nonsense)   | -   | [-] Rescue-ESE; [-] PESE  | [+] hnRNPA1 (89.05); [+] Fas-ESS                                       | AATTGATAGAAGCAGAAGATAGGCTATAAAAAAGATAATGG<br>CCATTATCTTTTTTATAGCCTATCTTCTGCTTCTATCAATT                   |
| c.8009C>G        | Missense p.Ser2670Trp    | VUS                 | -   | [-] PESE; [+] 9G8 (64.70)   | [+] hnRNPA1 (66.43); [+] 2 Sironi ESS; [+] 2 Fas-ESS; [+] PESS (24.86) | AATTGATAGAAGCAGAAGATGGGCTATAAAAAAGATAATGG<br>CCATTATCTTTTTTATAGCCCATCTTCTGCTTCTATCAATT                   |
| c.8009C>T        | Missense p.Ser2670Leu    | VUS                 | -   | [-] Rescue-ESE; [-] PESE  | [+] Sironi ESS   | AATTGATAGAAGCAGAAGATTGGCTATAAAAAAGATAATGG<br>CCATTATCTTTTTTATAGCCCATCTTCTGCTTCTATCAATT                   |
| c.8010_8032del   | Frameshift               | Causal (frameshift) | -   | [-]SC35 (76.09); [-] SRp40 (80.96); [-] 6 Rescue-ESE; [-] 4 Tra2-beta; [-] 2 9G8    | [-] 2 hnRNPA1 (65.48; 84.76); [-] 4 Sironi-ESS; [-] Fas-ESS; [-] PESS  | CGGAAATTGATAGAAGCAGAAGATCGGGATGACACAGCTGCAAAAACACT<br>AGTGTTTTGTCAGCTGTGTCATCCCGATCTTCTGCTTCTATCAATTTCCG |
| c.8019A>G        | Synonymous p.Lys2673=    | VUS                 | -   | [-] Tra2-beta (83.21); [+] Rescue-ESE; [+] PESE                                     | [+] hnRNPA1 (71.43); [-] PESS  | CAGAAGATCGGCTATAAAGAAGATAATGGAAAGGGATGA<br>TCATCCCTTTCCATTATCTTCTTATAGCCGATCTTCTG                        |
| c.8023A>G        | Missense p.Ile2675Val    | Causal (splicing)   | [+] Cryptic 5'SS (NNS: <0.4 →0.92)  | [-]SRp40 (80.96); [-] 2 Rescue-ESE; [-] 2 9G8 (65.17; 62.15); [-] Tra2-beta (86.88) | [+] 2 hnRNPA1 (66.67; 73.10); [+] Fas-ESS; [+] PESS                    | GAAGATCGGCTATAAAAAAGGTAATGGAAAGGGATGACACA<br>TGTGTCATCCCTTTCCATTACCTTTTTTATAGCCGATCTTC                   |

|                             |                               |   |  |   |   |   |
|-----------------------------|-------------------------------|---|--|---|---|---|
| c.8027T>C                   | Missense<br>p.Met2676Thr      | VUS   | -                                      | [-] SRp40 (80.96); [-] 3<br>Rescue-ESE  | [-] Sironi ESS; [-] Fas-ESS   | CGGCTATAAAAAAGATAACGGAAAGGGATGACACAGC<br>GCTGTGTCATCCCTTTCCGTTATCTTTTTTATAGCCG  |
| c.8027T>A                   | Missense<br>p.Met2676Lys      | VUS   | -                                      | [+] gG8 (61.34)   | [+] 2 hnRNPA1 (68.57;<br>76.43); [-] Fas-ESS                        | ATCGGCTATAAAAAAGATAAAGGAAAGGGATGACACAGCTG<br>CAGCTGTGTCATCCCTTTTCCTTTATCTTTTTTATAGCCGAT                                     |
| c.8035G>T                   | Missense<br>p.Asp2679Tyr      | Causal (splicing)   | [+] Cryptic 5'SS (NNS: <0.4<br>→ 0.96) | -   | [+] 2 hnRNPA1 (66.67;<br>70.48); [-] 2 Sironi ESS;[+] PESS          | GATAATGGAAGGTATGACACAGCTGCG<br>GCAGCTGTGCATACCTTTCCATTATC   |
| c.8039A>G                   | Missense<br>p.Asp2680Gly      | VUS   | -                                      | [-]SRp40 (78.98)  | [+]hnRNPA1 (67.38) [+] Sironi ESS; [+]Fas-ESS; [+]PESS              | AAAGATAATGGAAGGGATGGCACAGCTGCAAAAACACTTG<br>CAAGTGTTTTTGCAGCTGTGCCATCCCTTTCCATTATCTTT                                       |
| c.8042C>G                   | Missense<br>p.Thr2681Arg      | VUS   | -                                      | -   | [+] hnRNPA1 (71.43)   | GGAAAGGGATGACAGAGCTGCAAAAACAC<br>GTGTTTTTGCAGCTCTGTCATCCCTTTCC  |
| c.8072C>T                   | Missense<br>p.Ser2691Phe      | VUS   | -                                      | [-] 2 SF2 (76.08; 75.01);<br>[-] SC35 (77.81)   | [-] Sironi ESS; [-] Fas-ESS;<br>[+] 2 PESS                          | CAGTTGTTCTCTGTGTTTTTGACATAATTTTCATTGAGC<br>GCTCAATGAAATTATGTCAAAAACACAGAGAACAAGTG   |
| c.8084C>T                   | Missense<br>p.Ser2695Leu      | VUS   | -                                      | [-] PESE  | [+] hnRNPA1 (65.48); [+] 2 PESS                                     | CTGTGTTTCTGACATAATTTTATTGAGCGCAAATATATCTG<br>CAGATATATTTGCGCTCAATAAAATTATGTCCAGAAACACAG                                     |
| c.8111C>T                   | Missense<br>p.Ser2704Phe      | VUS   | -                                      | [-] SC35 (78.18);[-] SRp40<br>(84.79); [-] Rescue-ESE   | [-] Sironi ESS  | GCGCAAATATATCTGAACTTTTAGCAATAAAACATAGTAGTGC<br>GCACTACTAGTTTTATTGCTAAAAGTTTCAGATATATTTGCGC                                  |
| c.8165C>G                   | Missense<br>p.Thr2722Arg      | VUS   | -                                      | [-] 2 SF2 (77.62;79.62); [-] SRp40 (83.89); [-] SRp55 (80.71); [+]Rescue-ESE  |   | GGCCATTATTGAACTTAGAGATGGGTGGTATGCTG<br>CAGCATACCACCATCTCTAAGTTCAATAATGGCC   |
| c.8168A>G                   | Missense<br>p.Asp2723Gly      | Causal (missense-splicing)  | [+] Cryptic 5'SS (NNS: <0.4<br>→0.81)  | [-] 2 SF2 (77.62; 79.62); [-] Rescue-ESE; [-] gG8 (64.70)   | [-] Sironi ESS; [+] Fas-ESS   | GCCATTATTGAACTTACAGGTGGGTGGTATGCTGTTAAG<br>CTTAACAGCATACCACCACCTGTAAGTTCAATAATGGC   |
| c.8191C>T                   | Nonsense<br>p.Gln2731X        | Causal (nonsense)   |  | [-]SF2 (77.23)  | [+] hnRNPA1 (73.10); [-] Sironi ESS; [+] Fas-ESS                    | GGTGGTATGCTGTTAAGGCCTAGTTAGATCCTCCCCTCTTA<br>TAAGAGGGGAGGATCTAACTAGGCCTTAACAGCATACCACC                                      |
| c.8219insT                  | Frameshift<br>p.Leu2740?fs    | Causal (frameshift)   | [+] Cryptic 3'SS (NNS: <0.4<br>→ 0.44) | [+] SRp40 (79.04)   | [+] PESS  | CCTCCCCTCTTAGCTGTCTTTAAAGAAATGGCAGACTGAC<br>GTCAGTCTGCCATTCTTTAAAGACAGCTAAGAGGGGAGG   |
| c.8249_8250del              | Frameshift<br>p.Lys2750AsnfsX | Causal (frameshift)   |  | [-] 2 SF2 (80.15; 83.87); [-] SRp40 (86.29); [-] 2 gG8 (86.91; 62.15) [-] Tra2-beta (86.88); [-] 2 Rescue-ESE; [-] 4 PESE | [-] Sironi ESS; [-] Fas-ESS   | CAGACTGACAGTTGGTCAGAATTATTTCTTCATGGAGCAGA<br>TCTGCTCCATGAAGAATAATTTGACCAACTGTCAGTCTG  |
| c.8331G>A                   | Synonymous<br>p.Lys2777=      | VUS   | [-] 5'SS (NNS: 0.96 →<0.4)             | [+]Tra2-beta (83.21)  | [-] 3 hnRNPA1 (68.57; 66.67; 73.10); [-] Sironi ESS; [-] Fas-ESS    | CCAGAATCTCTTATGTTAAAGTAAATTAATTTGCACTCT<br>AGAGTGCAAATTAATTTACTTTTAACATAAGAGATTCTGG   |
| c.8331+1G>T                 | Intronic                      | Causal (splicing)   | [-] 5'SS (NNS: 0.96 →<0.4)             | [-] SRp40 (88.74)   | [-] 2 hnRNPA1 (66.67; 73.10); [-] Sironi ESS; [-] Fas-ESS; [-] PESS | CAGAATCTCTTATGTTAAAGTTAAATTAATTTGCACTCTTG<br>CAAGAGTGCAAATTAATTTAACTTTAACATAAGAGATTCTG                                      |
| c.8331+2T>C                 | Intronic                      | Causal (splicing)   | [-] 5'SS (NNS: 0.96 →<0.4)             | [+] gG8   | [-] Fas-ESS   | AGAATCTCTTATGTTAAAGGCAAATTAATTTGCACTCTTG<br>CCAAGAGTGCAAATTAATTTGCCTTTAACATAAGAGATTCT                                       |
| Artificial mutations        |                               | Bioinformatics <sup>2,3,4</sup>   |  |   |   | Primers for Site-directed Mutagenesis (5'→3')   |
| Microdeletions <sup>5</sup> |                               |   |  |   |   |   |
| c.7808_7837del30            |                               | -   |  |   |   | ATATTCTACTTTTATTGTTTCAGGGAGCTTATTCTAGAATTTGGGTTTA<br>TAAACCAAATCTAGAAATAAGCTCCCTGAACAAATAAAGTAGAATAT                        |
| c.7833_7862del30            |                               | -   |  |   |   | CAGGGCTCTGTGTGACACTCCAGGTGTGGATAATCACTATAGATGGATCATATGGAACT<br>AGTTTCCATATGATCCATCTATAGTGATTATCCACACCTGGAGTGTCACACAGAGCCCTG |
| c.7919_7948del30            |                               | -   |  |   |   | AACTGGCAGCTATGGAATGTGCCTTTCCATAAAGGGTGCTTCTTCAACTAAAATACAGGC<br>GCCTGTATTTTAGTTGAAGAAGCACCTTTTAGGAAGGACACATTCATAGCTGCCAGTT  |
| c.7944_7973del30            |                               | ESEs: [-] SF2 (IgM-BRCA1) (77.2); 3 SRp40 (79.4, 81.2,91.9); SRp55 (81.4); [-] 3 Rescue-ESEs;[-] 14 PESE; [-] 2 |  |   |   | TCCTAAGGAATTTGCTAATAGATGCCTAAGCAGGCAAGTTAAAGCATTACATTACGTAA   |

|                               |   |  |
|-------------------------------|---|--|
|                               | Tra2-β (81, 85.4)   | TTACGTAATGTAATGCTTTAACTTGCCCTGCTTAGGCATCTATTAGCAAATTCCTTAGGA   |
| c.7944_7953del10              | -   | TCCTAAGGAATTGCTAATAGATGCCTAAGGTGCTTCTTCACTAAAAACACAGGCAAGTT<br>AACTTGCCGTGATTTTAGTTGAAGAAGCACCTTAGGCATCTATTAGCAAATTCCTTAGGA  |
| c.7954_7963del10              | [-] SRp55 (81.4), SRp40 (81,2); [+]<br>SF2/ASF (73.8); [±] Rescue-ESEs; [+]<br>hnRNP A1 (67.6)  | TTTGCTAAATAGATGCCTAAGCCCAAGGAAGAACTAAAAACACAGGCAAGTTTAAAGCATT<br>TAATGCTTTAACTTGCCGTGATTTTAGTTCTCTTCTGGGCTTAGGCATCTATTAGCAAA |
| c.7964_7973del10              | [-] SRp40 (91.9); [-] 2 Tra2-β (81, 85.4); [-] 1 Rescue-ESE   | GATGCCCTAAGCCCAAGGAAGGGTGCCTTCTCCAGGCAAGTTTAAAGCATTACATTACGTAA<br>TTACGTAATGTAATGCTTTAACTTGCCTGGAAGAAGCACCTTCTTGGGCTTAGGCATC |
| c.7979_8008del30              | ESEs: [-] 2 SRp40; (78.98,78.14); SF2 (73.09); SF2 (lgM-BRCA1) (80.15, 83.87); SRp55 (84.62); [-] 10 Rescue-ESEs; [-] 8 PESE; [-] 4 9G8 (60.94, 68.52, 70.74, 86.91), [-] Tra2-β (86.88)              | TGCATTTTTGTTTTCACTTTTAGATCGGCTATAAAAAGATAATGGAAAG<br>CTTTCATTATCTTTTTATAGCCGATCTAAAGTGAAACAAAAATGCA                          |
| c.7979_7988del10              | ESEs: SRp40 (78.98) SF2 (73.09); SRp55 (84.62)  | AAATATGCATTTTTGTTTTCACTTTTAGATAATTGATAGAAGCAGAAGATCGGCTATAAA<br>TTTATAGCCGATCTTCTGCTTCTATCAATTATCTAAAAGTGAAACAAAAATGCATATTT  |
| c7989_7998del10               | -   | TTTTGTTTTCACTTTTAGATATGATACGGAAGCAGAAGATCGGCTATAAAAAAGATAATG<br>CATTATCTTTTTTATAGCCGATCTTCTGCTTCCGTATCATATCTAAAAGTGAAAACAAAA |
| c.7999_8008del10              | ESEs: SRp40 (78.14); SF2 (83.87); Tra2-β (86.88)  | ACTTTTAGATATGATACGGAATTGATAGACGGCTATAAAAAAGATAATGGAAAGGGATG<br>CATCCCTTTCCATTATCTTTTTTATAGCCGCTATCAATTTCCGTATCATATCTAAAAGT   |
| c.8004_8033del30              | ESEs: [-] SF2 (lgM-BRCA1) (80.15), SF2 (83.87), SC35 (76.09), SRp40 (80.96); [-] 12 Rescue-ESEs; [-] 7 PESE; [-] 4 9G8 (70.74, 86.91, 65.17, 62.15); [-] 5 Tra2-β (86.88, 94.14, 94.14, 83.21, 86.88) | ATGATACGGAATTGATAGAAGCAGGGATGACACAGCTGCAAAAAACACTT<br>AAGTGTTTTTGACGTGTGTCATCCCTGCTTCTATCAATTTCCGTATCAT                      |
| c.8004_8013del10              | ESEs: SF2 (83.87); Tra2-β (86.88); SC35 (76.09)   | TAGATATGATACGGAATTGATAGAAGCAGATAAAAAAGATAATGGAAAGGGATGACACA<br>TGTGTCATCCCTTTCCATTATCTTTTTTATCTGCTTCTATCAATTTCCGTATCATATCTA  |
| c.8014_8023del10              | -   | ACGGAATTGATAGAAGCAGAAGATCGGCTTAATGGAAAGGGATGACACAGCTGCAAAAA<br>TTTTTGACGTGTGTCATCCCTTTCCATTAGCCGATCTTCTGCTTCTATCAATTTCCGT    |
| c.8024_8033del10              | -   | ATAGAAGCAGAAGATCGGCTATAAAAAAGAGGATGACACAGCTGCAAAAAACACTTGTCT<br>AGAACAAGTGTTTTTGACGTGTGTCATCCCTTTTTTATAGCCGATCTTCTGCTTCTAT   |
| c.8274_8303del30              | -   | AGATTATCTTTCATGGAGCAGAAGTTGAAGCCCAAGATCTCTTATGTTA<br>TAACATAAGAGATTCTGGGGCTTCAAGTTCTGCTCCATGAAGAATAATCT                      |
| c.8299_8328del30              | -   | GGTGGGCTCTCCTGATGCCTGTACAAAGGTAAATTAATTTGCACTCTTGG<br>CCAAGAGTGCAAAATTAATTTACCTTTGTACAGGCATCAGGAGAGCCACC                     |
| c.[7984A>T; 8001C>T; 8003G>A] | ESE disruptions: SF2 (73.09); SF2 (83.87); SRp40 (78.98); SRp55 (84.62)   | ACTTTTAGATATGATTTCGGAATTGATAGAAGTAAAGATCGGCTATAAA<br>TTTATAGCCGATCTTTTACTTCTATCAATTTCCGAATCATATCTAAAAGT                      |
| c.7984A>T                     | ESE disruptions: SF2 (73.09); SRp40 (78.98); SRp55 (84.62)  | TTTTCACTTTTAGATATGATTTCGGAATTGATAGAAGCAGA<br>TCTGCTTCTATCAATTTCCGAATCATATCTAAAAGTGAAAA                                       |
| c.8001C>T                     | ESE disruptions: SF2 (83.87)  | GATACGGAATTGATAGAAGTAGAAGATCGGCTATAAAAA<br>TTTTTTATAGCCGATCTTCTACTTCTATCAATTTCCGTATC   |
| c.8003G>A                     | ESE disruptions: SF2 (83.87)  | TACGGAATTGATAGAAGCAAGATCGGCTATAAAAAAGA<br>TCTTTTTTATAGCCGATCTTTTGCTTCTATCAATTTCCGTA  |
| c.7976+21_7976+14odel         | -   | TAAAAACAGGCAAGTTTAAAGCATTACATATCCCTAGCCCCCATTAAAGAGAGATCACA<br>TGTGATCTCTCTTAAATGGGGCTAGGGATATGTAATGCTTTAAACTTGCCGTGATTTTTA  |
| c.7976+136_7976+24odel        | -   | TGACAGTTGCCATCCCACTGCTGTTCTCATATGAGTTACTAATTTGATCCACTATTTG<br>CAAATAGTGATCAAATTAGTAAGTCTATGAGAACAGCAGTGTGGGATGGCAACTGTCA     |
| c.7976+231_7977-141del        | ESEs: [-] 3 SRp40; [-] 3 SC35; [-] SF2; [-] SRp55; [-] 4 9G8; [-] 10 Tra2-beta; [-] 10 Rescue-ESE; [-] 3 PESE<br>ESSs: [-] 4 hnRNP A1; [-] 5 Sironi ESS; [-] 5 Fas-ESS; [-] 18 PESS                   | CCACCCCTCCTTAACCTCTTGATGATGAATTCCTAGCTACAAAAATTTTTAATTCTCAG<br>CTGAGAATTAATAATTTTGTAGCTAGGAATTCATACATCAAGAGGTTAAGGAGGGGGTGG  |
| c.7977-150_7977-21del         | -   | AAAATAATTGATATTTAACAATATGAAACATTTTGTGTTTCACTTTTAGATATGATACG<br>CGTATCATATCTAAAAGTGAAACAAAAATGTTTCATATGTTAAAAATATCAATATTTTT   |
| c.7976+2C>T                   | Switch to a canonical GT site (NNSplice= 1.0)   | TTCTTCAACTAAAAACAGGTAAGTTTAAAGCATTACATTA<br>TAATGTAATGCTTTAACTTACCTGTATTTTAGTTGAAGAA   |

- 1- Spliceogenic variants are highlighted in yellow.
- 2- HSF: Human Splicing Finder. NNS: NNSplice. [+] and [-] symbols indicate creation or disruption, respectively, of splicing regulatory sequences. The most relevant data are indicated in bold-type.
- 3- **Cut-offs of the splicing programs:**  
**Splice Sites:**  
NNSPLICE (values 0-1): Cut-off=0.4 for both 5' and 3' splice sites (Reese et al., 1997).  
HSF: Human Splicing Finder matrices (default values, cut-offs cannot be defined) (Desmet et al., 2009).  
MaxEnt: 0 for 5' and 3' splice sites (Yeo and Burge, 2004).  
**Enhancers and Silencers:**  
ESEfinder cut-offs (HSF scale 0-100): SF2/ASF: 72.98 / SF2/ASF (lgM-BRCA1): 70.51 / SRp40: 78.08 / SC35: 75.05 / SRp55: 73.86 (Cartegni et al., 2003)  
ESE motifs from HSF, cut-offs values (0-100): Tra2: 75.964 / 9G8: 59.245

Algorithms Rescue-ESE, Fas-ESS, PESE and PESS predict generation or disruption of motifs with presumed ESE or ESS activity (Fairbrother et al., 2002; Zhang and Chasin, 2004; Wang et al., 2004)

Sironi silencers (Sironi et al., 2004), cut-offs values (0-100): Motif 1: 60 / Motif 2: 60 / Motif 3: 60.

hnRNP motifs: hnRNP A1: 65.476

- 4- Bioinformatics data of microdeletions with no impact on splicing are not presented, except for the intronic deletion c.7976+231\_7977-141del where a high concentration of Tra2-beta motifs were observed.
- 5- Positive microdeletions and artificial point mutations are shown in red.

**S3 Table.** Relative quantification of transcripts induced by DNA variants and microdeletions of *BRCA2* exons 17 and 18 in MCF-7 cells.

| DNA variants                    | Canonical Transcript <sup>1</sup> | Exon 17 skipping  | Ex17-del20        | Ex17-del69  | Exon17-insAG      | lvs16-ins8  | Ex17-del1   | Other aberrant transcripts <sup>3</sup> |
|---------------------------------|-----------------------------------|-------------------|-------------------|-------------|-------------------|-------------|-------------|---|
| <b>EXON 17</b>                  |                                   |                   |                   |             |                   |             |             |   |
| c.7806-9T>G                     |                                   | <b>41.5% ±0.3</b> |                   | 22.2% ±0.2  |                   | 36.3% ±0.5  |             |   |
| c.7806-2A>G                     |                                   | 20.1% ±0.3        | <b>51.8% ±0.3</b> | 28.1% ±0.1  |                   |             |             |   |
| c.7806-1G>A                     |                                   |                   |                   |             |                   |             | <b>100%</b> |   |
| c.7806-1G>T                     |                                   |                   | <b>100%</b>       |             |                   |             |             |   |
| c.7806insAG                     |                                   | 5.1% ±0.5         |                   | 2.3% ±0.1   | <b>92.6% ±0.6</b> |             |             |   |
| c.7875A>G                       | <b>95.3% ±1.1</b>                 |                   |                   | 4.7% ±1.1   |                   |             |             |   |
| c.7975A>G                       | <b>73.8% ±0.1</b>                 | 26.2% ±0.1        |                   |             |                   |             |             |   |
| c.7976G>C                       |                                   | <b>100%)</b>      |                   |             |                   |             |             |   |
| c.7976G>A                       |                                   | <b>100%</b>       |                   |             |                   |             |             |   |
| c.7976+1G>A                     |                                   | <b>100%</b>       |                   |             |                   |             |             |   |
| <b>Microdeletions</b>           |                                   |                   |                   |             |                   |             |             |   |
| c.7944_7973del                  | <b>61.6% ±0.2</b>                 | 38.4 ±0.2         |                   |             |                   |             |             |   |
| c.7944_7953del                  | <b>100%</b>                       |                   |                   |             |                   |             |             |   |
| c.7954_7963del                  | <b>97.8%±0.5</b>                  | 1.0±0.1           |                   |             |                   |             |             | 1.2%±0.5                                |
| c.7964_7973del                  | <b>95.4%±0.2</b>                  | 4.6%±0.2          |                   |             |                   |             |             |   |
| <b>EXON 18</b>                  |                                   |                   |                   |             |                   |             |             |   |
| DNA variants                    | Canonical Transcript              | Exon 18 skipping  | Ex18-ins6         | Ex18-del309 | Ex18-del298       | Ex18-del164 | Ex18-del191 | Other aberrant transcripts <sup>3</sup> |
| Wild type Minigene <sup>2</sup> | <b>99.3 ±0.8</b>                  | 0.7% ±0.8         |                   |             |                   |             |             |   |
| c.7977-7C>G                     |                                   | 21.6% ±0.5        | <b>78.4% ±0.5</b> |             |                   |             |             |   |
| c.7977-6T>G                     | <b>66.7% ±0.8</b>                 | 31% ±0.8          |                   |             |                   |             | 2.3%        |   |
| c.7977-3_7978del                |                                   | <b>90% ±1.1</b>   |                   |             |                   |             | 10% ±1.1    |   |
| c.7977-2A>T                     |                                   | 93.3% ±0.1        |                   |             |                   |             | 6.7% ±0.1   |   |
| c.7977-1G>C                     |                                   | <b>89.8% ±1.8</b> |                   |             |                   |             | 10.2% ±1.8  | -                                       |
| c.7977-1G>T                     |                                   | <b>91.5% ±0.1</b> |                   |             |                   |             | 7%±0.1      | ex18-del236=1.5%                        |
| c.7985C>G                       |                                   | <b>90.2% ±4.4</b> |                   |             |                   |             | 5.0% ±0.1   | 4.7% ±4.5                               |
| c.7985C>T                       | <b>96.7 ±1.3</b>                  | 3.3% ±1.3         |                   |             |                   |             |             |   |
| c.7988A>T                       | <b>84.2% ±1.5</b>                 | 8.6% ±1.7         |                   |             |                   |             |             | 7.2% ± 0.4                              |
| c.7992T>A                       | <b>68.6% ±1.0</b>                 | 31.4% ±1.0        |                   |             |                   |             |             |   |
| c.8007A>G                       | <b>84.8% ±0.6</b>                 | 15.2%±0.6         |                   |             |                   |             |             |   |
| c.8009C>A                       | 4.0 ±3.4                          | <b>91.2% ±4.4</b> |                   |             |                   |             | 4.8% ±3.9   | -                                       |
| c.8009C>T                       | <b>76.6% ±6.1</b>                 | 23.4% ±6.1        |                   |             |                   |             |             | -                                       |

|                       |                   |                   |  |                 |                    |            |           |  |
|-----------------------|-------------------|-------------------|--|-----------------|--------------------|------------|-----------|--|
| c.8009C>G             | <b>79.9 ±1.2</b>  | 20.1% ±1.2        |  |                 |                    |            |           |  |
| c.8023A>G             |                   | 0.7% ±0.1         |  | <b>93% ±0.3</b> |                    |            |           | 1.0% ±0.1/<br>4.2±0.2                      |
| c.8035G>T             | 2.4% ± 0.5        |                   |  |                 | <b>93.6% ± 0.4</b> |            |           | 4.0±0.2                                    |
| c.8042C>G             | <b>97.7% ±0.1</b> | 2.3% ±0.1         |  |                 |                    |            |           |  |
| c.8072C>T             | <b>94.9% ±1.2</b> | 5.1% ±1.2         |  |                 |                    |            |           |  |
| c.8168A>G             | <b>69.6% ±0.9</b> | 4.5% ±0.4         |  |                 |                    | 25.9% ±0.5 |           |  |
| c.8249_50del          | <b>93.0 ±1.2</b>  | 7.0 ±1.2          |  |                 |                    |            |           |  |
| c.8331G>A             | 40.7% ±0.6        | <b>52.0% ±0.4</b> |  |                 |                    |            |           | 2.1% ±0.1 / 1.7<br>±0.1 / 3.5%±0.2         |
| c.8331+1G>T           |                   | <b>81.0 ±0.3</b>  |  |                 |                    |            |           | 6.1%±0.2% /<br>3.7% ±0.3 /<br>6.4% ±0.4    |
| c.8331+2T>C           |                   | <b>87.1%±0.3</b>  |  |                 |                    |            |           | [ex17-del151<br>+ex18 skipping]<br>(12.9%) |
| <b>Microdeletions</b> |                   |                   |  |                 |                    |            |           |  |
| c.7979_8008del        | 47.2% ±0.4        | <b>49.6% ±0.4</b> |  |                 |                    |            | 3.2% ±0.1 |  |
| <i>c.7979_7988del</i> | <b>79.8% ±0.7</b> | 20.2%±0.7         |  |                 |                    |            |           |  |
| <i>c.7989_7998del</i> | <b>100%</b>       |                   |  |                 |                    |            |           |  |
| <i>c.7999_8008del</i> | 48.3% ±0.4        | <b>51.7% ±0.4</b> |  |                 |                    |            |           |  |
| c.8004_8033del        | 30.6% ±0.2        | <b>64.7% ±0.0</b> |  |                 |                    |            | 4.7% ±0.1 |  |
| <i>c.8004_8013del</i> | 16.3% ±0.1        | <b>78.1% ±0.1</b> |  |                 |                    |            | 5.6%      |  |
| <i>c.8014_8023del</i> | <b>100%</b>       |                   |  |                 |                    |            |           |  |
| <i>c.8024_8033del</i> | <b>100%</b>       |                   |  |                 |                    |            |           |  |

Relative proportions (%) of each transcript with their standard deviations (±%) are indicated. Principal transcripts are shown in bold type.

<sup>1</sup> Size of the canonical transcript 1012 nt after cDNA amplification with primers RTBR2\_ex16-FW and RTpSAD-RV.

<sup>2</sup> The wild type minigenes showed a minor transcript with exon 18 skipping ranging from 0% to 2.6%.

<sup>3</sup> Other aberrant transcripts: [Ex17del151+Ex18 skipping], [Ex18 sk + ivs17ins58], [Ex18del156], [~710 nt],[ 775 nt], [822 nt], [867 nt], [878 nt], [1021 nt]

**S4 Table.** Analysis of the 52 assayed variants according to the guidelines of the American College of Medical Genetics and Genomics (ACMG).

| DNA variant <sup>1</sup> | Previous classification    | ACMG Criteria <sup>2</sup>       | Potential ACMG Classification | # database records |
|--------------------------|----------------------------|----------------------------------|-------------------------------|--------------------|
| <b>EXON 17</b>           |                            |                                  |                               |                    |
| c.7806-40A>G             | VUS                        | PP3, BS3                         | VUS                           | 28                 |
| c.7806-9T>G              | VUS                        | PS3, PM2, PM4, PP3               | Likely Pathogenic (SP.)       | 1                  |
| c.7806-2A>G              | Causal                     | PVS1 (SP.), PS3, PM2, PM4, PP3   | Pathogenic (SP.)              | 7                  |
| c.7806-1G>A              | Causal                     | PVS1 (SP.), PS3, PM2, PP3        | Pathogenic (SP.)              | 1                  |
| c.7806-1G>T              | Causal                     | PVS1 (SP.), PS3, PM2, PP3        | Pathogenic (SP.)              | 1                  |
| c.7806-1_7806-2dup       | Causal (frameshift)        | PVS1 (SP.), PS3, PM2, PP3        | Pathogenic (SP.)              | 2                  |
| c.7819A>C                | VUS                        | PM2, PP3, BS3                    | VUS                           | 1                  |
| c.7875A>G                | VUS                        | PM2, PP3, BS3                    | VUS                           | 2                  |
| c.7940T>C                | VUS                        | PM2, PP3, BS3                    | VUS                           | 6                  |
| c.7947A>G                | VUS                        | PP3, BS3                         | VUS                           | 1                  |
| c.7952G>T                | VUS                        | PM2, PP3, BS3                    | VUS                           | 1                  |
| c.7971A>G                | VUS                        | PM2, PP3, BS3                    | VUS                           | 2                  |
| c.7975A>G                | VUS                        | PM2, PP3                         | VUS                           | 17                 |
| c.7976G>C                | VUS                        | PS3, PM2, PP3                    | Likely Pathogenic (SP.)       | 2                  |
| c.7976G>A                | VUS                        | PS3, PP3, PP5                    | Likely Pathogenic (SP.)       | 18                 |
| c.7976+1G>A              | VUS                        | PVS1 (SP.), PS3, PM2, PP3        | Pathogenic (SP.)              | 3                  |
| c.7976+35C>A             | Neutral                    | PP3, BS3                         | VUS                           | 16                 |
| <b>EXON 18</b>           |                            |                                  |                               |                    |
| c.7977-119A>T            | VUS                        | PP3, BS3                         | VUS                           | 1                  |
| c.7977-7C>G              | VUS                        | PS3, PM2, PM4, PP3               | Likely Pathogenic (SP.)       | 4                  |
| c.7977-6T>G              | VUS                        | PM2, PP3                         | VUS                           | 1                  |
| c.7977-3_7978del         | VUS                        | PVS1 (SP.), PS3, PM2, PP3        | Pathogenic (SP.)              | 1                  |
| c.7977-2A>T              | VUS                        | PVS1 (SP.), PS3, PM2, PP3        | Pathogenic (SP.)              | 1                  |
| c.7977-1G>T              | VUS                        | PVS1 (SP.), PS3, PM2, PP3        | Pathogenic (SP.)              | 1                  |
| c.7977-1G>C              | VUS/causal                 | PVS1 (SP.), PS3, PM2, PP3        | Pathogenic (SP.)              | 18                 |
| c.7984dupA               | Causal (frameshift)        | PVS1 (Fr.), PM2, PP3             | Pathogenic (Fr.)              | 1                  |
| c.7985C>G                | VUS                        | PS3, PM2, PP3                    | Likely Pathogenic (SP.)       | 1                  |
| c.7985C>T                | VUS                        | PM2, PP3, BS3                    | VUS                           | 5                  |
| c.7988A>T                | VUS                        | PM2, PP3                         | VUS                           | 9                  |
| c.7992T>A                | VUS                        | PP3                              | VUS                           | 6                  |
| c.8007A>G                | VUS                        | PP3                              | VUS                           | 2                  |
| c.8009C>A                | Causal (nonsense)          | PVS1 (Non.), PS3 (SP.), PM2, PP3 | Pathogenic (SP.)              | 1                  |
| c.8009C>G                | VUS                        | PM2, PP3                         | VUS                           | 6                  |
| c.8009C>T                | VUS                        | PM2, PP3                         | VUS                           | 17                 |
| c.8010_8032del           | Causal (frameshift)        | PVS1 (Fr.), PM2, PP3             | Pathogenic (Fr.)              | 1                  |
| c.8019A>G                | VUS                        | PM2, PP3, BS3                    | VUS                           | 1                  |
| c.8023A>G                | Causal (splicing)          | PS3, PM2, PP3, PP5               | Likely Pathogenic (SP.)       | 5                  |
| c.8027T>C                | VUS                        | PP3, BS3                         | VUS                           | 3                  |
| c.8027T>A                | VUS                        | PM2, PP3, BS3                    | VUS                           | 1                  |
| c.8035G>T                | Causal (splicing)          | PS3,PM2,PP3                      | Likely Pathogenic (SP.)       | 2                  |
| c.8039A>G                | VUS                        | PM2, PP3, BS3                    | VUS                           | 2                  |
| c.8042C>G                | VUS                        | PM2, PP3, BS3                    | VUS                           | 1                  |
| c.8072C>T                | VUS                        | PM2, PP3                         | VUS                           | 4                  |
| c.8084C>T                | VUS                        | PM2, PP3, BS3                    | VUS                           | 4                  |
| c.8111C>T                | VUS                        | PP3, BS3                         | VUS                           | 5                  |
| c.8165C>G                | VUS                        | PM2, PP3, BS3                    | VUS                           | 6                  |
| c.8168A>G                | Causal (missense-splicing) | PS3 (Mis.), PM2, PP3, PP5        | Likely Pathogenic (Mis.-SP.)  | 13                 |
| c.8191C>T                | Causal (nonsense)          | PVS1 (Non.), PM2, PP3            | Pathogenic (Non.)             | 3                  |

|                |                     |                      |                         |   |
|----------------|---------------------|----------------------|-------------------------|---|
| c.8219insT     | Causal (frameshift) | PVS1 (Fr.), PM2, PP3 | Pathogenic (Fr.)        | 1 |
| c.8249_8250del | Causal (frameshift) | PVS1 (Fr.), PM2, PP3 | Pathogenic (Fr.-SP.)    | 6 |
| c.8331G>A      | VUS                 | PS3, PM2, PP3, PP5   | Likely Pathogenic (SP.) | 1 |
| c.8331+1G>T    | Causal (splicing)   | PVS1 (SP.), PS3, PM2 | Pathogenic (SP.)        | 1 |
| c.8331+2T>C    | Causal (splicing)   | PVS1 (SP.), PS3, PM2 | Pathogenic (SP.)        | 1 |

<sup>1</sup> Variants with impact on splicing are yellow-shadowed

<sup>2</sup> ACMG criteria used in this study:

**Very strong evidence of pathogenicity**

**PVS1:** null variant (nonsense, frameshift, canonical  $\pm 1$  or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease. When this criterion is applied, the type of mutation is indicated between brackets: SP, splicing; Non., nonsense; Fr, frameshift; Mis., missense.

**Strong evidence of pathogenicity**

**PS3:** Well-established in vitro or in vivo functional studies (splicing assays) supportive of a damaging effect on the gene or gene product.

**Moderate evidence of pathogenicity**

**PM2:** Absent from controls. All variants were checked in the Exome Aggregation Consortium database (exome data from 60,706 unrelated individuals; <http://exac.broadinstitute.org/>)

**PM4:** Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants (this item has been used for variants that induce in-frame transcripts such as ex18-ins6)

**Supporting evidence of pathogenicity**

**PP3:** Multiple lines of computational evidence support a deleterious effect on the gene or gene product. In this study, all variants were checked with NNSplice and Human Splicing Finder 3.0 that contains several splicing algorithm (see also S2 Table).

**PP5:** Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation (only used for missense, nonsense and frameshift variants).

**Strong evidence of benign impact**

**BS3:** Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing

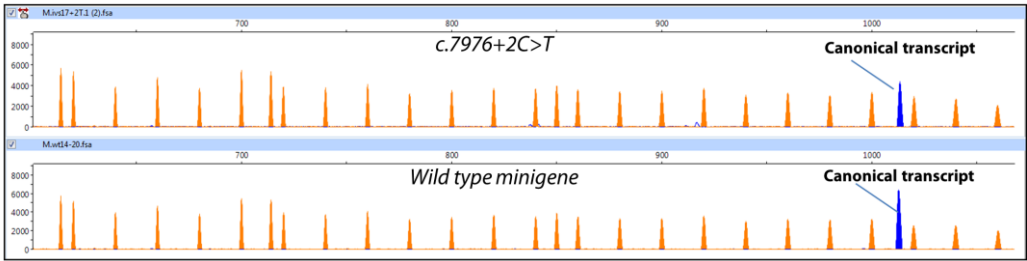
**Reference: “Guidelines of the American College of Medical Genetics and Genomics (ACMG) “**

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. ; 2015;17: 405–423. doi:10.1038/gim.2015.30

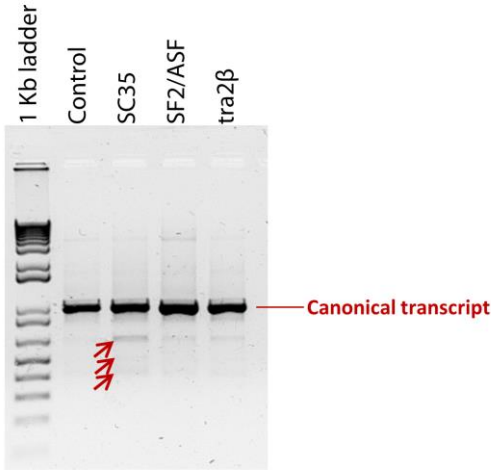
**S5 Table.** Distribution of pathogenic/likely pathogenic variants by type according to the BRCA Share database and this study.

|                  | Pathogenic variants |         | Total (%)  |
|------------------|---------------------|---------|------------|
| Type of mutation | Exon 17             | Exon 18 |            |
| Frameshift       | 5                   | 24      | 29 (44.6%) |
| Nonsense         | 2                   | 8       | 10 (15.4%) |
| Missense         | 1                   | 5       | 6 (9.2%)   |
| Splicing         | 8                   | 10      | 20 (30.8%) |
| TOTAL            | 16                  | 47      | 65         |

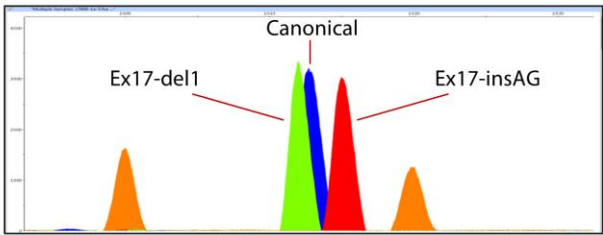
S1 Figure



S2 Figure



S3 Figure





## ARTÍCULO 2

Eugenia Fraile-Bethencourt, Alberto Valenzuela-Palomo, Beatriz Díez-Gómez, Mar Infante, Mercedes Durán, Germán Marcos, Enrique Lastra, Susana Gómez-Barrero, Eladio A. Velasco.

### **Genetic dissection of the *BRCA2* promoter and transcriptional impact of DNA variants.**

Breast Cancer Research and Treatment 2018; 171:53–63

<https://doi.org/10.1007/s10549-018-4826-7>

*Impact factor 2017: 3.605*

**Introducción:** Las mutaciones en el promotor pueden afectar a la transcripción génica y estar asociadas a la aparición de enfermedades humanas. Sin embargo, no es común incluir el rastreo del promotor en los estudios genéticos de los genes asociados a cáncer de mama. Nuestro objetivo se centra en el estudio del promotor de *BRCA2* para determinar una posible correlación entre la alteración de la transcripción y el cáncer de mama.

**Materiales y métodos:** Se secuenciaron el núcleo del promotor y el promotor proximal de *BRCA2* en 95 pacientes de cáncer de mama de alto riesgo. Se generó un inserto del promotor de *BRCA2* (desde la posición -938 hasta la +312 desde el sitio de inicio de la transcripción, TSS) que fue clonado en el vector de luciferasa pGL4.10. Las variantes y deleciones del promotor fueron introducidas en el constructo mediante mutagénesis dirigida. Los ensayos funcionales se realizaron en células MCF-7 y el resultado fue cuantificado mediante ensayos *Dual-Luciferase* y RT-PCR semi-cuantitativa.

**Resultados:** Se detectaron tres variantes en pacientes de alto riesgo de cáncer de mama: rs3092989 (G>A), rs206118 (A>G) y rs563971900 (C>T). El mapeo funcional a través de 13 microdeleciones solapantes mostró cuatro segmentos que reducían significativamente la tasa de transcripción (posiciones desde el TSS): -59\_-10del/μdel3 (16% de actividad con respecto al wt), -104\_-55del/μdel4 (62%), -239\_-190del/μdel7 (39%) y -464\_-415/μdel12 (78%); sugiriendo la presencia de elementos activadores de la transcripción en dichas regiones. Además, seis microdeleciones provocaron un aumento significativo en la expresión de luciferasa: +32\_+81del/μdel1 (356%), -14\_+36del/μdel2 (180%), -194\_-145del/μdel6

(154%), -284\_-235del/  $\mu$ del8 (168%), -329\_-280del/ $\mu$ del9 (111%) y -509\_-460del/ $\mu$ del13 (139%); lo cual indicaría la presencia de elementos represores. El ensayo funcional de 15 variantes de promotor (incluidas las encontradas en pacientes) mostró que 10 de ellas alteraban significativamente la expresión, siete aumentándola (113-163%) y tres disminuyéndola (72-83%). Ocho de las variantes estaban localizadas en el *cluster* 121 de hipersensibilidad a DNasa (TSS -185 a +105), donde se encuentran la mayoría de los elementos activadores de la transcripción.

**Conclusiones:** La expresión de *BRCA2* es altamente sensible a variaciones en el promotor, pues la mayoría de las variantes ensayadas inducía cambios relevantes. Además, hemos mapeado regiones críticas en la secuencia del promotor de *BRCA2* que podrían constituir dianas para variantes reguladoras. Tres variantes disminuyeron moderadamente la actividad transcripcional del promotor, sin embargo, para poder confirmar su patogenicidad son necesarios más análisis. En conjunto, nuestros datos refuerzan la necesidad de incluir el estudio del promotor en el rastreo de los genes de cáncer de mama con el objetivo de detectar nuevas variantes patogénicas.



# Genetic dissection of the BRCA2 promoter and transcriptional impact of DNA variants

Eugenia Fraile-Bethencourt<sup>1</sup> · Alberto Valenzuela-Palomo<sup>1</sup> · Beatriz Díez-Gómez<sup>1</sup> · Mar Infante<sup>2</sup> · Mercedes Durán<sup>2</sup> · Germán Marcos<sup>3</sup> · Enrique Lastra<sup>4</sup> · Susana Gómez-Barrero<sup>5</sup> · Eladio A. Velasco<sup>1</sup>

Received: 2 November 2017 / Accepted: 16 April 2018 / Published online: 15 May 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

**Purpose** Promoter mutations may affect transcription and can be associated with human diseases. However, the promoters of the breast cancer (BC) genes are not regularly screened. Our goal was to investigate the *BRCA2* promoter in order to study a possible correlation between impaired transcription and disease.

**Methods** The proximal and core promoter of the *BRCA2* gene was sequenced in 95 high-risk BC patients. A *BRCA2*-promoter insert [− 938 to +312 from the transcription start site (TSS)] was generated and cloned into the firefly luciferase vector pGL4.10. Promoter variants and deletions were introduced by site-directed mutagenesis and quantified by Dual-Luciferase assays and semi-quantitative RT-PCR.

**Results** Three different variants were detected in high-risk BC patients: rs3092989, rs206118, and rs563971900. Functional mapping of 13 overlapping deletions revealed four down-regulating segments (TSS positions): −59\_−10del/μdel3 (16% of activity of the wild-type construct), −104\_−55del/μdel4 (62%), −239\_−190del/μdel7 (39%), −464\_−415/μdel12 (78%), suggesting the presence therein of putative transcriptional activator motifs. Additionally, six microdeletions rendered luciferase overexpression: +32\_+81del/μdel1 (356%), −14\_+36del/μdel2 (180%), −194\_−145del/μdel6 (154%), −284\_−235del/μdel8 (168%), −329\_−280del/μdel9 (111%), and −509\_−460del/μdel13 (139%), which is indicative of repressor elements. Functional assays of 15 promoter variants (including those detected in patients) showed that ten of them significantly altered expression with seven up-regulating (113–163%) and three down-regulating (rs551887850\_G, rs570548398\_T, rs55880202\_T; 72–83%) SNPs. Eight of them were located in an ENCODE-DNase Hypersensitive Cluster (TSS − 185 to +105) where most active transcriptional motifs are known to be placed.

**Conclusions** *BRCA2* expression is highly sensitive to promoter variations as most of them induced relevant changes. Moreover, we mapped critical regions of the *BRCA2* promoter that may constitute potential targets for regulatory variants. Three SNPs moderately decreased luciferase activity, but confirmation of its potential pathogenicity requires further analysis. These data reinforce the need to screen the promoter regions of breast cancer genes with a view to discovering novel deleterious mutations.

**Keywords** Breast cancer · Ovarian cancer · Susceptibility genes · BRCA2 · Transcription · Promoter · Regulatory mutations · Luciferase assays

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10549-018-4826-7>) contains supplementary material, which is available to authorized users.

✉ Eladio A. Velasco  
eavelsam@ibgm.uva.es

<sup>1</sup> Grupo de Splicing y Cáncer, Instituto de Biología y Genética Molecular (IBGM), Consejo Superior de Investigaciones Científicas (CSIC-UVA), Sanz y Forés 3, 47003 Valladolid, Spain

<sup>2</sup> Cancer Genetics, Instituto de Biología y Genética Molecular (UVA-CSIC), Valladolid, Spain

<sup>3</sup> Servicio de Oncología, Hospital Río Hortega, Valladolid, Spain

<sup>4</sup> Servicio de Oncología, Hospital Universitario de Burgos, Burgos, Spain

<sup>5</sup> VISAVET-Universidad Complutense de Madrid, Madrid, Spain

## Introduction

*BRCA1* (OMIM #113705) and *BRCA2* (OMIM #600185) encode proteins of the DNA repair pathway and constitute the two most prominent tumor suppressor genes linked to an inherited susceptibility to breast and ovarian cancer [1, 2]. Inactivating germline mutations in both genes significantly increase the lifetime risk of breast and ovarian cancer [3, 4] as well as other types of tumors [5]. Hereditary breast and ovarian cancer (HBOC) is characterized by a high genetic heterogeneity where more than 20 susceptibility genes have been discovered [2]. Deleterious mutations in *BRCA1* and *BRCA2* are found in ~16% of HBOC families [6]. Testing for *BRCA1* and *BRCA2* mutations in high-risk individuals has become routine in family cancer clinics and genetic counseling units. Therefore, BRCA mutation carriers can benefit from risk assessment, management strategies, and even specific targeted therapies such as Poly(ADP-ribose) polymerase (PARP) inhibitors [7, 8].

Apart from protein-disrupting mutations, sequence variations can also alter transcription initiation, pre-mRNA splicing, RNA polyadenylation, RNA structure, mRNA stability (3'UTR), or other gene expression steps [9–16]. Interestingly, nearly 90% of disease-associated SNPs are placed outside protein-coding regions, suggesting a relevant role of the non-coding sequence variations [17].

We herein focus on transcription initiation, specifically on the promoter sequence, which despite being a critical part of the gene is poorly studied. In fact, mutation screening of the BRCA promoters is not included in the daily routine of genetic testing laboratories. However, DNA variants at the promoter may substantially modify transcriptional activity and alter protein levels, which might be related with the disease [9, 18]. The core promoter is defined as the minimal sequence that is necessary for an accurate transcription initiation to take place. It is at this core promoter where general transcription factors (TF) recognize specific sequences (6–12pb) and recruit RNA polymerase II, which starts RNA synthesis at the transcription start site (TSS). The size of the core promoter varies but usually spans from –40 to +50 nucleotides from the TSS [19]. Altogether, transcription is a highly complex process where dysregulation can lead to disease. In fact, 1.1% of disease-causing SNPs occur in the promoter altering transcription initiation or decreasing mRNA levels [20]. For instance,  $\beta$ -thalassemia may be caused by one SNP in the *HBB* promoter (–28A→C) which modifies the essential TATA-box [21].

*BRCA2* core promoter is located from position –66 to +129 [22]. It is, like most of human promoters, rich in CG and does not contain a TATA-box near the TSS. However,

this region contains some relevant TF binding sites, such as E-box (TSS –18), Ets/E2F (TSS –58), and SP1 (TSS –6) [22]. We intended to investigate the role of DNA variants of the *BRCA2* promoter in breast cancer susceptibility. For this purpose, we sequenced the *BRCA2* promoter of 95 high-risk BC patients with previous negative BRCA test. Then, we functionally studied this promoter in the pGL4.10 vector by luciferase reporter assays and tested several SNPs to know at what extent they influence the transcription rate. Moreover, we performed microdeletions along the promoter in order to map the essential regulatory regions.

## Materials and methods

### Patients

Ninety-five high-risk BC patients fulfilled the following criteria: three or most BC cases in the same family (87 patients) and breast + ovarian cancers in the same family or patient (8 cases). DNA was purified from peripheral blood lymphocytes by standard procedures. The complete coding and flanking intronic sequences of *BRCA1* and *BRCA2* were screened by heteroduplex analysis with capillary array electrophoresis [23] and/or by multiplex ligation-dependent probe amplification [24]. No pathogenic BRCA mutations were detected in either of the selected BC patients. Ethical approval for this study was obtained from the Ethics Review Committee of the Hospital Universitario Río Hortega (Valadolid, 6/11/2014). All patients signed the written informed consent of the breast/ovarian cancer prevention program from Castilla y León (Spain). The methods were carried out in accordance with relevant guidelines and regulations.

In order to identify genetic variations in the proximal and core promoter (–500 to +50) [9], about 100 ng of template DNA of each patient was amplified with primers Pr\_BR2-FW5 (5'-ATGAGCGAGTTAAGATGGGT-3') and Pr\_BR2-RV5 (5'-CACCACCACTAACCCGCTC-3') (size: 772 bp) and Phusion High-Fidelity polymerase according to the manufacturer's instructions. PCR products were purified with the AccuPrep® PCR purification kit (Bioneer, Daejeon, Korea) and then sequenced with BigDye Terminators v3.1 (Life Technologies, Carlsbad, CA) and with the primers above mentioned. To identify DNA variants, patient and the reference promoter sequences of *BRCA2* (ENST00000380152.7) were aligned with Clustal Omega [25].

### Cloning of the *BRCA2* promoter

An insert of the *BRCA2* promoter (1250 bp; TSS –938 to +312) was amplified with Phusion High-Fidelity Polymerase (Thermo Fisher Scientific, Waltham, MA, USA) and

primers 5'-CTAACTGGCCGGTACCTGAGCTCGCTAGCCAGGTGATCCGAATCCTAAGAA-3' and 5'-GCCGCCGAGGCCAGATCTTGATATCCTCGACAGAGACAAAAGGGCAAGAAG-3' which contained tails (underlined) for cloning into pGL4.10 (Promega, Madison, WI). The PCR product (megaprimer) was gel purified and then inserted into pGL4.10 by overlapping extension PCR [26] in a final volume of 50 µl with 850 ng of megaprimer (previously denatured at 98 °C for 5 min), 100 ng of pGL4.10, 4% DMSO, and Phusion Hot Start Polymerase under these PCR conditions: 98 °C, 30 s, +35 cycles x [98 °C, 10 s/60 °C, 15 s/72 °C, 3 min], and 72 °C, 10 min. Colonies were screened by PCR, and recombinant clones were confirmed by sequencing at MacroGen (Seoul, Korea).

Mutations were introduced with the QuikChange Lighting kit (Agilent, Santa Clara, CA). The wild-type (wt) construct pGL4.10-BR2 was used as template to generate 15 promoter SNPs (Table 1) and 13 50-nt overlapping microdeletions (Supplemental Table S1).

### Luciferase assays

Approximately  $2 \times 10^5$  MCF7 cells were grown to 90% confluency in 0.5 mL of medium (MEME, 10% Fetal Bovine Serum, 2 mM glutamine, 1% Non-essential amino acids, and 1% Penicillin/Streptomycin) in 4-well plates (Nunc, Roskilde, Denmark). Cells were transiently transfected using Lipofectamine LTX (Life Technologies), with 500 ng of the pGL4.10-constructs plus 25 ng of the control Renilla plasmid (pGL4.74; Promega) to normalize for transfection efficiency. Twenty-four hours after transfection, luminescence was measured with the Dual-Luciferase® Reporter Assay System (Promega) in a microplate luminometer (Berthold, Germany). Luciferase activities are representative of at least three independent transfection experiments each at least in triplicate. Student's *t* test was used to determine statistical significance between mutant and control constructs.

### Results and discussion

To identify potential harmful variants, a fragment of the *BRCA2* promoter (TSS – 565 to + 207) was sequenced in a cohort of 95 BRCA negative high-risk breast cancer patients. Three different changes were identified: rs3092989 (G>A) and rs206118 (c.-52A>G) and one rare variant (Minor Allele Frequency [MAF] < 0.01), rs563971900 (C>T). According to the 1000 Genomes project ([http://www.ensembl.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000139618;r=13:32315474-32400266](http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000139618;r=13:32315474-32400266)), rs3092989 and rs206118 are common variants with MAFs of 0.16 and 0.149, respectively. By contrary, allele T of rs563971900 was found only in one out of 190 chromosomes (0.005)

while in the 1000 genomes project was reported in 2/5008 chromosomes (MAF=0.0004) confirming that it is indeed a rare variant. Moreover, allele T was not found in the control Spanish population (0/214 chromosomes) used in the 1000 genomes project (<https://goo.gl/12Xg4S>). SNPs rs3092989 and rs563971900 (TSS – 254G>A and – 63C>T, respectively) are located in an important region of the promoter that might impair transcription.

### Functional study of the *BRCA2* promoter

To study the potential impact of the *BRCA2* promoter variants on transcriptional activity, we developed a luminescence assay. The *BRCA2* promoter was cloned into the pGL4.10 vector (Fig. 1) and was functionally validated by a luciferase assay.

We then generated thirteen 50-bp microdeletions (µdel) to map the presence of regulatory sequences of the promoter, spanning from – 509 to + 81 from the TSS (Fig. 1 and Supp. Fig. S1), which covered the segment where most transcriptional mutations are located (TSS – 500 to + 50) [9]. Their transcriptional impacts were subsequently tested by Dual-Luciferase assays (Fig. 2; ≥ 9 quantifications). Microdeletions 5, 10, and 11 did not show remarkable changes in transcriptional activity since they showed similar luminescence to the wild-type construct. Thus, it can be concluded that these µdels may not include essential sequences to promote the transcription of the *BRCA2* gene. On the other hand, µdels 1, 2, 6, 8, 9, and 13 were significantly associated with luciferase overexpression of 356.3, 179.7, 153.6, 168.3, 111.5, and 138.8%, respectively, with respect to the wild-type activity, suggesting the presence of binding motifs for transcriptional repressors, which, once deleted, do enhance transcription (Supplemental Table S1). Conversely, µdels 3, 4, 7, and 12 clearly showed a reduced transcriptional activity (16, 62, 39, and 78%, respectively;  $p < 0.000001$ ), which might reflect the loss of functional motifs for transcriptional activators. Indeed, microdeletion 3, with more than 80% reduction of luminescence, might represent a null allele. Microdeletion 3 spans from nt – 59 to – 10 from the TSS, where the core promoter (– 40 to + 50) is located, which contains binding sites for general transcription factors, such as TFIID, TFIIA, and TFIIB, and RNA polymerase II [9]. In summary, this strategy is useful to map DNA segments at the promoter where essential binding motifs for either repressors or activators are placed and functionally active. Fine mapping of these regulatory elements could be achieved by complementary internal deletions.

### Functional characterization of promoter variants

We then proceeded to assay the DNA variants detected in patients (rs3092989\_A, rs563971900\_T, and

**Table 1** Allelic frequencies, bioinformatics analysis, and mutagenesis primers of *BRCA2* promoter DNA variants

| SNP_allele <sup>a</sup> | Allele frequencies <sup>b</sup> | Position (TSS) | Bioinformatics analysis <sup>c</sup>   |   | Primers: 5'→3' sequences  |
|-------------------------|---------------------------------|----------------|--|---|---|
|                         |                                 |                | PROMO  | JASPAR  |   |
| rs545887954_T           | C: 0.9998<br>T: 0.0002          | -492C>T        | T: [-]RXRalpha (8.66), NF1/CTF (7.59), NFY (3.05)  | No changes  | AGGTCTTGAACACCTGCTACTCAATAG<br>CAGAACACAGCTACTG<br>CAGTAGCTGTCTGCTATTGAGTAGCA<br>GGTGTCAAGACCT  |
| rs541454524_G           | T: 0.9986<br>G: 0.0014          | -467T>G        | G: [+] TFII-I; [+] STAT4 (1.47 (T) - 4.41 (G))   | No changes  | AGCAGAACAGCTACTGGAACGAAA<br>ATCCTCTGATTTCAAAT<br>ATTGAAATCAGAGGATTTTCGTTCCA<br>GTAGCTGTCTGCT  |
| rs148060832_A           | G: 0.9998<br>A: 0.0002          | -262G>A        | No changes   | No changes  | CCACCCAAACATGAGCTGAAGCA<br>AAAAGAAAGGGATGG<br>CCATCCCTTCTTTTGTCTTCAGCTCA<br>TGTTGGGTGGG   |
| rs3092989_A             | G: 0.840<br>A: 0.160            | -248 G>A       | A: [-] GR-alpha (0.21), TFII-I (9.51)  | A: [-] MZF1(var.2) (9.84)   | GAGCTGGAGCAAAAAGAAAGAGAT<br>GGGGACTTGGAGTAGG  |
| rs36221751_A            | G: 0.9982<br>A: 0.0018          | -174G>A        | No changes   | A: NRL (9.58→10.93)   | CCTACTCCAAGTCCCCCATCTCTTTCT<br>TTTGTCCAGCTC<br>ACTCTAAGGGTCAGCGAAAAGAG<br>AACACACACTCCAG  |
| rs565737252_A           | T: 0.9996<br>A: 0.0004          | -159T>A        | A: [-] TFII-I (6.58)   | A: [-] NKX2-8 (9.083)   | CTGGAGTGTGTCTCTCTTTTCGCTGA<br>CCCTTAAGAGT<br>GCGAGAGAGAACACACACACCAG<br>CTCCCGCTTATTCG  |
| rs533226610_A           | G: 0.9996<br>A: 0.0004          | -120G>A        | A: [-] NF-1 (7.45), FOXF3 (9.51) [+/-] GR-beta (1.68), TFII (4.76), STAT4 (4.41), c-Ets-1 (9.19) | A: [-] Hic1 (9.55); [-] NFIX (6.37), NFIC (7.59), TEAD1 (10.61), TEAD4 (10.98), TEAD3 (12.53) | CGAATAAAGCGGAGCTGGTGTGT<br>GTGTCTCTCTCGC<br>GGTCAGATACTGACGGTTGGAATGCCT<br>GACAAAGGAATTTCC  |
| rs551887850_G           | A: 0.9998<br>G: 0.0002          | -119A>G        | G: [-] XBP-1 (4.89); [+] RXRalpha (5.94)   | G: [-] Hic1 (9.55)  | GGAAATTCCTTGTACGGCATTCACACC<br>GTCAGTATCTGACC<br>TCAGATACTGACGGTTGGGGTGCCTGA<br>CAAGGAAATTC   |
| rs11571571_C            | C: 0.9998<br>T: 0.0002          | -94 T>C        | C: [-] TFIIID (9.55); [+] MEF2A (7.98)   | C: [-] SOX10 (6.81)   | GGAAATTCCTTGTACGGCACCCCA<br>ACCGTCAGTATCTGA<br>GGAAATTCCTTTCGCCACACCGAGAAA<br>TACCCGACGGGCC<br>GGCCGCTGCGGCTATTTCTCGGTGTGG<br>CGAAAGGAAATTC |

**Table 1** (continued)

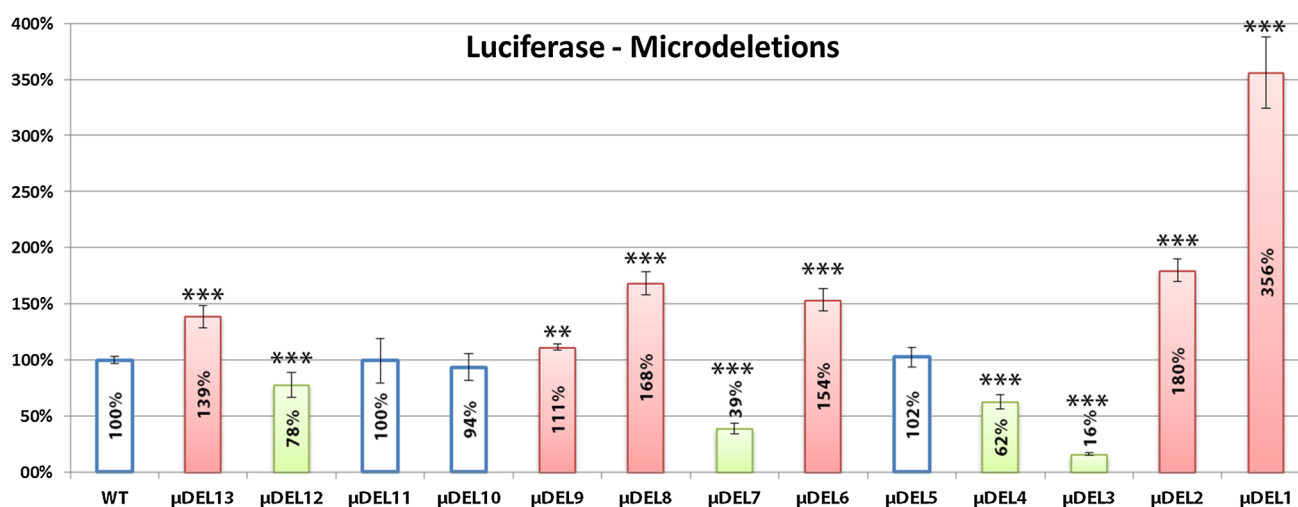
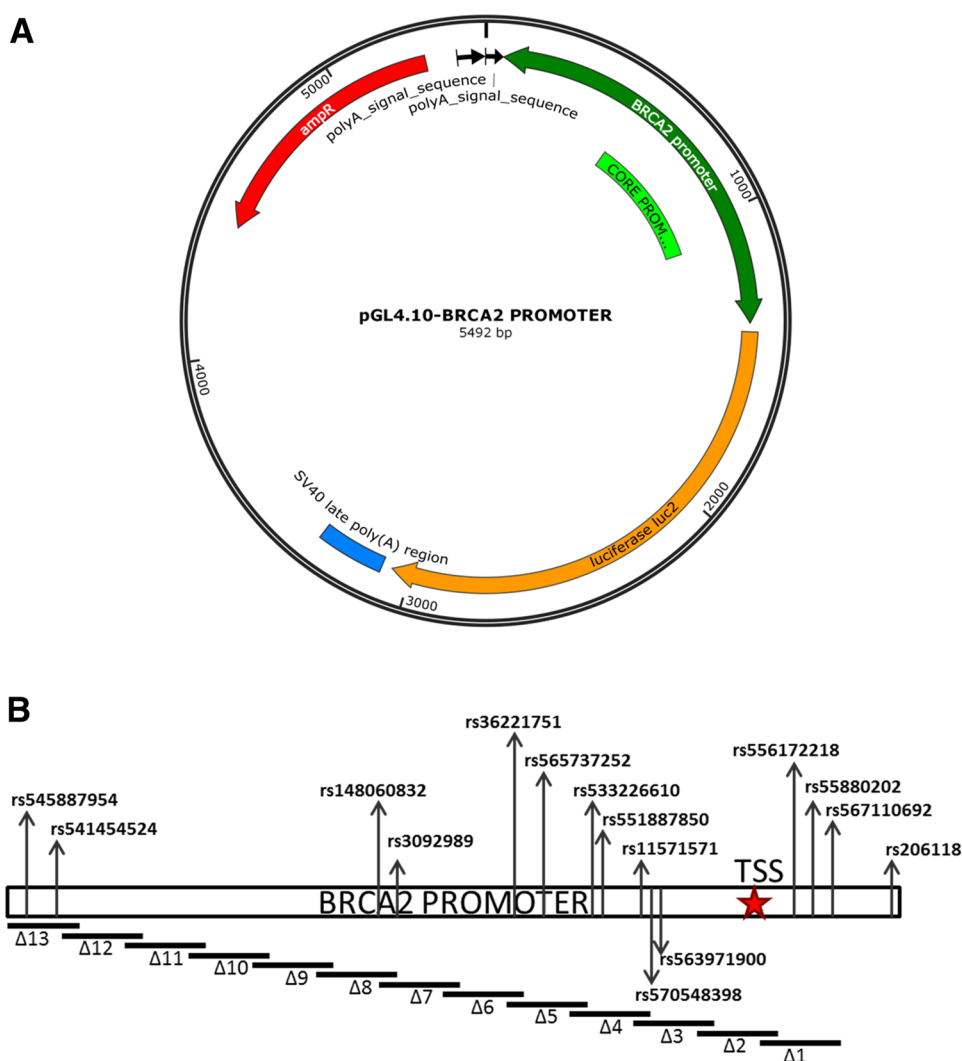
| SNP_allele <sup>a</sup>             | Allele frequencies <sup>b</sup> | Position (TSS) | Bioinformatics analysis <sup>c</sup>                                     |  | Primers: 5'→3' sequences   |
|-------------------------------------|---------------------------------|----------------|--|--|--|
|                                     |                                 |                | PROMO  | JASPAR   |  |
| <u>rs570548398_T</u>                | C: 0.9998<br>T: 0.0002          | -77C>T         | T: [-] RXRalpha(6.56); [+]<br>(8.07)                                     | T: [-] GCM1 (10.93), GCM2 (10.85); [+]<br>TEAD3 (7.61), SNAI2 (8.96) | TCGCCACACTGAGAAATACCTGCA<br>GCGGCCACCCAGGCTT<br>AGGCTGGTGGCCGCTGCAGGT<br>ATTCTCAGTGTGGCGA    |
| <u>rs563971900_T</u>                | C: 0.9996<br>T: 0.0004          | -63C>T         | T: [-] RXRalpha (6.56); [+]<br>(8.07); [↑] AP2alphaA (0.23 (C)-4.44 (T)) | T: [-] Klf4 (10.81)  | AATACCCGACGCGGCCACCTAGG<br>CCTGACTTCCGGGTGGT<br>ACCACCCGAAATCAGGCCTAGGT<br>GGGCCGCTGCGGTATT  |
| <u>rs556172218_T</u><br>(c.-188C>T) | C: 0.999<br>T: 0.001            | +46C>T         | No changes   | No changes   | TAGGCGCAGAGCGGAGCTGCTG<br>TGGCACTGCTGCGCC<br>GGCGCAGCAGTGCCACACAGCAGCTC<br>CGCCTCTGCCGCCTA   |
| <u>rs55880202_T</u><br>(c.-175C>T)  | C: 0.994<br>T: 0.006            | +59C>T         | T: [-] GCF (8.26); [+]<br>(9.51), PR B, PR A                             | No changes   | CGGAGCGCTGTGGCACTGTTGCG<br>CCTTCTGCTGGCCTC<br>GAGGCGCAGCAGAGGCGCAACAGT<br>GCCACAGCGGCTCCG    |
| <u>rs567110692_A</u><br>(c.-162G>A) | G: 0.998<br>A: 0.002            | +72G>A         | A: [-] GCF (8.26)  | No changes   | GCACTGCTGCGCCTCTGCTACGCCTCG<br>GGTGTCTTTTGC<br>GCAAAAGACACCCGAGGCGCTAGCA<br>GAGGCGCAGCAGTGC  |
| <u>rs206118_A</u><br>(c.-52A>G)     | A: 0.851<br>G: 0.149            | +182A>G        | A: [-] GCF (7.19), AP2alphaA (3.23)                                      | No changes   | TCCGGCCAAAAAAGAACTGCACCT<br>CTGGAGCGGGTTAGTGG<br>CCACTAACCCGCTCCAGAGGTGCAGTT<br>CTTTTGGCCCGA |

<sup>a</sup>Alleles detected in patients are underlined, rs563971900 allele T was detected in one out of 190 chromosomes. Bold and italics variants indicate up- or down-regulation, respectively

<sup>b</sup>Allelic frequencies were taken from the 1000 genomes project at the Ensembl browser (2504 genomes, 5008 chromosomes)

<sup>c</sup>Putative binding sites of transcription factors: [+]  
Creation or [-]  
disruption of a binding motif; [↑]  
increase of a motif score. ALGGEN/PROMO: AP2alphaA transcription factor AP-2 alpha (activating enhancer-binding protein 2 alpha), C/EBPbeta CCAAT-enhancer-binding protein beta, c-*Ets-1* ETS proto-oncogene 1, transcription factor, E2F1 E2F transcription factor (retinoblastoma-associated protein 1), FOXO3 forkhead box P3, GCF, GCF2 GC-rich sequence DNA-binding factor, GR-alpha glucocorticoid receptor (transcriptional activation and repression activity) isoform alpha, GR-beta glucocorticoid receptor isoform beta (repressor of alpha), MEF2A myocyte enhancer factor 2A, NF-1 nuclear factor 1, NF1/CTF CAAT box transcription factor, NFY nuclear transcription factor Y, RXRalpha retinoid X receptor alpha, STAT4 signal transducer and activator of transcription 4, TFIIID transcription factor II D (make up the RNA polymerase II preinitiation complex), TFII-I general transcription factor II-1 (activator of BRCA1), TFIIH transcription factor III (binds to TNA pol III), PR A and PR B progesterone receptor (encoded by the same gene), XBP-1 X-box binding protein. JASPAR: GCM1 glial cells missing homolog 1, GCM2 glial cells missing homolog 2, Hic-1 hypermethylated in cancer 1 protein (mouse), Klf4 Kruppel-like factor 4 (mouse), KLF5 Kruppel-like factor 5, MZF1 myeloid zinc finger 1, NFIC nuclear factor 1 C-type, NFIX nuclear factor 1 X-type, NKX2-8 NK2 homeobox 8, NREL neural retina-specific leucine zipper protein, SOX10 SRY-box 10, SNAI2 Snail family transcriptional repressor 2, SP1 Sp1 transcription factor (binds to GC-rich motifs of many promoters), SP2 Sp2 transcription factor (binds to GC box promoters elements and selectively activates mRNA synthesis from genes that contain functional recognition sites), TEAD1 TEA domain transcription factor 1, TEAD3 TEA domain transcription factor 3, TEAD4 TEA domain transcription factor

**Fig. 1** Structure of the pGL4.10-prBR2 clone (a) and map of SNPs and microdeletions (b). The construct was drawn with the program SnapGene Viewer version 3.3.3



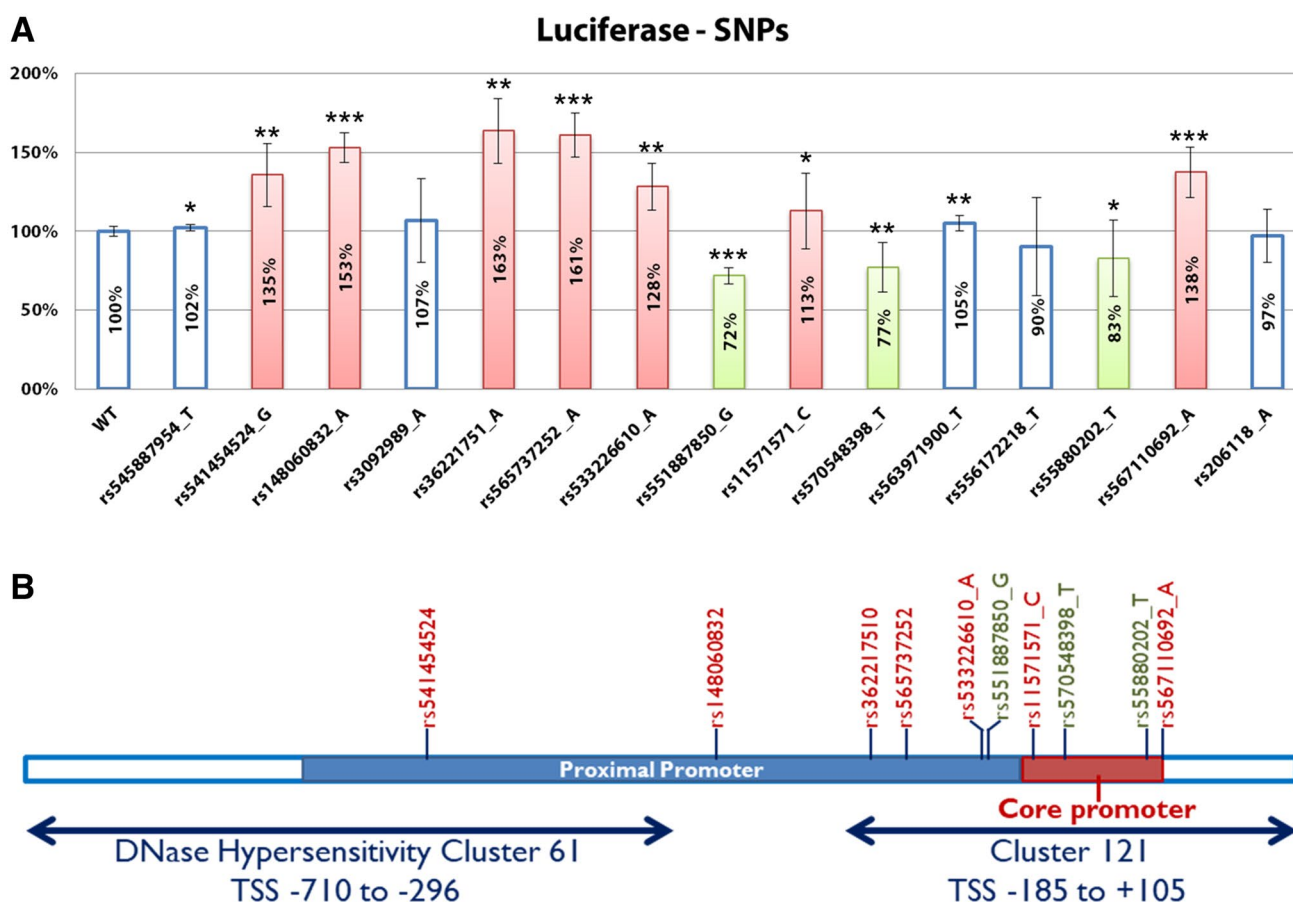
**Fig. 2** Mapping of regulatory sequences of the BRCA2 promoter by luciferase assays of internal deletions. White bars indicate no effect, red bars indicate overexpression ( $\geq 110\%$ ) and green bars indicate

underexpression ( $\leq 90\%$ ). Statistical significance of luciferase assays: \* $p < 0.05$ ; \*\* $p < 0.0001$ ; \*\*\* $p < 0.000001$

rs206118\_A), as well as 12 Ensembl rare variants (MAF < 0.01) that mapped between -500 and +75 from the TSS: rs545887954\_T, rs541454524\_G, rs148060832\_A, rs36221751\_A, rs565737252\_A, rs533226610\_A, rs551887850\_G, rs11571571\_C, rs570548398\_T, rs556172218\_T, rs55880202\_T, and rs567110692\_A. All these constructs and the wt plasmid pGL4.10-prBR2 were transfected into MCF-7 cells (Fig. 3a). The three patients' variants (rs3092989\_A, rs563971900\_T, and rs206118\_A) together with another two reported SNPs (rs545887954\_T and rs556172218\_T) did not display any important changes ( $\pm 10\%$ ) with respect to the wild-type promoter. We therefore concluded that these sequence variations did not affect *BRCA2* transcription and may not be associated with breast cancer risk, at least at this gene expression step.

However, we found that seven rare variants (rs541454524\_G, rs148060832\_A, rs36221751\_A, rs565737252\_A, rs533226610\_A, rs11571571\_C, and rs567110692\_A; Fig. 3a) significantly up-regulated

luciferase expression (from 113 to 163%;  $p < 0.05$ ,  $p < 0.0001$ ,  $p < 0.000001$ ), suggesting that they might be associated either with the disruption of a transcriptional repressor motif or with the creation of an activator motif. Finally, three SNPs (rs551887850\_G, rs570548398\_T, and rs55880202\_T; green bars in Fig. 3a) significantly down-regulated expression ( $-17$  to  $-27\%$ ;  $p < 0.000001$ ;  $p < 0.0001$  and  $p < 0.05$ , respectively), so that these nucleotide changes might either disrupt a transcriptional activator motif or create a repressor one. In a recently accepted manuscript of transcriptional effect of promoter variants [27], rs206118 and rs563971900 were also tested by luciferase assays in MDA-MB231 cells. Remarkably, similar to our results, rs206118 did not change promoter activity (97%), whereas rs563971900 showed a slight decrease with respect to the wild-type construct (72 vs. 105% observed in our assay), which might be due to the use of different host cells or other factors described below.



**Fig. 3** Impact of 15 *BRCA2* promoter variants on transcriptional activity. **a** Quantification of promoter activity by Dual-Luciferase assays. Statistical significance: \* $p < 0.05$ ; \*\* $p < 0.0001$ ; \*\*\* $p < 0.000001$ . White bars indicate no effect, red bars indicate overexpression ( $\geq 110\%$ ) while green bars indicate underexpression

( $\leq 90\%$ ). **b** Map of regulatory SNPs and DNase hypersensitivity clusters identified by the ENCODE project at the UCSC genome browser human GRCh37/hg19 Assembly (<https://goo.gl/DFH2y3>). Red and green variants indicate up- and down-regulating SNPs, respectively

Transcriptional activity is correlated with less condensed chromatin regions that are hypersensitive to cleavage by DNase I. Interestingly, 9 up- or down-regulated SNPs were placed in two DNase I Hypersensitivity clusters mapped by the ENCODE (Encyclopedia of DNA Elements) project (Table 1; Fig. 3b) [28]. Furthermore, five up-regulating and the three down-regulating SNPs were within cluster 121 (TSS – 185 to + 105) that constitutes a critical promoter segment where most TF binding sites are located. Consequently, any nucleotide change within this interval is prone to dysregulate transcription. These data suggest that these SNPs might disrupt repressor or activator motifs, although none of them altered TF binding sites mapped by CHIP-seq (ENCODE).

Bioinformatics analysis with PROMO and Jaspar programs [29, 30] did not reveal a common pattern between either overexpressing or underexpressing genetic variants (Table 1), and, in general, predictions showed low accuracy even though the maximum dissimilarity rate was reduced to 10%. Conclusively, the implication of a specific TF in expression anomalies should be demonstrated by DNA-TF-binding assays, such as the electrophoretic mobility shift assay (EMSA) [9].

To lend further support to these results, luciferase expression of the wt and mutant constructs was measured at the RNA level. All constructs were transfected again and RNA was isolated to perform a semi-quantitative PCR and fragment analysis in a DNA sequencer as previously described [31, 32] (Supplemental Methods; Supp. Fig. S2). Most variants (12/15) induced overexpression of the luciferase mRNA, while two of them reduced expression (rs3092989\_A and rs570548398\_T) and one (rs556172218\_T) induced transcript levels similar to the wild-type construct. These experiments replicated the results in nine variants: one underexpressing (rs570548398\_T), one wild-type-like (rs556172218\_T), and seven overexpressing SNPs (rs541454524\_G, rs148060832\_A, rs36221751\_A, rs565737252\_A, rs533226610\_A, rs11571571\_C, and rs567110692\_A). Both expression profiles of luciferase and RNA yields (Fig. 3a and Supp. Fig. 2) were very similar, supporting thus our results. Anyway, these results should be corroborated in patient RNA or even at the protein level whenever possible. The pGL4.10 constructs should only be considered as a method for the initial characterization of transcription anomalies. In fact, reporter activity of a particular single-nucleotide substitution can be affected by a wide range of additional factors, such as quality and concentration of the construct, the co-transfected control reporter, transfection method, number of constructs within the cell, growth history of host cells and also, importantly, the chromatin conformation of gene regulatory regions since it is likely to affect TF binding [33].

## Biological significance of *BRCA2* up- and down-regulation

According to the luminescence data, three variants (rs551887850\_G, rs570548398\_T, and rs55880202\_T) caused moderate decreases (17 to 27%) of the transcriptional activity with respect to the wild-type construct. Therefore, an important fraction of the canonical transcript is expressed, so that they should not be classified as null alleles. Previous studies on the mismatch repair genes (MMR) suggested that a 50% reduction in promoter activity of MLH1 and MSH2 might be an ideal cutoff to define abrogated allele-specific expression [34]. With regard to the BRCA genes, a recent report on the spliceogenic mutation c.[594-2A>C; 641A>G] proposed that any BRCA1 allele that retains 20–30% of its tumor suppressor function does not confer breast cancer risk [35]. Unfortunately, none of these three down-regulating SNPs were included in the most comprehensive breast cancer association study using OncoArray where 65 new risk loci were identified [36]. Besides, no data of these promoter SNPs were found at the NHGRI-EBI Catalog of published genome-wide association studies (GWAS) (<http://www.ebi.ac.uk/gwas/>, date last accessed June 2017), so that, nothing can be concluded regarding the breast cancer-associated risk of rs551887850\_G, rs570548398\_T, and rs55880202\_T. Only rs3092989 (107%), rs36221751 (163%), and rs206118 (97%) were included in the OncoArray study but they were not associated with breast cancer risk [36] (<http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/>). We also investigated *BRCA2*-eQTL (expression Quantitative Trait Loci) associations of the 15 assayed SNPs at the GTEx database (<https://www.gtexportal.org/>) [37]. Only SNPs rs3092989 and rs206118 were found within the GTEx datasets, where just rs206118 displayed a significant eQTL association in esophagus mucosa. Neither of these or any other variants showed *BRCA2*-eQTL in breast tissue or other disease targets (ovary, pancreas, and prostate). Other *BRCA2* SNPs (rs9562605, rs9567552, rs1799943, or rs11571583) did show eQTL associations in some tissues, such as thyroid or lung, but they are placed in introns 1 and 2 and exon 2, outside the region of interest of this study.

*Up-regulating SNPs.* One might speculate that, given the tumor suppressor nature of *BRCA2*, overexpressing SNPs most probably constitute protective alleles by decreasing disease-risk [38]. For example, the protective allele of SNP-rs1059111 of the neurofilament gene *NEFL* was correlated with *NEFL* up-regulation and a reduced risk to neuroblastoma [39]. Interestingly, these authors also showed that *NEFL* overexpression triggered cell differentiation and impaired proliferation. Besides, high levels of *NEFL* in primary neuroblastoma were associated with a better overall survival. Moreover, allele A of *BRCA2* variant c.–26G>A

(rs1799943) presumably creates a binding site for C/EBP alpha and enhances *BRCA2* expression by twice, providing a protective effect against breast cancer in the heterozygous genotype [40]. Likewise, the minor allele A of the *BRCA2* intronic SNP rs4942440 increased expression and was associated with a reduced risk of breast cancer [41]. These data would endorse a protective effect of the seven *BRCA2* overexpressing SNPs. However, they are rare variants (MAF << 0.01; Table 1) and GWAS studies of any of them have not been reported yet, so the relative risk of breast cancer cannot be estimated and consequently their potential protective effect cannot be confirmed.

On the contrary, it should be highlighted that overexpression of DNA repair genes was associated with metastatic progression of melanoma and other tumors (including breast cancer), therapeutic resistance and poor clinical outcome [42, 43]. This a priori apparent contradiction can be explained as follows. Though genetic instability is critical for the transformation of normal into tumor cells, paradoxically, a metastatic cell requires efficient DNA repair systems to resist chemotherapeutic agents as well as to have the capacity of invasiveness and growing in other tissues. This can be achieved through the overexpression of specific DNA repair genes, such as *RAD51*, which is a key factor for homologous recombination and DNA repair of double strands breaks. Remarkably, it has been found that *RAD51* expression is increased in a wide range of human tumors, most likely contributing to drug resistance [44]. Moreover, *BRCA2* overexpression was previously suggested to play a role in the aggressiveness of sporadic breast tumors [45]. Taken together, these *BRCA2* activating SNPs might display a double-sided effect: on one hand, they might protect against carcinogenesis and, on the other hand, an adverse consequence might be generated as up-regulation of DNA repair genes in metastatic cells confers chemoresistance increasing thus their survival.

## Concluding remarks

In the present study, we have shown that transcriptional activity is very sensitive to promoter variants since most of them (10/15 in luminescence assays) significantly modulated luciferase expression. Remarkably, eight of them mapped in cluster 121 of ENCODE-DNase hypersensitive regions from TSS – 185 to + 105 that is enriched in TF motifs. It is therefore reasonable to think that these changes may affect positively or negatively breast cancer risk. Functional mapping of the *BRCA2* promoter provided valuable data concerning the presence of TF binding sites, unveiling segments with either repressors ( $\mu$ del 1, 2, 6, 8, 9, and 13) or activators ( $\mu$ del 3, 4, 7, and 12). In fact, microdeletion 3 (from – 59 to – 10 from the TSS) absolutely abolished

luciferase expression ( $p < 10^{-25}$ ) indicating the presence of critical binding motifs. Indeed, it removes part of the core promoter (– 40 to + 50) that contains motifs for general TF and RNA polymerase II binding [9]. Also, microdeletion 7 (– 239\_– 190) significantly reduced transcription ( $> 60\%$ ,  $p < 10^{-17}$ ) suggesting the presence of activators within it. Therefore, both promoter sequences constitute targets where potential harmful variants can bring about transcription suppression, leading to null alleles and haploinsufficiency. Moreover, these down-regulating fragments should be scanned in patients with a view to searching for regulatory mutations.

We have only found three promoter variants in our cohort of patients probably due to the small size of the sample. None of these variants affected the transcriptional activity. However, according to Ensembl data, there exists a high density of SNPs in the *BRCA2* promoter that might regulate transcriptional activity. Actually, association studies with these SNPs in large BC patient cohorts will help to achieve a precise risk estimation of each of them. In this regard, it would be very interesting if massive sequencing projects of breast cancer patients included also the promoter regions of susceptibility genes. By so doing, very valuable data of promoter variants in thousands of samples would be generated and thus it would be possible to make precise risk estimations.

According to the luciferase data, most variants (10 out of 15) significantly affected transcription. None of the variants abolished transcription, since only a slight reduction of the transcription rate (23–27%) was confirmed for rs570548398\_T in both luciferase and semi-quantitative RT-PCR experiments. Another three variants only displayed lower luciferase expression in one of the assays so their transcriptional impact could not be proved. It is complex to define a null allele that should be gene-specific since, for example, 20–30% of transcript is enough to maintain *BRCA1* activity [35] while a 50% reduction of *MLH1* expression abrogates mismatch repair [34]. On the other hand, most alleles enhanced transcriptional activity but their impact on BC susceptibility as a protector allele or in other processes should be further investigated. Elucidating the roles of these up- and down-regulating SNPs will suppose an important step toward refining the knowledge on the genetic landscape of breast cancer.

Altogether these data support the need for the genetic screening and functional analyses of promoter mutations in hereditary cancers to improve their molecular diagnostics as well as to better understand the role of transcriptional regulation in these diseases. For this purpose, better bioinformatics tools, as well as more straightforward and high-throughput transcriptional assays to check variant impact, are required. Identification of null variants of the *BRCA2* or other BC susceptibility genes would broaden the number of patients and

families who may benefit from prevention and prophylactic protocols, or even from therapeutic measures such as the use of PARP inhibitors in BRCA-deficient tumors.

**Acknowledgements** We are grateful to the breast/ovarian cancer patients and clinicians who participated in this study. EAV's lab was supported by grants from the Spanish Ministry of Economy and Competitiveness, Plan Nacional de I + D + I 2013–2016, ISCIII (Fis: PI13/01749) co-funded by FEDER from Regional Development European Funds (European Union), and Grant CSI090U14 from the Consejería de Educación (ORDEN EDU/122/2014), Junta de Castilla y León. EFB was supported by a predoctoral fellowship from the University of Valladolid and Banco de Santander (2015–2019).

**Data availability** The datasets generated and/or analyzed during the current study are available in the Figshare repository, <https://figshare.com/s/e7c982e6afb9907d56a0>.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interests.

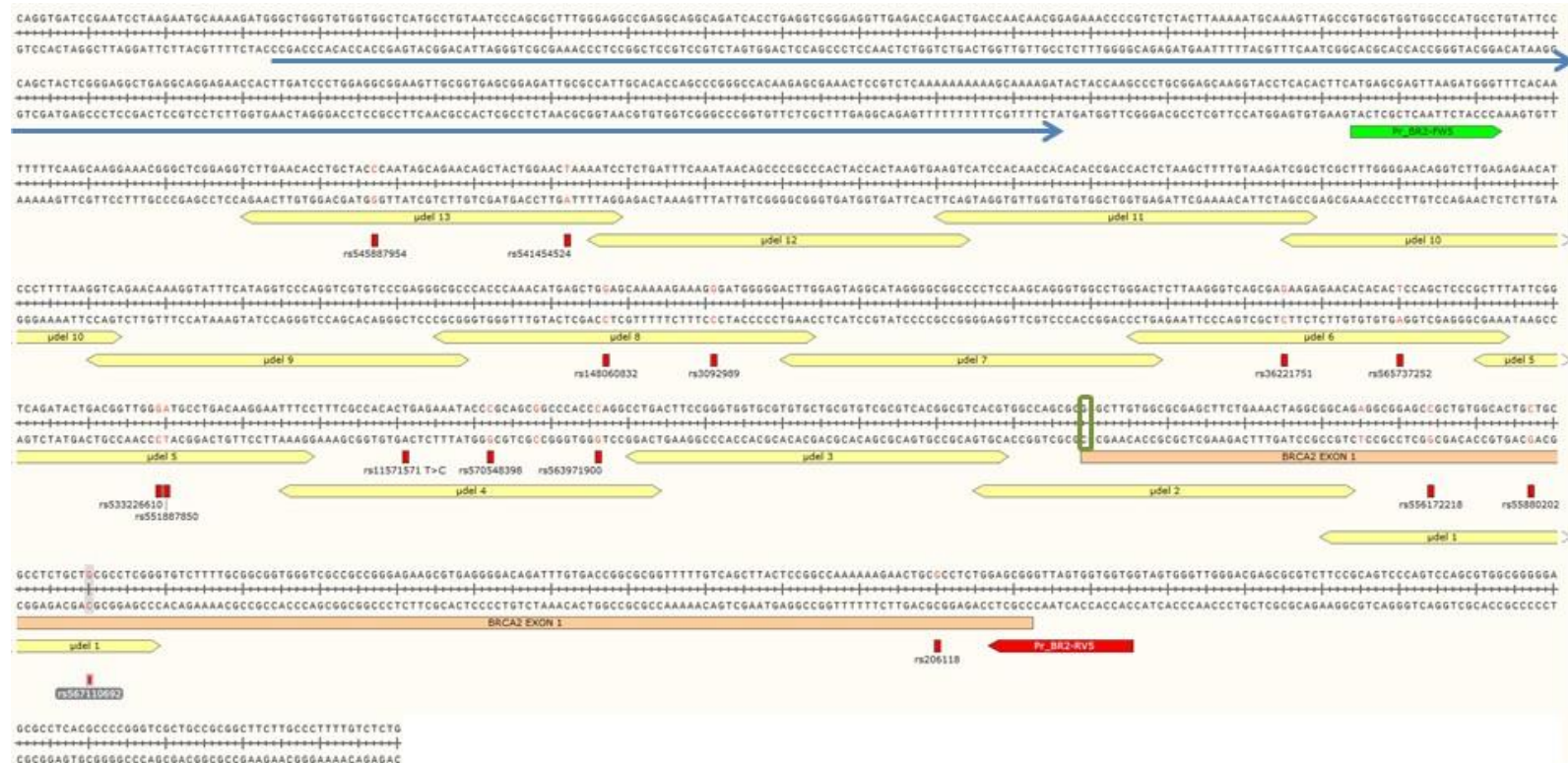
**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

- Roy R, Chun J, Powell SN (2012) BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer* 12:68–78. <https://doi.org/10.1038/nrc3181>
- Nielsen FC, van Overeem Hansen T, Sørensen CS (2016) Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat Rev Cancer* 16:599–612. <https://doi.org/10.1038/nrc.2016.72>
- Miki Y, Swensen J, Shattuck-Eidens D et al (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266:66–71
- Wooster R, Bignell G, Lancaster J et al (1995) Identification of the breast-cancer susceptibility gene Brca2. *Nature* 378:789–792
- Foulkes WD (2008) Inherited susceptibility to common cancers. *N Engl J Med* 359:2143–2153. <https://doi.org/10.1056/NEJMra0802968>
- Stratton MR, Rahman N (2008) The emerging landscape of breast cancer susceptibility. *Nat Genet* 40:17–22. <https://doi.org/10.1038/ng.2007.53>
- Balmaña J, Domchek SM, Tutt A, Garber JE (2011) Stumbling blocks on the path to personalized medicine in breast cancer: the case of PARP inhibitors for BRCA1/2-associated cancers. *Cancer Discov* 1:29–34. <https://doi.org/10.1158/2159-8274.CD-11-0048>
- Turner NC, Ashworth A (2011) Biomarkers of PARP inhibitor sensitivity. *Breast Cancer Res Treat* 127:283–286. <https://doi.org/10.1007/s10549-011-1375-8>
- De Vooght KMK, Wijk R, Van, Van Solinge WW (2009) Management of gene promoter mutations in molecular diagnostics. *Clin Chem* 55:698–708. <https://doi.org/10.1373/clinchem.2008.120931>
- Halvorsen M, Martin JS, Broadaway S, Laederach A (2010) Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet* 6:1–11. <https://doi.org/10.1371/journal.pgen.1001074>
- Brewster BL, Rossiello F, French JD et al (2012) Identification of fifteen novel germline variants in the BRCA1 3'UTR reveals a variant in a breast cancer case that introduces a functional miR-103 target site. *Hum Mutat* 33:1665–1675. <https://doi.org/10.1002/humu.22159>
- Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V et al (2017) Functional classification of DNA variants by hybrid mini-genomes: identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. *PLoS Genet* 13:e1006691. <https://doi.org/10.1371/journal.pgen.1006691>
- Sanz DJ, Acedo A, Infante M et al (2010) A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin Cancer Res* 16:1957–1967. <https://doi.org/10.1158/1078-0432.CCR-09-2564>
- Quiles F, Menéndez M, Tornero E et al (2016) Investigating the effect of 28 BRCA1 and BRCA2 mutations on their related transcribed mRNA. *Breast Cancer Res Treat*. <https://doi.org/10.1007/s10549-015-3676-9>
- Théry JC, Krieger S, Gaildrat P et al (2011) Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. *Eur J Hum Genet* 19:1052–1058. <https://doi.org/10.1038/ejhg.2011.100>
- Lheureux S, Lambert B, Krieger S et al (2011) Two novel variants in the 3'UTR of the BRCA1 gene in familial breast and/or ovarian cancer. *Breast Cancer Res Treat* 125:885–891. <https://doi.org/10.1007/s10549-010-1165-8>
- Hindorff LA, Sethupathy P, Junkins HA et al (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 106:9362–9367. <https://doi.org/10.1073/pnas.0903103106>
- Shim S-Y, Jeong HJ, Park HJ et al (2017) Functional variation of SHP-2 promoter is associated with preterm birth and delayed myelination and motor development in preterm infants. *Sci Rep* 7:6052. <https://doi.org/10.1038/s41598-017-06401-x>
- Haberle V, Lenhard B (2015) Promoter architectures and developmental gene regulation. *Semin Cell Dev Biol*. <https://doi.org/10.1016/j.semcdb.2016.01.014>
- Cooper DN (2002) Human gene mutation in pathology and evolution. *J Inherit Metab Dis* 25:157–182. <https://doi.org/10.1023/A:1015621710660>
- Poncz M, Ballantine M, Solowiejczyk D et al (1982)  $\beta$ -Thalassemia in a Kurdish Jew: single base changes in the T-A-T-A box. *J Biol Chem* 257:5994–5996
- Davis PL, Miron A, Andersen LM et al (1999) Isolation and initial characterization of the BRCA2 promoter. *Oncogene* 18:6000–6012
- Velasco E, Infante M, Durán M et al (2007) Heteroduplex analysis by capillary array electrophoresis for rapid mutation detection in large multiexon genes. *Nat Protoc* 2:237–246. <https://doi.org/10.1038/nprot.2006.482>
- de la Hoya M, Gutierrez-Enriquez S, Velasco E et al (2006) Genomic rearrangements at the BRCA1 locus in Spanish families with breast/ovarian cancer. *Clin Chem* 52:1480–1485. <https://doi.org/10.1373/clinchem.2006.070110>
- Li W, Cowley A, Uludag M et al (2015) The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* 43:W580–W584. <https://doi.org/10.1093/nar/gkv279>
- Bryksin AV, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques* 48:463–465. <https://doi.org/10.2144/000113418>
- Dos Santos ES, Caputo SM, Castera L et al (2017) Assessment of the functional impact of germline BRCA1/2 variants located in

- non-coding regions in families with breast and/or ovarian cancer predisposition. *Breast Cancer Res Treat*. <https://doi.org/10.1007/s10549-017-4602-0>
28. Dunham I, Kundaje A, Aldred SF et al (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74. <https://doi.org/10.1038/nature11247>
  29. Farré D, Roset R, Huerta M et al (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res* 31:3651–3653
  30. Sandelin A, Alkema W, Engström P et al (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* 32:D91–D94. <https://doi.org/10.1093/nar/gkh012>
  31. Acedo A, Sanz DJ, Durán M et al (2012) Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res* 14:R87. <https://doi.org/10.1186/bcr3202>
  32. Acedo A, Hernández-Moro C, Curiel-García Á et al (2015) Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat* 36:210–221. <https://doi.org/10.1002/humu.22725>
  33. Karimi M, Goldie LC, Cruickshank MN et al (2009) A critical assessment of the factors affecting reporter gene assays for promoter SNP function: a reassessment of -308 TNF polymorphism function using a novel integrated reporter system. *Eur J Hum Genet* 17:1454–1462. <https://doi.org/10.1038/ejhg.2009.80>
  34. Liu Q, Thompson BA, Ward RL et al (2016) Understanding the pathogenicity of noncoding mismatch repair gene promoter variants in Lynch syndrome. *Hum Mutat* 37:417–426. <https://doi.org/10.1002/humu.22971>
  35. de la Hoya M, Soukarieh O, López-Perolio I et al (2016) Combined genetic and splicing analysis of BRCA1 c.[594-2A> C; 641A> G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum Mol Genet* 25:2256–2268. <https://doi.org/10.1093/hmg/ddw094>
  36. Michailidou K, Lindström S, Dennis J et al (2017) Association analysis identifies 65 new breast cancer risk loci. *Nature* 551:92–94. <https://doi.org/10.1038/nature24284>
  37. GTEx Consortium (2013) The genotype-tissue expression (GTEx) project. *Nat Genet* 45:580–585. <https://doi.org/10.1038/ng.2653>
  38. Harper AR, Nayee S, Topol EJ (2015) Protective alleles and modifier variants in human health and disease. *Nat Rev Genet* 16:689–701. <https://doi.org/10.1038/nrg4017>
  39. Capasso M, Diskin S, Cimmino F et al (2014) Common genetic variants in NEFL influence gene expression and neuroblastoma risk. *Cancer Res* 74:6913–6924. <https://doi.org/10.1158/0008-5472.CAN-14-0431>
  40. Gochhait S, Bukhari SIA, Bairwa N et al (2007) Implication of BRCA2 -26G> A 5' untranslated region polymorphism in susceptibility to sporadic breast cancer and its modulation by p53 codon 72 Arg> Pro polymorphism. *Breast Cancer Res* 9:R71. <https://doi.org/10.1186/bcr1780>
  41. Maia A-T, Antoniou AC, O'Reilly M et al (2012) Effects of BRCA2 cis-regulation in normal breast and cancer risk amongst BRCA2 mutation carriers. *Breast Cancer Res* 14:R63. <https://doi.org/10.1186/bcr3169>
  42. Kauffmann A, Rosselli F, Lazar V et al (2007) High expression of DNA repair pathways is associated with metastasis in melanoma patients. *Oncogene* 27:565–573. <https://doi.org/10.1038/sj.onc.1210700>
  43. Sarasin A, Dessen P (2010) DNA repair pathways and human metastatic malignant melanoma. *Curr Mol Med* 10:413–418
  44. Klein HL (2008) The consequences of Rad51 overexpression for normal and tumor cells. *DNA Repair* 7:686–693. <https://doi.org/10.1016/j.dnarep.2007.12.008>
  45. Bièche I, Noguès C, Lidereau R (1999) Overexpression of BRCA2 gene in sporadic breast tumours. *Oncogene* 18:5232–5238. <https://doi.org/10.1038/sj.onc.1202903>



### Supplemental Figure S1. Sequence of the *BRCA2* promoter.

The sequence of the cloned *BRCA2* promoter (1250 bp) is indicated in upper case; the blue arrow marks an Alu repetitive sequence of the Sp family; sequence variations are shown as red small squares; microdeletions are shown as double-arrows in yellow, the Ensembl transcription start site (TSS, green-boxed) is also indicated in *BRCA2* exon 1; Forward (green; Pr-BR2-FW5) and reverse (red; Pr-BR2-RV5) sequencing primers are shown.

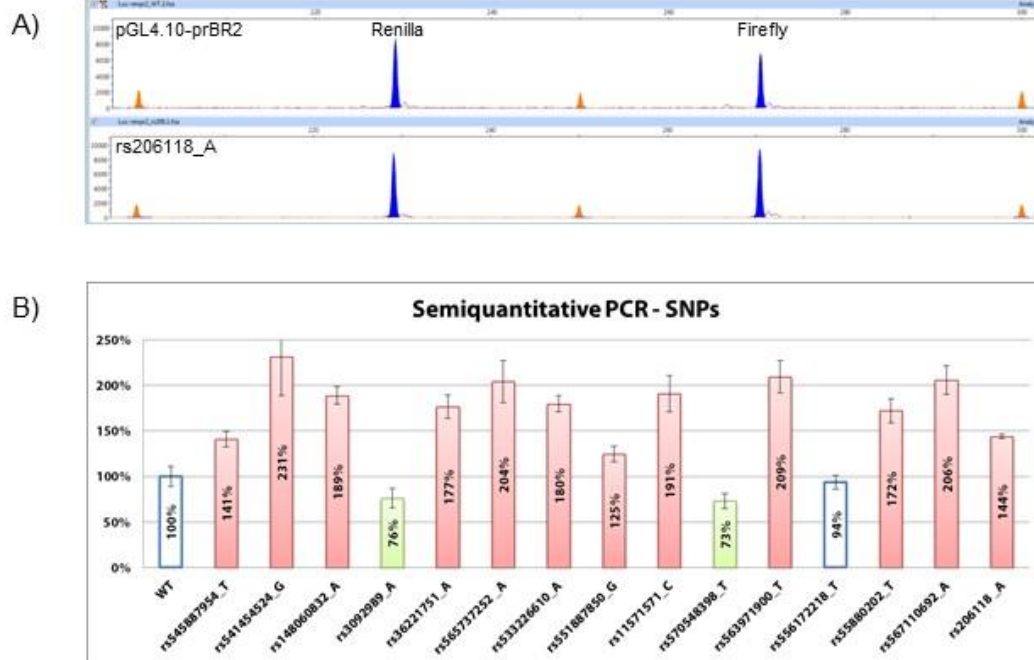
## Supplemental Methods

### RNA quantification: Semiquantitative PCR

Semi-quantitative fluorescent PCRs were undertaken to quantify Firefly and Renilla transcripts. Twenty ng of cDNA were amplified in triplicate in 25- $\mu$ l reactions with Platinum Taq (Life Technologies) and primers for Firefly (FAM-5' CGCTATGGGCTGAATACAAA 3' and 5'-TGGTAGTCGGTCTTGCTATC-3', 272bp) and Renilla (FAM- 5'-GAAGAGGGCGAGAAAATGG-3' and 5'-GAAGGTAGGCGTTGTAGTTG-3', 229bp), with the following thermocycling protocol: 94°C, 2 min +26 cycles x [94°C, 30 sec/60°C, 30 sec/72°C, 30 sec]. FAM-labeled products were run with LIZ-500 Size Standard at MacroGen and analyzed with the Peak Scanner software V1.0 as described elsewhere [1, 2], where peak areas were used to calculate Firefly/Renilla ratios.

1. Acedo A, Sanz DJ, Durán M, et al (2012) Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res* 14:R87 . doi: 10.1186/bcr3202

2. Acedo A, Hernández-Moro C, Curiel-García Á, et al (2015) Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat* 36:210–21 . doi: 10.1002/humu.22725.



**Supplemental Figure S2. Semi-quantitative PCR of RNA from Renilla and Firefly luciferases.** A) Example of fragment analysis of semi-quantitative fluorescent RT-PCRs of the wild type and the rs206118 (allele A) constructs. Renilla and firefly peaks are shown as blue peaks while the Liz-500 size standard is shown as orange peaks. B) Impact of SNPs on the promoter activity by semi-quantitative-PCR.

**Supplemental Table S1.** Location and primer sequences of microdeletions.

| 50 nt-<br>Microdeletions | Position<br>(TSS) | 5'→3' sequences  |
|--------------------------|-------------------|--|
| Pr_BR2-mdel13            | -509_-<br>460del  | ATTTTTCAGCAAGGAAACGGGCTCGGAGGTCTGATTTCAAATAACAGCCCCGCCACTA<br>TAGTGGGCGGGGCTGTTATTTGAAATCAGACCTCCGAGCCGTTTCCTTGCTTGAAAAAT    |
| Pr_BR2-mdel12            | -464_-<br>415del  | ACCCAATAGCAGAACAGCTACTGGAATAAATCCACAACCACACACCGACCACTCTAAGC<br>GCTTAGAGTGGTCGGTGTGTGGTTGTGGATTAGTTCAGTAGCTGTTCTGCTATTGGGT    |
| Pr_BR2-mdel11            | -419_-<br>370del  | AATAACAGCCCCGCCCACTACCACTAAGTGTGCTTTGGGGAACAGGTCTTGAGAGAACA<br>TGTTCTCTCAAGACCTGTTCCCAAGCGACACTTAGTGGTAGTGGGCGGGCTGTTATT     |
| Pr_BR2-mdel10            | -374_-<br>325del  | ACACACCGACCACTCTAAGCTTTTGTAAAGAGAACAAGGTATTTTCATAGGTCCCAGGTCG<br>CGACCTGGGACCTATGAAATACCTTTGTCTCTTACAAAAGCTTAGAGTGGTCGGTGTGT |
| Pr_BR2-mdel9             | -329_-<br>280del  | GAACAGGTCCTTGAGAGAACATCCCTTTTAACACCCAAACATGAGCTGGAGCAAAAAGAAA<br>TTTCTTTTGTCTCCAGCTCATGTTGGGTGTAAAAGGGATGTCTCTCAAGACCTGTTC   |
| Pr_BR2-mdel8             | -284_-<br>235del  | ATTTTCATAGGTCCCAGGTCGTGTCGCCAGGGAGTAGGCATAGGGGCGGCCCTCCAAGCA<br>TGCTTGAGGGGCGCCCTATGCCTACTCCCTCGGGACACGACCTGGGACCTATGAAAT    |
| Pr_BR2-mdel7             | -239_-<br>190del  | TGAGCTGGAGCAAAAAGAAAGGGATGGGGGCTTAAGGTCAGCGAGAAGAGAACACACAC<br>GTGTGTGTTCTCTTCTGCTGACCCCTAAGCCCCATCCCTTTCTTTTGTCTCCAGCTCA    |
| Pr_BR2-mdel6             | -194_-<br>145del  | AGGGGCGGCCCTCCAAGCAGGTTGGCTGATTTCGGTCAGATACTGACGGTTGGGATGCC<br>GGCATCCCAACCGTCAGTATCTGACCGAATCAGGCCACCCTGCTTGAGGGGCGGCCCT    |
| Pr_BR2-mdel5             | -149_-<br>100del  | AGCGAGAAGAGAACACACACTCCAGCTCCCTTTCGCCACACTGAGAAATACCCGACGCGG<br>CCGCTGCGGGTATTTCTCAGTGTGGCGAAAGGAGCTGGAGTGTGTCTCTCTCCTCCT    |
| Pr_BR2-mdel 4            | -104_-55del       | ATACTGACGGTTGGGATGCCTGACAAGGAACCTCCGGGTGGTGCCTGCTGCGTGTGCGC<br>GCGACACGCAGCACACGCACCACCGGAAGTTCTTGTGAGGCATCCCAACCGTCAGTAT    |
| Pr_BR2-mdel3             | -59_-10del        | CTGAGAAATACCCGACGCGGCCACCCAGGGGCCAGCGCGGGCTTGTGGCGGAGCTTCT<br>AGAAGCTCGCGCCACAAGCCCGCGCTGGCCCTGGGTGGGCGGCTGCGGGTATTTCTCAG    |
| Pr_BR2-mdel2             | -14_+36del        | GTGCGTGTGCTGCGTGTGCGGTACGGCGTAGGCGGAGCCGCTGTGGCACTGCTGCGCCT<br>AGGCGCAGCAGTGCCACAGCGGCTCCGCCTACGCCGTGACGCGACACGCAGCACACGCAC  |
| Pr_BR2-mdel1             | +32_+81del        | GGCTTGTTGGCGCGAGCTTCTGAAACTAGGCTGTCTTTTGCGGCGGTGGGTGCGCGCCGGG<br>CCCGGCGGCGACCCACCGCCGAAAAGACAGCTAGTTTCAGAAGCTCGCGCCACAAGCC  |

### ARTÍCULO 3

Eugenia Fraile-Bethencourt, Alberto Valenzuela-Palomo, Beatriz Díez-Gómez, Alberto Acedo and Eladio A. Velasco.

#### Identification of Eight Spliceogenic Variants in *BRCA2* Exon 16 by Minigene Assays.

Frontiers in Genetics, 2018; 9:188. <https://doi.org/10.3389/fgene.2018.00188>

*Impact factor 2017: 4.151*

**Introducción:** Los estudios genéticos realizados en *BRCA1* y *BRCA2* han identificado un gran número de variantes de significado clínico desconocido cuya interpretación funcional y clínica supone un reto en el consejo genético. Una importante fracción de estas variantes de DNA podría estar alterando el proceso de *splicing* en los genes de susceptibilidad a cáncer. Nuestro grupo ha analizado, principalmente mediante minigenes, más de 200 variantes en 19 exones de *BRCA2*, obteniendo que el 54% alteraban el *splicing*. Los minigenes son herramientas útiles para ensayar variantes de *splicing* cuando no se dispone de muestras de RNA de los pacientes. Nuestro objetivo se centra en analizar el exón 16 de *BRCA2* con el fin de caracterizar las variantes de *splicing* registradas en las bases de datos.

**Resultados:** Se analizaron 72 variantes procedentes de las bases de datos BIC y UMD con NNSplice y *Human Splicing Finder*. Se seleccionaron 12 que mostraban una alteración de los elementos esenciales de *splicing*: ocho variantes alteraban sitios canónicos de *splicing* y cuatro modificaban elementos reguladores. Las 12 variantes candidatas se introdujeron en un minigen de *BRCA2* con siete exones (14-20) mediante mutagénesis dirigida y fueron transfectadas en células MCF-7. Siete de las variantes analizadas (seis intrónicas y una *missense*) inducían patrones de *splicing* completamente aberrantes: c.7618-2A>T, c.7618-2A>G, c.7618-1G>C, c.7618-1G>A, c.7805G>C, c.7805C+1G>A y c.7805C+3A>C. Por otro lado, una de ellas (c.7802A>G) provocaba un efecto parcial. Se encontraron al menos 10 isoformas diferentes. Las principales fueron:  $\Delta 16p_{44}$  (aceptor alternativo 44 nt *downstream*, variantes del aceptor),  $\Delta 16$  (pérdida del exón 16; variantes del donador),  $\Delta 16p_{55}$  (aceptor

alternativo 55 nt *downstream*),  $\Delta 16q_4$  (donador alternativo 4 nt *upstream*),  $\Delta 16q_{100}$  (donador alternativo 100 nt *upstream*) y  $\nabla 16q^{20}$  (donador alternativo 20 nt *downstream*). Encontramos también dos transcritos minoritarios ( $\Delta 16p_{93}$  y  $\Delta 16,17p_{69}$ ) y otros dos no caracterizados de 893 y 954 nucleótidos. Las isoformas  $\Delta 16p_{44}$ ,  $\Delta 16$ ,  $\Delta 16p_{55}$ ,  $\Delta 16q_4$ ,  $\Delta 16q_{100}$  y  $\nabla 16q^{20}$  introducen codones de *stop* prematuros, los cuales, potencialmente, inactivarían BRCA2. De acuerdo con las guías propuestas por ACMG (*American College of Medical Genetic and Genomics*), las ocho variantes mencionadas podrían ser clasificadas como patogénicas o probablemente patogénicas. Por otro lado, siguiendo las normas del consorcio ENIGMA (*Evidence-based Network for the Interpretation of Germline Mutant Alleles*), siete variantes serían clasificadas como clase 4 y una de ellas (c.7802A>G) como clase 3.

**Conclusiones:** Este estudio recalca la importancia de los ensayos funcionales de *splicing* para la clasificación clínica de variantes genéticas. Nuestros resultados proporcionan nuevos datos sobre variantes que afectan al proceso de *splicing* en el exón 16 de *BRCA2* y están directamente relacionadas con la susceptibilidad a cáncer de mama.



# Identification of Eight Spliceogenic Variants in BRCA2 Exon 16 by Minigene Assays

Eugenia Fraile-Bethencourt<sup>1</sup>, Alberto Valenzuela-Palomo<sup>1</sup>, Beatriz Díez-Gómez<sup>1</sup>, Alberto Acedo<sup>1,2†</sup> and Eladio A. Velasco<sup>1\*</sup>

<sup>1</sup> Splicing and Genetic Susceptibility to Cancer, Instituto de Biología y Genética Molecular, Consejo Superior de Investigaciones Científicas, Universidad de Valladolid, Valladolid, Spain, <sup>2</sup> Biome Makers Inc., San Francisco, CA, United States

## OPEN ACCESS

### Edited by:

Paolo Peterlongo,  
IFOM – The FIRCC Institute  
of Molecular Oncology, Italy

### Reviewed by:

Logan Walker,  
University of Otago, New Zealand  
John Frederick Pearson,  
University of Otago, New Zealand

### \*Correspondence:

Eladio A. Velasco  
eavelsam@ibgm.uva.es

### † Present address:

Alberto Acedo,  
AC-Gen Reading Life SL,  
Valladolid, Spain

### Specialty section:

This article was submitted to  
Cancer Genetics,  
a section of the journal  
Frontiers in Genetics

**Received:** 16 March 2018

**Accepted:** 08 May 2018

**Published:** 24 May 2018

### Citation:

Fraile-Bethencourt E,  
Valenzuela-Palomo A,  
Díez-Gómez B, Acedo A and  
Velasco EA (2018) Identification  
of Eight Spliceogenic Variants  
in BRCA2 Exon 16 by Minigene  
Assays. *Front. Genet.* 9:188.  
doi: 10.3389/fgene.2018.00188

Genetic testing of *BRCA1* and *BRCA2* identifies a large number of variants of uncertain clinical significance whose functional and clinical interpretations pose a challenge for genetic counseling. Interestingly, a relevant fraction of DNA variants can disrupt the splicing process in cancer susceptibility genes. We have tested more than 200 variants throughout 19 *BRCA2* exons mostly by minigene assays, 54% of which displayed aberrant splicing, thus confirming the utility of this assay to check genetic variants in the absence of patient RNA. Our goal was to investigate *BRCA2* exon 16 with a view to characterizing spliceogenic variants recorded at the mutational databases. Seventy-two different BIC and UMD variants were analyzed with NNSplice and Human Splicing Finder, 12 of which were selected because they were predicted to disrupt essential splice motifs: canonical splice sites (ss; eight variants) and exonic/intronic splicing enhancers (four variants). These 12 candidate variants were introduced into the *BRCA2* minigene with seven exons (14–20) by site-directed mutagenesis and then transfected into MCF-7 cells. Seven variants (six intronic and one missense) induced complete abnormal splicing patterns: c.7618-2A>T, c.7618-2A>G, c.7618-1G>C, c.7618-1G>A, c.7805G>C, c.7805+1G>A, and c.7805+3A>C, as well as a partial anomalous outcome by c.7802A>G. They generated at least 10 different transcripts:  $\Delta 16p_{44}$  (alternative 3'ss 44-nt downstream; acceptor variants),  $\Delta 16$  (exon 16-skipping; donor variants),  $\Delta 16p_{55}$  (alternative 3'ss 55-nt downstream),  $\Delta 16q_4$  (alternative 5'ss 4-nt upstream),  $\Delta 16q_{100}$  (alternative 5'ss 4-nt upstream),  $\nabla 16q^{20}$  (alternative 5'ss 20-nt downstream), as well as minor ( $\Delta 16p_{93}$  and  $\Delta 16,17p_{69}$ ) and uncharacterized transcripts of 893 and 954 nucleotides. Isoforms  $\Delta 16p_{44}$ ,  $\Delta 16$ ,  $\Delta 16p_{55}$ ,  $\Delta 16q_4$ ,  $\Delta 16q_{100}$ , and  $\nabla 16q^{20}$  introduced premature termination codons which presumably inactivate *BRCA2*. According to the guidelines the American College of Medical Genetics and Genomics these eight variants could be classified as pathogenic or likely pathogenic whereas the Evidence-based Network for the Interpretation of Germline Mutant Alleles rules suggested seven class 4 and one class 3 variants. In conclusion, our study highlights the relevance of splicing functional assays by hybrid minigenes for the clinical classification of genetic variations. Hence, we provide new data about spliceogenic variants of *BRCA2* exon 16 that are directly correlated with breast cancer susceptibility.

**Keywords:** breast cancer, *BRCA2*, DNA variants, splicing, hybrid minigenes

## INTRODUCTION

Hereditary Breast and Ovarian Cancer (HBOC) represents 5–10% of all breast cancers. Nowadays, more than 25 HBOC susceptibility genes have been identified, most of them involved in DNA repair pathways (Nielsen et al., 2016). Deleterious variants of the most prevalent genes *BRCA1* (MIM# 113705) and *BRCA2* (MIM# 600185) confer up to 87% of risk to develop breast cancer by the age of 70 years (Petrucelli et al., 2013). Apart from specific founder deleterious mutations (Levy-Lahad et al., 1997; Infante et al., 2013), there have been described thousands of different *BRCA1/2* variants at the mutation databases. According to Universal Mutation Database (UMD, <http://www.umd.be>; date last accessed 2017/06/16) 2,495 and 3,454 different variants have been detected in *BRCA1* and *BRCA2*, respectively, where a relevant fraction of them has been classified as variants of uncertain significance (VUS). These pose a challenge in clinical genetics since mutation carriers could benefit from preventive and prophylactic measures as well as new targeted therapies such as the Poly-ADP Ribose Polymerase Inhibitors (Ricks et al., 2015).

Standard approaches tend to classify DNA variants from the protein point of view. In this way, nonsense variants and frameshift insertions and deletions are automatically classified as pathogenic if they truncate critical protein domains [Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) class 5<sup>1</sup>]. However, upstream gene expression mechanisms, such as splicing, can be disrupted by DNA changes. In fact, splicing is a critical highly regulated process involved in many cell functions whose disruption has been directly related with disease, being common in cancer (Wang and Cooper, 2007; Douglas and Wood, 2011). Likewise, spliceogenic variants are more common than they are thought, and they are not restricted to the sequences of the canonical donor and acceptor sites since it has been suggested that up to 50% of exon variants could also affect splicing (López-Bigas et al., 2005). This can be explained by the wide range of splicing regulatory elements (SREs) that control this process, which include the conserved splice sites (5'ss and 3'ss), the branch point, polypyrimidine track, exonic/intronic splicing enhancers (ESEs/ISEs) and exonic/intronic splicing silencers (ESSs/ISSs) (Grodecká et al., 2017), as well as other regulatory components or the RNA secondary structure (Soemedi et al., 2017). Thus, all these factors cooperate with splicing factors and the spliceosome, to accurately remove introns (Will and Lührmann, 2011).

Interestingly, spliceogenic variants are often found in *BRCA2*. Our previous results showed that more than a half of tested *BRCA2* variants impaired splicing (Acedo et al., 2012, 2015; Fraile-Bethencourt et al., 2017). Moreover, the minigene technology was confirmed as a reliable tool to functionally assay potential splicing variants. Here, we aimed to check *BRCA2* exon 16 candidate variants to characterize the splicing effects using the pSAD-based minigene MGBR2\_14-20, previously

employed to assay DNA variants of exons 17 and 18 (Fraile-Bethencourt et al., 2017). We have assayed 12 likely spliceogenic variants from HBOC patients reported in databases and selected after bioinformatics predictions. Wild-type (wt) and mutant minigenes assays showed that eight variants altered the splicing. Thus, we provide valuable information of spliceogenic *BRCA2* exon 16 variants that could be classified following ENIGMA and American College of Medical Genetics and Genomics (ACMG) guidelines (Richards et al., 2015).

## MATERIALS AND METHODS

Ethical approval for this study was obtained from the Ethics Review Committee of the Hospital Universitario Río Hortega de Valladolid (6/11/2014).

### Variant Collection and *In Silico* Analyses

*BRCA2* introns 15 and 16 and exon 16 variants were collected from the BIC database<sup>2</sup> and the BRCA Share Database (UMD, date last accessed 2017/06/16; <http://www.umd.be/BRCA2/>) (Beroud et al., 2016). Variant descriptions were according to the *BRCA2* GenBank sequence NM000059.1 and the guidelines of the Human Genome Variation Society (HGVS<sup>3</sup>).

Wild-type and mutant sequences were analyzed with NNSPLICE<sup>4</sup> (Reese et al., 1997) and Human Splicing Finder version 3.0 (HSF<sup>5</sup>) (Desmet et al., 2009), which includes algorithms to detect splice sites, branch point, silencers, and enhancers (Fairbrother et al., 2002; Cartegni et al., 2003; Sironi et al., 2004; Wang et al., 2004; Yeo and Burge, 2004; Zhang and Chasin, 2004).

### Minigene and Mutagenesis

MGBR2\_ex14-20 was assembled as previously described (Fraile-Bethencourt et al., 2017). DNA variants and deletions were introduced by the QuikChange Lightning Kit (Agilent, Santa Clara, CA, United States). The wt minigene MGBR2\_ex14-20 was used as template to generate 12 BIC/BRCA Share DNA variants and 4 microdeletions (Table 1). They were checked by SANGER sequencing at the MacroGen Spain facility (MacroGen, Madrid, Spain).

### Transfection of Eukaryotic Cells

MCF-7 cells were plated ( $\sim 2 \times 10^5$  cells/well) and grown to 90% confluency in 0.5 mL of medium (MEME, 10% fetal bovine serum, 2 mM glutamine, 1% non-essential amino acids, and 1% penicillin/streptomycin) in four-well plates (Nunc, Roskilde, Denmark). Transfections were made with 1  $\mu$ g of minigene and 2  $\mu$ L of low toxicity Lipofectamine (Life Technologies, Carlsbad, CA, United States) in Gibco<sup>TM</sup> Opti-Mem<sup>TM</sup> (Thermo Fisher Scientific, Waltham, MA, United States). Cells were incubated with 300  $\mu$ g/mL of cycloheximide (Sigma-Aldrich, St. Louis,

<sup>1</sup><https://enigmaconsortium.org/library/general-documents/enigma-classification-criteria/>

<sup>2</sup><https://research.nhgri.nih.gov/projects/bic/Member/index.shtml>

<sup>3</sup><http://www.hgvs.org/mutnomen/>

<sup>4</sup>[http://www.fruitfly.org/seq\\_tools/splice.html](http://www.fruitfly.org/seq_tools/splice.html)

<sup>5</sup><http://www.umd.be/HSF3/>

**TABLE 1 |** Mutagenesis primers of candidate splicing variants.

| HGVS variants  | Primers (5' → 3')  |
|----------------|--|
| c.7618-2A>T    | GTGTGTTTATTTTGTGTGCTGTATACGTATGGCG<br>CGCCATACGTATACAGCAACACAAAATAAACACAC  |
| c.7618-2A>G    | GTGTGTTTATTTTGTGTGGCTGTATACGTATGGCG<br>CGCCATACGTATACAGCCACACAAAATAAACACAC                                       |
| c.7618-1G>A    | GTGTGTGTTTATTTTGTGTAACGTATACGTATGGCGTTTC<br>GAAACGCCATACGTATACAGTTACACAAAATAAACACACAC                            |
| c.7618-1G>C    | GTGTGTGTTTATTTTGTGTACCTGTATACGTATGGCGTTTC<br>GAAACGCCATACGTATACAGGTACACAAAATAAACACACAC                           |
| c.7625C>G      | TTATTTTGTGTAGCTGTATAGGTATGGCGTTTCTAAACATT<br>AATGTTTAGAAACGCCATACCTATACAGCTACACAAAATAA                           |
| c.7738C>T      | ATGGACTGGAAGGAATATAGTTGGCTGATGGTGGATGG<br>CCATCCACCATCAGCCAATATATCTTTCCAGTCCAT                                   |
| c.7753G>A      | AATACAGTTGGCTGATGGTAGATGGCTCATACCCCTCCAAT<br>ATTGGAGGGTATGAGCCATCTACCATCAGCCAATGTATT                             |
| c.7772A>G      | GGATGGCTCATACCCCTCAGTGATGGAAGGCTGGAAAAAG<br>CTTTTCCAGCCTTTCCATCACTGGAGGGTATGAGCCATCC                             |
| c.7802A>G      | GCTGGAAGAAGAATTTTGTAGGTACTCTATGCAAAAAG<br>CTTTTGCATAGAGTACCTACAAAATCTCTTTTCCAGC                                  |
| c.7805G>C      | GGAAAGAAGAATTTTATACGTACTCTATGCAAAAAGATT<br>AATCTTTTGCATAGAGTACGTATAAAATCTCTTTTCC                                 |
| c.7805+1G>A    | GGAAAGAAGAATTTTATAGATACTCTATGCAAAAAGATTG<br>CAATCTTTTGCATAGAGTATCTATAAAATCTCTTTTCC                               |
| c.7805+3A>C    | AAGAAGAATTTATAGGTCCTCTATGCAAAAAGATTG<br>CAATCTTTTGCATAGAGGACCTATAAAATCTCTCTT                                     |
| Microdeletions | Primers (5' → 3')  |
| c.7620_7649del | TTGTGTGTGTTTATTTTGTGTAGCTAAAAATTAACAGCAAA<br>AATGCAGAG<br>CTCTGCATTTTGTCTGTTAATTTTGTAGCTACACAAAATAAC<br>ACACACAA |
| c.7645_7674del | GTATACGTATGGCGTTTCTAAACATTCTTTTCAGTTTCACAC<br>TGAAGATT<br>AATCTTCAGTGTGAAACTGAAAGAATGTTTAGAAACGCCA<br>TACGTATAC  |
| c.7748_7772del | ACTGGAAAAGGAATACAGTTGGCTGGAAGGCTGGAAAAAG<br>AAGAATTTTA<br>TAAATCTCTCTTTTCCAGCCTTTCCAGCCAATGTATTCTT<br>TTTCCAGT   |
| c.7773_7802del | ATGGTGGATGGCTCATACCCCTCAATAGGTACTCTATGCAA<br>AAAGATTGT<br>ACAATCTTTTGCATAGAGTACCTATTGGAGGGTATGAGCC<br>ATCCACCAT  |

MO, United States) for 4 h to inhibit nonsense-mediated decay (NMD). RNA was purified with the Genematrix Universal RNA Purification Kit (EURx, Gdansk, Poland) with on-column DNase I digestion to degrade genomic DNA that could interfere with RT-PCR.

## RT-PCR of Minigenes

Approximately 400 ng of RNA was retrotranscribed using RevertAid H Minus First Strand cDNA Synthesis Kit (Life Technologies, Carlsbad, CA, United States) and the gene-specific primer RTPSPL3-RV (5'-TGAGGAGTGAATTGGTCGAA-3').

Samples were incubated at 42°C for 1 h, and reactions were inactivated at 70°C for 5 min. Then, 40 ng of cDNA was amplified in 50 µL reaction with pMAD\_607FW (Patent P201231427, CSIC) and RTBR2\_ex17RV2 (5'-GGCTTAGGCATCTATTAGCA-3') or with RT\_ex15FW (5'-CGAATTAAGAAGAAACAAAGG-3') and pSAD\_RT\_RV (Patent P201231427, CSIC) using Platinum Taq DNA polymerase (Life Technologies, Carlsbad, CA, United States) (size of transcripts: 1018 and 1250 nt, respectively). Samples were denatured at 94°C for 2 min, followed by 35 cycles consisting of 94°C for 30 s, Td-2°C for 30 s, and 72°C (1 min/kb), and a final extension step at 72°C for 5 min. Sequencing reactions were performed by the sequencing facility of MacroGen Spain. Semi-quantitative fluorescent 26 cycles PCRs were done in triplicate with primers pMAD\_607FW-FAM and RTBR2\_ex17RV2 using Platinum Taq DNA polymerase (Life Technologies, Carlsbad, CA, United States). FAM-labeled products were run with Genescan LIZ-1200 as size standard (Life Technologies, Carlsbad, CA, United States) at the MacroGen facility and analyzed with the Peak Scanner software V1.0. Only peaks with heights ≥50 relative fluorescence unit (RFU) were considered. Mean peak areas of each transcript of three runs were used to quantify the relative abundance of each transcript.

## RESULTS

### Bioinformatics Analysis of Splicing Variants

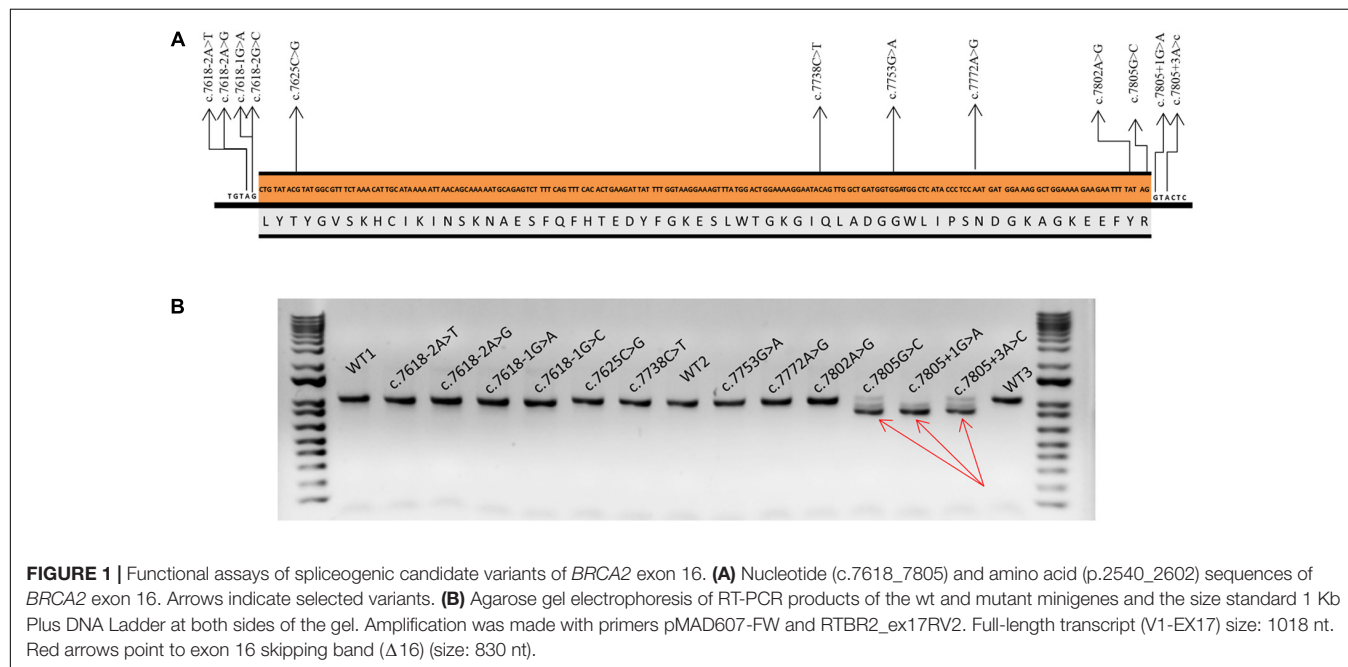
Seventy-two variants were collected from the BIC and UMD databases. Among them, 35 had been previously classified as VUS and 34 as pathogenic or likely pathogenic. In order to select possible spliceogenic variants, they were analyzed by the splicing prediction software NNSplice and HSF (Supplementary Table S1). Selections were made following the next criteria: (a) ss creation or disruption; (b) branch point disruption; and (c) ESS creation (hnRNPA1). Curiously, NNSplice did not recognize exon 16 canonical 5'ss. In contrast, a very strong 100-nt upstream cryptic donor (NNSplice score: 0.99) was identified at position c.7706\_7707. The MaxEnt results showed a weak canonical donor (4.68) and a strong cryptic donor (8.92) (Table 2).

Twelve variants were selected (Figure 1): six intronic (c.7618-2A>T, c.7618-2A>G, c.7618-1G>C, c.7618-1G>A, c.7805+1G>A, and c.7805+3A>C), five missense (c.7625C>G, c.7753G>A, c.7772A>G, c.7802A>G, and c.7805G>C), and one nonsense (c.7738C>T) variants. Four missense variants (c.7625C>G, c.7753G>A, c.7772A>G, and c.7802A>G) and c.7805+3A>C had been previously classified as VUS. Intronic variants (c.7618-2, -1 and c.7805+1, +3) and c.7805G > C disrupted the canonical ss, whereas variants c.7625C>G and c.7753G>A created new ss. DNA change c.7802A>G was selected because of its proximity to the canonical donor site and the presumable generation of an alternative "gt" donor site 4-nt upstream (underlined, TTTGTAGgtactc). Finally, bioinformatics results of c.7738C>T and c.7772A>G suggested the creation of one ESS (hnRNPA1) (Table 2).

**TABLE 2 |** Bioinformatics analysis of potential splicing variants of *BRCA2* exon 16.

| DNA variants             | NNSplice <sup>1</sup>                          | MaxEnt <sup>1</sup>      | ESE finder <sup>2</sup>                | ESE HSF <sup>2</sup> | hnRNPA1 <sup>2</sup> |
|--------------------------|--|--------------------------|--|----------------------|----------------------|
| Ex16-Canonical ss        | 3'SS: 0.69<br>5'SS <0.4                        | 3'SS: 7.11<br>5'SS: 4.68 |  |                      |                      |
| c.7618-2A>T              | [−] 3'SS: 0.69 → <0.4                          | [−] 3'SS: 7.11 → −1.24   |  |                      | [−] (70.24)          |
| c.7618-2A>G              | [−] 3'SS: 0.69 → <0.4                          | [−] 3'SS: 7.11 → −0.83   | [−] SRp55 (83.78)                      |                      | [−] (70.24)          |
| c.7618-1G>A              | [−] 3'SS: 0.69 → <0.4                          | [−] 3'SS: 7.11 → −1.63   | [−] SRp55 (77.06)                      |                      | [−] (70.24)          |
| c.7618-1G>C              | [−] 3'SS: 0.69 → <0.4                          | [−] 3'SS: 7.11 → −0.94   | [−] SRp55 (77.06)                      |                      | [−] (70.24)          |
| c.7625C>G (p.Thr2542Arg) | [+] 3'SS (<0.4 → 0.94); [+] 5'SS (<0.4 → 0.91) | [+] 5'SS (2.23 → 7.37)   | [−] SRp55(89.80)                       | [−] 9G8 (62.28)      | [+] (88.33)          |
| c.7738C>T (p.Gln2580Ter) |  |                          | [−] SC35 (81.01)                       |                      | [+] (70.24)          |
| c.7753G>A (p.Gly2585Arg) |  | [+] 5'SS (−1.33 → 3.06)  |  |                      | [+] (74.05)          |
| c.7772A>G (p.Asn2591Ser) |  |                          | [+] SF2/ASF (IgM) (74.85)              | [−] 9G8 (60.94)      | [+] (69.52)          |
| c.7802A>G (p.Tyr2601Cys) |  |                          | [−] SRp40 (83.89)                      |                      | [−] (66.43)          |
| c.7805G>C (p.Arg2602Thr) |  | [−] 5'SS: 4.68 → −2.79   | [+] SRp55 (89.80)                      |                      | [−] (69.05)          |
| c.7805+1G>A              |  | [−] 5'SS: 4.68 → −3.50   | [−] SRp40 (83.89);<br>[−] SC35 (82.36) |                      |                      |
| c.7805+3A>C              |  | [−] 5'SS: 4.68 → −3.85   |  |                      |                      |

[+] and [−] symbols indicate creation or disruption of splicing motifs, respectively. Thresholds of the splicing programs: <sup>1</sup>Splice sites (ss): NNSPLICE (values 0–1): Cut-offs = 0.4 for both 5'- and 3' ss (ss disruption <0.4, ss creation >0.4); MaxEnt: 3.0 for 5' and 3' ss (variation threshold ± 30% according to HSF). <sup>2</sup>Enhancers and silencers HSF: Human Splicing Finder matrices (default values, <http://www.umd.be/HSF3/technicaltips.html>); ESEfinder cut-offs (HSF scale, normalized to 0–100): SF2/ASF: 72.98/SF2/ASF (IgM-BRCA1): 70.51/SRp40: 78.08/SC35: 75.05/SRp55: 73.86; ESE motifs from HSF, cut-offs values (0–100): Tra2: 75.964/9G8: 59.245; hnRNP motifs: hnRNPA1: 65.476 (these values were considered the limits for the disruption or creation of a splicing regulatory element).



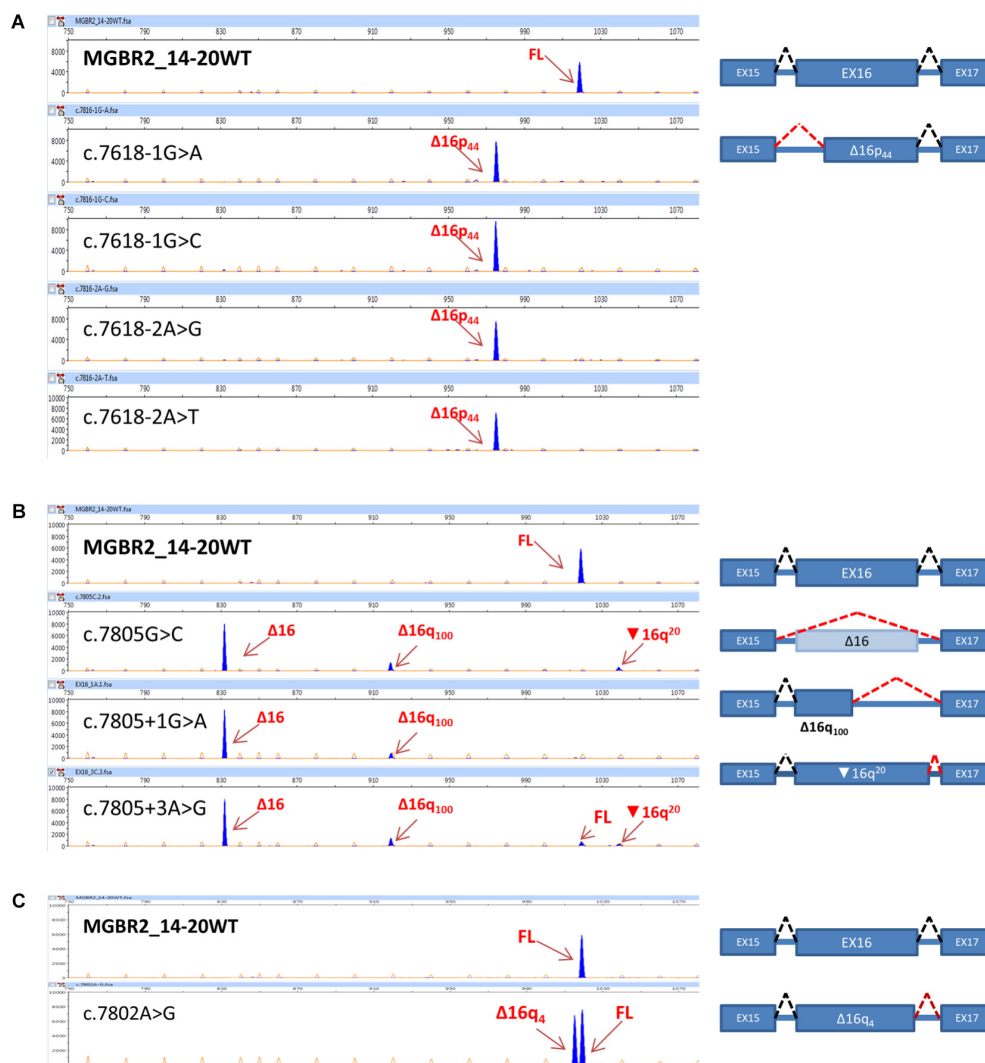
## Splicing Functional Assays of DNA Variants

The minigene MGBR2\_ex14-20 had been already shown as a robust tool to assay possible spliceogenic variants contained in any of those exons and flanking introns (Fraile-Bethencourt et al., 2017). The wt construct produced a full-length transcript of the expected size (1806 nt), sequence, and structure (V1-BRCA2 exons 14-20-V2). To map the presence of putative splicing enhancers, a set of four overlapping exonic microdeletions

were generated, which spanned 55-nt of the 5'- and 3'-ends (Fairbrother et al., 2004). This strategy had been previously shown to increase the accuracy of predictions of ESE disrupting variants (Acedo et al., 2015; Fraile-Bethencourt et al., 2017). None of the microdeletions induced splicing anomalies suggesting that this exon is not controlled by ESEs (data not shown). Consequently, ESE-disrupting variants, as unique selection criterion, were not chosen for subsequent functional tests (Table 2).

Selected variants were introduced into the minigene and functionally assayed in MCF-7 cells. Agarose electrophoresis clearly showed that three of them (c.7805G>C, c.7805+1G>A, and c.7805+3A>C) induced aberrant splicing patterns (Figure 1). However, the high resolution and sensitivity of fluorescent capillary electrophoresis allowed us to identify a total of eight variants, including the three previous ones, that disrupted splicing: c.7618-2A>T, c.7618-2A>G, c.7618-1G>C, c.7618-1G>A, c.7802A>G, c.7805G>C, c.7805+1G>A, and c.7805+3A>C (Figure 2). Actually, this approach is able to detect rare transcripts with a relative abundance below 1% or can resolve transcripts that differ only in a few nucleotides (e.g., only 4-nt between the canonical and  $\Delta 16q4$  isoforms). A total of at least 10 different aberrant transcripts were characterized

by fragment analysis and sequencing:  $\Delta 16p44$  (44-nt deletion; alternate 3'ss 44-nt downstream),  $\Delta 16p55$  (55-nt del; alternate 3'ss 55-nt downstream),  $\Delta 16$  (exon 16 skipping),  $\Delta 16q4$  (4-nt del; alternate 5'ss 4-nt upstream),  $\Delta 16q100$  (100-nt del; alternate 5'ss 100-nt upstream),  $\nabla 16q20$  (20-nt insertion; alternate 5'ss 20-nt downstream), minor ( $\Delta 16p93$  and  $\Delta 16,17p69$ ), and uncharacterized transcripts of 893 and 954 nt (Figure 2 and Table 3). On the one hand, fragments analysis and sequences revealed that 3'ss disrupting variants (positions -2 and -1) provoked the use of a cryptic acceptor 44-nt downstream ( $\Delta 16p44$ ) within exon 16 (Figure 2A). Interestingly, this cryptic 3'ss was not recognized either by NNSplice or MaxEnt. The loss of 44-nt at 5' of exon 16 would suppose a frameshift deletion and a premature termination codon (PTC) (p.L2540Qfs\*11).



**FIGURE 2 |** Fluorescent capillary electrophoresis of transcripts from *BRCA2* exon 16 variants. On the left, screenshots of electropherograms are shown. cDNA was amplified with primers FAM-labeled pMAD\_607FW and RTBR2\_ex17RV2. Arrows indicate transcripts (blue peaks). Full-length transcript: 1018 nt. Size standard was Genescan LIZ 1200 (orange/faint peaks). Fragments were analyzed with the Peak Scanner software v1.0. Fragment sizes (bp) and relative fluorescent units are indicated on the x- and y-axes, respectively. On the right, diagrams of the splicing patterns are shown. Boxes represent exons, discontinued black lines represent canonical splicing, and discontinue red lines represent aberrant splicing. (A) Acceptor site variants. (B) Donor site variants. (C) Alternative donor variant.

**TABLE 3 |** Quantification of transcripts of spliceogenic variants of exon 16 by fluorescent capillary electrophoresis.

|              | Canonical       | $\Delta 16p_{44}$ | $\Delta 16p_{55}$ | $\Delta 16$     | $\Delta 16q_4$  | $\Delta 16q_{100}$ | $\nabla 16q^{20}$ | Others                                   |
|--------------|-----------------|-------------------|-------------------|-----------------|-----------------|--------------------|-------------------|--|
| MGBR2_14-20  | 100%            |                   |                   |                 |                 |                    |                   |  |
| DNA variants |                 |                   |                   |                 |                 |                    |                   |  |
| c.7618-2A>T  |                 | 96.9 $\pm$ 0.4%   | 1.8 $\pm$ 0.3%    | 1.3 $\pm$ 0.03% |                 |                    |                   |  |
| c.7618-2A>G  |                 | 97.2 $\pm$ 0.4%   |                   |                 |                 |                    |                   | 2.8 $\pm$ 0.4% (954 nt)                  |
| c.7618-1G>A  |                 | 91.5 $\pm$ 0.2%   | 4.7 $\pm$ 0.2%    | 0.7 $\pm$ 0.03% |                 |                    |                   | 1.7 $\pm$ 0.03% ( $\Delta 16p_{93}$ );   |
|              |                 |                   |                   |                 |                 |                    |                   | 1.4 $\pm$ 0.04% ( $\Delta 16,17p_{69}$ ) |
| c.7618-1G>C  |                 | 92.6 $\pm$ 0.3%   | 1.9 $\pm$ 0.2%    | 2.8 $\pm$ 0.05% |                 |                    |                   | 1.1 $\pm$ 0.05% ( $\Delta 16p_{93}$ )    |
|              |                 |                   |                   |                 |                 |                    |                   | 0.7 $\pm$ 0.03% (893 nt)                 |
|              |                 |                   |                   |                 |                 |                    |                   | 0.9 $\pm$ 0.07% ( $\Delta 16,17p_{69}$ ) |
| c.7625C>G    | 100%            |                   |                   |                 |                 |                    |                   |  |
| c.7738C>T    | 100%            |                   |                   |                 |                 |                    |                   |  |
| c.7753G>A    | 100%            |                   |                   |                 |                 |                    |                   |  |
| c.7772A>G    | 100%            |                   |                   |                 |                 |                    |                   |  |
| c.7802A>G    | 54.3 $\pm$ 0.3% |                   |                   |                 | 45.7 $\pm$ 0.3% |                    |                   |  |
| c.7805G>C    |                 |                   |                   | 77.6 $\pm$ 0.6% |                 | 14.4 $\pm$ 0.3%    | 6.5 $\pm$ 0.3%    | 1.5 $\pm$ 0.06% ( $\Delta 16,17p_{69}$ ) |
| c.7805+1G>A  |                 |                   |                   | 88.0 $\pm$ 0.2% |                 | 10.1 $\pm$ 0.1%    |                   | 1.9 $\pm$ 0.02% ( $\Delta 16,17p_{69}$ ) |
| c.7805+3A>C  | 7.6 $\pm$ 0.2%  |                   |                   | 75.3 $\pm$ 0.5% |                 | 13.3 $\pm$ 0.2%    | 3.8 $\pm$ 0.3%    |  |

HGVS-RNA effect of transcripts:  $\Delta 16p_{44}$ , r.7618\_7661del;  $\Delta 16q_4$ , r.7802\_7805del;  $\Delta 16$ , r.7618\_7805del;  $\Delta 16q_{100}$ , r.7706\_7805del;  $\Delta 16p_{55}$ , r.7618\_7672del;  $\nabla 16q^{20}$ , r.7805\_7806ins7805+1\_7805+20;  $\Delta 16p_{93}$ , r.7618\_7710del;  $\Delta 16,17p_{69}$ , r.7618\_7874del.

On the other hand, 5'ss variants (positions +1 and +3) produced exon 16 skipping ( $\Delta 16$ ), which means a frameshift deletion through the loss of 188-nt from r.7618 to r.7805 (**Figure 2B**). Consequently, BRCA2 would be truncated with a PTC four codons downstream (p.L2540Gfs\*4). Last exon nucleotide variant (c.7805G>C) induced the same outcome ( $\Delta 16$ ) highlighting the importance of this position conservation (G in nearly 80% in all exons) in exon recognition (Zhang, 1998). Fragment analysis of variants c.7805G>C, c.7805+1G>A, and c.7805+3A>C also showed  $\sim 14\%$  of transcript  $\Delta 16q_{100}$ , which corresponds with the use of the previously mentioned cryptic 5'ss within exon 16 (NNSplice: 0.99; MaxEnt: 8.92), provoking r.7706\_7805del (p.K2570Lfs\*45) (**Table 3** and **Figure 2B**). Finally, missense variant c.7802A>G created a new 5'ss, which resulted in  $\sim 45\%$  of the aberrant transcript  $\Delta 16q_4$  (**Figure 2C**). The loss of four nucleotides would introduce a PTC into the protein (p.Y2601Wfs\*46). Thus, our results showed clearly how these eight variants disrupted splicing. Moreover, seven of them (c.7618-2A>T; c.7618-2A>G; c.7618-1G>C; c.7618-1G>A; c.7805G>C; c.7805+1G>A; c.7805+3A>C) generated more than  $\sim 92\%$  of frameshift transcripts.

## DISCUSSION

Nowadays, with the advent of new generation sequencing technologies and, namely, cancer-gene panels (Slavin et al., 2015), thousands of variants are being described. However, their classifications as neutral or deleterious variants pose a challenge in Human Genetics. In fact, some deleterious variants can be missed because they are synonymous or intronic. Moreover, a significant fraction of BRCA2 variants are considered VUS and require additional proofs to be reclassified, including functional tests. Here, we have shown that the minigene MGBR2\_14-20 is a

robust tool to functionally assay candidate spliceogenic variants of the BRCA2 exon 16. Until now, we have comprehensively studied candidate splicing variants from 20 out of 27 BRCA2 exons (Sanz et al., 2010; Acedo et al., 2012, 2015; Fraile-Bethencourt et al., 2017). Thus, we have found six intronic and two missense BRCA2 variants which alter the splicing and could confer cancer risk.

BRCA2 exon 16 codifies from Leucine 2540 to Arginine 2602 (p.2540\_2602). Interestingly, according to the International Agency for Research on Cancer (IARC<sup>6</sup>), this is a conserved region, since there is  $\sim 22\%$  of ultra-conserved aminoacids from human to sea urchin and  $\sim 54\%$  between mammals. Furthermore, this protein segment belongs to FANCD2- and DSS1-binding domains. Fanconi Anemia group D2 (FANCD2) protein binds to aminoacids from position p.2350 to p.2545 of BRCA2 and it has been suggested to have a role in the repair process (Hussain et al., 2004). DSS1 (Delete in Split hand/Split foot) protein, which binds to BRCA2 at positions p.2467\_2957 (Marston et al., 1999), is an essential element of BRCA2 stability, since its loss supposes a dramatic decrease of BRCA2 levels (Li et al., 2006). Altogether, this highlights the value of exon 16 in BRCA2 function. Moreover, exon 16 skipping supposes a frame-shift deletion and the generation of a PTC (p.L2540Gfs\*4), which would truncate the protein and subsequently loss the C-terminal region that would compromise BRCA2 function.

This study, based on minigene technology, provides detailed information about the impact on splicing of 12 BRCA2 exon 16 variants. Aberrant splicing outcomes were found in eight of these variants, six intronic and two missense changes. Intriguingly, none of the aberrant transcripts described here was previously reported as natural alternative splicing events of the BRCA2

<sup>6</sup>[http://agvgd.hci.utah.edu/BRCA2\\_Spur.html](http://agvgd.hci.utah.edu/BRCA2_Spur.html)

gene (Fackenthal et al., 2016). Among them, seven (c.7618-2A>T, c.7618-2A>G, c.7618-1G>A, c.7618-1G>C, c.7805G>C, c.7805+1G>A, and c.7805+3A>C) provoked more than ~92% of frameshift transcripts. Interestingly, previous studies of variant c.7618-1G>A in lymphoblastoid cells showed that 3'ss disruption induced transcripts  $\Delta 16p_{44}$  and  $\Delta 16,17p_{69}$  (Whiley et al., 2011). Here we found both transcripts, but also other minor ones:  $\Delta 16p_{55}$ ,  $\Delta 16p_{93}$ ,  $\Delta 16$  (Table 3). Additionally, according to our data  $\Delta 16p_{44}$  is the main transcript (~91%) that other authors also identified but described as a minor transcript in agarose gels (Whiley et al., 2011). These differences could be due to: (i) the cell line; (ii) the use of cycloheximide to inhibit the NMD; (iii) the fact that we work with a single-mutant allele, avoiding the wt counterpart effect; and (iv) the high sensitivity of fluorescent capillary electrophoresis, which can detect rare transcripts versus agarose electrophoresis. In any case, both results show that c.7618-1G>A severely disrupted splicing. On the other hand, variant c.7805G>C was previously reported to result in  $\Delta 16$  and  $\Delta 16q_{100}$ , with the total absence of the canonical transcript (Bonnet et al., 2008). This outcome matches our results (Table 3):  $\Delta 16$  as the main transcript (~78%), followed by  $\Delta 16q_{100}$  (~14%), and the lack of the full-length transcript. It is also worthy to mention that we detected other minor transcripts due to the high sensitivity of fluorescent capillary electrophoresis ( $\Delta 16q_{20}$  at ~6.5% and  $\Delta 16,17p_{69}$  at ~1.5%) that otherwise could not be easily detected on agarose gels. In any case, the spliceogenic effects of variants c.7618-1G>A and c.7805G>C were supported by our data.

Variant c.7802A>G probably generated the most conflicting result since it triggered ~54% of canonical transcript and ~46% of  $\Delta 16q_4$ , so that its interpretation is more complex. The transcript  $\Delta 16q_4$ , caused by the use of a new 5'ss, generated a frameshift deletion and the protein truncation by a PTC 46 codons downstream (p.Y2601Wfs\*46). However, it is still unclear if ~54% of full-length transcript can preserve BRCA2 function, given that, for example, 20–30% of BRCA1 transcript is able to maintain BRCA1 activity (de la Hoya et al., 2016). It is also important to keep in mind that full-length transcript carries a missense variant (p.Y2601W) that, according to IARC alignment<sup>7</sup>, Tyrosine 2601 is highly conserved from human to sea urchin, suggesting an important function in the protein. Moreover, PolyPhen-2 (Adzhubei et al., 2010) predicted that this aminoacid change is damaging with the maximum score (1.0). Curiously, c.7802A>G was reported a family with a significant history of primary cancers (colorectal, lymphoma, and breast cancers) which carried biallelic BRCA2 mutations (c.7802A>G and c.1845\_1856delCT). However, patients did not present the typical FA phenotype, which suggested that p.Y2601W BRCA2 maintained at least enough BRCA2 activity to prevent early childhood FA features (Degrolard-Courcet et al., 2014). Nevertheless, this missense change remains classified as VUS in ClinVar<sup>8</sup>.

On the other hand, variant c.7625C>G was previously computed to disrupt one SRp55 motif (Pettigrew et al., 2008),

although functional mapping by microdeletions indicated that exon 16 is likely not regulated by splicing enhancers. Nevertheless, this change was selected because it presumably created new strong 3' and 5'ss as well, both with a NNSplice score >0.9 (Table 2). However, c.7625C>G only produced the full-length transcript without any splicing anomaly. The protein would even carry the missense variant p.T2542R. However, consistent with PolyPhen, this change might be considered as benign with a score of 0.0, which could be explained by the low conservation of the affected threonine. Anyway, further functional and association studies must be performed to interpret this variant. Other variant that resulted in a normal splicing pattern was the nonsense variant c.7738C>T (p.Q2580X), that *a priori* had been classified as pathogenic. In this case, the protein would be truncated at codon 2580 losing 839 aminoacids of the C-terminal where the DSS1-binding site, the DNA-binding domain, the RAD51C-binding site, and the cyclin-dependent kinase (CDK) phosphorylation site are located (Roy et al., 2012). Interestingly, this variant was found in an Italian non-Ashkenazi BRCA1 and BRCA2 double heterozygote family (Musolino et al., 2005).

According to the ACMG guidelines (Table 4; Richards et al., 2015), five variants (c.7618-2A>T, c.7618-2A>G, c.7618-1G>A, c.7618-1G>C, and c.7805+1G>A) can be classified as pathogenic as they match criteria PVS1 (very strong evidence of pathogenicity: null variant – nonsense, frameshift, canonical  $\pm 1$  or 2 ss, initiation codon, single or multiexon deletion – in a gene where LOF is a known mechanism of disease), PS3 (strong evidence: well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product), PM2 (moderate evidence: absent from controls in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium), PP3 (supporting evidence: multiple lines of computational evidence support a deleterious effect on the gene or gene product: conservation, evolutionary, splicing impact, etc.), and PP5 (reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation). On the other hand, variants c.7802A>G, c.7805G>C, and c.7805+3A>C were classified as likely pathogenic as they match criteria PS3, PM2, PP3, and PP5.

Similarly, following the ENIGMA rules for variant classification<sup>9</sup>, all variants, except for c.7802A>G, should be reclassified as class 4 (likely pathogenic) because they are “considered extremely likely to alter splicing based on position” and are “predicted bioinformatically to alter the use of the native donor/acceptor site.” Conversely, minigenes are not considered robust approaches to functionally test these variants yet (“... results from construct-based mRNA assays alone are not considered sufficiently robust to be used as evidence for variant classification ...”). However, this specific minigene with BRCA2 exons 14–20 was confirmed as a robust tool since it reproduced patient RNA results from eight variants (Fraile-Bethencourt et al., 2017), and also c.7618-1G>A and c.7805G>C of this study, so

<sup>7</sup>[http://agvgd.hci.utah.edu/BRCA2\\_Spur.html](http://agvgd.hci.utah.edu/BRCA2_Spur.html)

<sup>8</sup><https://www.ncbi.nlm.nih.gov/clinvar/variation/185651/#clinical-assertions>

<sup>9</sup><https://enigmaconsortium.org/library/general-documents/enigma-classification-criteria/>

**TABLE 4 |** Classification of variants according to the ENIGMA and ACMG rules.

| DNA variants | Main aberrant transcripts <sup>1</sup> | HGVS RNA/protein effects                                   | Previous classification <sup>2</sup> | ACMG classification <sup>3</sup>      | Enigma classification <sup>4</sup> |
|--------------|--|--|--------------------------------------|---------------------------------------|------------------------------------|
| c.7618-2A>T  | Δ16p <sub>44</sub>                     | r.7618_7661del/p.L2540Qfs*11                               | 5 – Causal                           | Pathogenic: PVS1, PS3, PM2, PP3, PP5  | Class 4                            |
| c.7618-2A>G  | Δ16p <sub>44</sub>                     | r.7618_7661del/p.L2540Qfs*11                               | 5 – Causal                           | Pathogenic: PVS1, PS3, PM2, PP3, PP5  | Class 4                            |
| c.7618-1G>A  | Δ16p <sub>44</sub>                     | r.7618_7661del/p.L2540Qfs*11                               | 5 – Causal                           | Pathogenic: PVS1, PS3, PM2, PP3, PP5  | Class 4                            |
| c.7618-1G>C  | Δ16p <sub>44</sub>                     | r.7618_7661del/p.L2540Qfs*11                               | 5 – Causal                           | Pathogenic: PVS1, PS3, PM2, PP3, PP5  | Class 4                            |
| c.7802A>G    | Δ16q <sub>4</sub>                      | r.7802_7805del/p.Y2601Wfs*46                               | 3 – VUS                              | Likely pathogenic: PS3, PM2, PP3, PP5 | Class 3                            |
| c.7805G>C    | Δ16, Δ16q <sub>100</sub>               | r.[7618_7805del; 7706_7805del]/p.[L2540Gfs*4; K2570Lfs*45] | 5 – Causal                           | Likely pathogenic: PS3, PM2, PP3, PP5 | Class 4                            |
| c.7805+1G>A  | Δ16, Δ16q <sub>100</sub>               | r.[7618_7805del; 7706_7805del]/p.[L2540Gfs*4; K2570Lfs*45] | 5 – Causal                           | Pathogenic: PVS1, PS3, PM2, PP3, PP5  | Class 4                            |
| c.7805+3A>C  | Δ16, Δ16q <sub>100</sub>               | r.[7618_7805del; 7706_7805del]/p.[L2540Gfs*4; K2570Lfs*45] | 3 – VUS                              | Likely pathogenic: PS3, PM2, PP3, PP5 | Class 4                            |

<sup>1</sup>Transcripts were annotated according to previous reports of the ENIGMA consortium (Colombo et al., 2014). Δ, skipping or deletion; ▼, insertion; p: alternative acceptor site; q, alternative donor site; subscript number, number of deleted nt; superscript number, number of inserted nt. Thus, Δ16 indicates exon 16 skipping and Δ16p<sub>44</sub> indicates loss of 44 nt at the exon 16 acceptor site (or new acceptor site 44 nt upstream). All transcripts and their quantification data are described in Table 2. <sup>2</sup>Previous classifications according to the BRCA share (c.7618-2A>T, c.7618-2A>G, c.7618-1G>A, c.7618-1G>C, c.7802A>G, and c.7805G>C) and the BIC mutation databases (c.7805+1G>A and c.7805+3A>C). <sup>3</sup>ACMG criteria: PVS1, null variant (nonsense, frameshift, canonical ± 1 or 2 ss, etc.) in a gene where LOF is a known mechanism of disease; PS3, well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product; PM2, absent from controls in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium; PP3, multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.); PP5, reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation. <sup>4</sup>ENIGMA. Class 4: Nucleotide positions that are considered extremely likely to alter splicing and/or variants are predicted bioinformatically to alter the use of the native donor/acceptor site. Class 3: "In the absence of clinical evidence to assign an alternative classification, variant allele tested for mRNA aberrations ... is found to produce mRNA transcript(s) predicted to encode intact full-length protein ..."

that these seven class 4 variants could be even reclassified as class 5. Finally, c.7802A>G was classified as class 3 because it did not meet the above standards and induce a partial aberrant outcome with more than 50% of the canonical transcript.

In summary, we detected eight spliceogenic *BRCA2* exon 16 variants that should be classified as pathogenic or likely pathogenic according to the ACMG guidelines (Table 4). Moreover, they account for 22% of causal variants of exon 16 and 11% of all recorded variants of this exon at the mutation databases. Taken together this and our previous studies, we have tested 283 *BRCA1/2* variants under the splicing perspective, 154 of which induced anomalous patterns and 111 could be classified as pathogenic or likely pathogenic. These data remark the importance of variants of splicing regulatory sequences, which are often underestimated because most of them are placed in non-coding regions of the protein. Until now, genetic family-based studies have set up the impact of some variants on cancer risk. However, because of the exponential increment in the number of variants, their low frequencies and different nature, functional assays are strictly required. In this context, minigene technology constitutes a robust tool which can be used to functionally test spliceogenic candidate variants of any disease-gene without the interference of the counterpart wt allele. Certainly, pSAD-based minigenes represented valuable tools to functionally check variants of the *SERPINA1* (severe alpha-1 antitrypsin deficiency) and *CHD7* (Charge Syndrome) genes (Lara et al., 2014; Villate et al., 2018). RNA assays provide essential data for the initial characterization

of VUS and improve the genetic counseling of hereditary diseases.

## AUTHOR CONTRIBUTIONS

EF-B contributed to the bioinformatics analysis, minigene construction, manuscript writing, and performed most of the splicing functional assays. BD-G and AV-P participated in minigene construction, mutagenesis experiments, and functional assays. AA participated in minigene construction and functional mapping experiments. EV conceived the study and the experimental design, supervised all the experiments, and wrote the manuscript. All authors contributed to data interpretation, revisions of the manuscript, and approved the final version of the manuscript.

## FUNDING

EV's lab was supported by grants from the Spanish Ministry of Economy and Competitiveness, Plan Nacional de I+D+I 2013–2016, ISCIII (Grants: PI13/01749 and PI17/00227) co-funded by FEDER from European Regional Development Funds (European Union), and grant CSI090U14 from the Consejería de Educación (ORDEN EDU/122/2014) and Junta de Castilla y León. EF-B was supported by a predoctoral fellowship from the University of Valladolid and Banco Santander (2015–2019).

## ACKNOWLEDGMENTS

We acknowledge the support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

## REFERENCES

- Acedo, A., Hernández-Moro, C., Curiel-García, Á., Díez-Gómez, B., and Velasco, E. A. (2015). Functional classification of *BRCA2* DNA variants by splicing assays in a large minigene with 9 exons. *Hum. Mutat.* 36, 210–221. doi: 10.1002/humu.22725
- Acedo, A., Sanz, D. J., Durán, M., Infante, M., Pérez-Cabornero, L., Miner, C., et al. (2012). Comprehensive splicing functional analysis of DNA variants of the *BRCA2* gene by hybrid minigenes. *Breast Cancer Res.* 14:R87. doi: 10.1186/bcr3202
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Beroud, C., Letovsky, S. I., Braastad, C. D., Caputo, S. M., Beaudoux, O., Bignon, Y. J., et al. (2016). BRCA share: a collection of clinical BRCA gene variants. *Hum. Mutat.* 37, 1318–1328. doi: 10.1002/humu.23113
- Bonnet, C., Krieger, S., Vezain, M., Rousselin, A., Tournier, I., Martins, A., et al. (2008). Screening *BRCA1* and *BRCA2* unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *J. Med. Genet.* 45, 438–446. doi: 10.1136/jmg.2007.056895
- Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 31, 3568–3571. doi: 10.1093/nar/gkg616
- Colombo, M., Blok, M. J., Whitley, P., Santamariña, M., Gutiérrez-Enríquez, S., Romero, A., et al. (2014). Comprehensive annotation of splice junctions supports pervasive alternative splicing at the *BRCA1* locus: a report from the ENIGMA consortium. *Hum. Mol. Genet.* 23, 3666–3680. doi: 10.1093/hmg/ddu075
- de la Hoya, M., Soukari, O., López-Perolio, I., Vega, A., Walker, L. C., van Ierland, Y., et al. (2016). Combined genetic and splicing analysis of *BRCA1* c.[594-2A > C; 641A > G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum. Mol. Genet.* 25, 2256–2268. doi: 10.1093/hmg/ddw094
- Degrolard-Courcet, E., Sokolowska, J., Padeano, M. M., Guiu, S., Bronner, M., Chery, C., et al. (2014). Development of primary early-onset colorectal cancers due to biallelic mutations of the *FANCD1/BRCA2* gene. *Eur. J. Hum. Genet.* 22, 979–987. doi: 10.1038/ejhg.2013.278
- Desmet, F.-O., Hamroun, D., Lalonde, M., Collod-Bérout, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 37:e67. doi: 10.1093/nar/gkp215
- Douglas, A. G., and Wood, M. J. (2011). RNA splicing: disease and therapy. *Brief Funct. Genomics* 10, 151–164. doi: 10.1093/bfpg/elfr020
- Fackenthal, J. D., Yoshimatsu, T., Zhang, B., de Garibay, G. R., Colombo, M., De Vecchi, G., et al. (2016). Naturally occurring *BRCA2* alternative mRNA splicing events in clinically relevant samples. *J. Med. Genet.* 53, 548–558. doi: 10.1136/jmedgenet-2015-103570
- Fairbrother, W. G., Holste, D., Burge, C. B., and Sharp, P. A. (2004). Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol.* 2:E268. doi: 10.1371/journal.pbio.0020268
- Fairbrother, W. G., Yeh, R.-F., Sharp, P. A., and Burge, C. B. (2002). Predictive identification of exonic splicing enhancers in human genes. *Science* 297, 1007–1013. doi: 10.1126/science.1073774
- Fraile-Bethencourt, E., Díez-Gómez, B., Velásquez-Zapata, V., Acedo, A., Sanz, D. J., and Velasco, E. A. (2017). Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of *BRCA2* exons 17 and 18. *PLoS Genet.* 13:e1006691. doi: 10.1371/journal.pgen.1006691
- Grodecká, L., Buratti, E., and Freiberger, T. (2017). Mutations of pre-mRNA splicing regulatory elements: Are predictions moving forward to clinical diagnostics? *Int. J. Mol. Sci.* 18:E1668. doi: 10.3390/ijms18081668
- Hussain, S., Wilson, J. B., Medhurst, A. L., Hejna, J., Witt, E., Ananth, S., et al. (2004). Direct interaction of *FANCD2* with *BRCA2* in DNA damage response pathways. *Hum. Mol. Genet.* 13, 1241–1248. doi: 10.1093/hmg/ddh135
- Infante, M., Durán, M., Acedo, A., Sánchez-Tapia, E. M., Díez-Gómez, B., Barroso, A., et al. (2013). The highly prevalent *BRCA2* mutation c.2808\_2811del (3036delACAA) is located in a mutational hotspot and has multiple origins. *Carcinogenesis* 34, 2505–2511. doi: 10.1093/carcin/bgt272
- Lara, B., Martínez, M. T., Blanco, I., Hernández-Moro, C., Velasco, E. A., Ferrarotti, I., et al. (2014). Severe alpha-1 antitrypsin deficiency in composite heterozygotes inheriting a new splicing mutation QOMadrid. *Respir. Res.* 15:125. doi: 10.1186/s12931-014-0125-y
- Levy-Lahad, E., Catane, R., Eisenberg, S., Kaufman, B., Hornreich, G., Lishinsky, E., et al. (1997). Founder *BRCA1* and *BRCA2* mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. *Am. J. Hum. Genet.* 60, 1059–1067.
- Li, J., Zou, C., Bai, Y., Wazer, D. E., Band, V., and Gao, Q. (2006). DSS1 is required for the stability of *BRCA2*. *Oncogene* 25, 1186–1194. doi: 10.1038/sj.onc.1209153
- López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 579, 1900–1903. doi: 10.1016/j.febslet.2005.02.047
- Marston, N. J., Richards, W. J., Hughes, D., Bertwistle, D., Marshall, C. J., and Ashworth, A. (1999). Interaction between the product of the breast cancer susceptibility gene *BRCA2* and DSS1, a protein functionally conserved from yeast to mammals. *Mol. Cell. Biol.* 19, 4633–4642. doi: 10.1128/MCB.19.7.4633
- Musolino, A., Naldi, N., Michiara, M., Bella, M. A., Zanelli, P., Bortesi, B., et al. (2005). A breast cancer patient from Italy with germline mutations in both the *BRCA1* and *BRCA2* genes. *Breast Cancer Res. Treat.* 91, 203–205. doi: 10.1007/s10549-004-7704-4
- Nielsen, F. C., van Overeem Hansen, T., and Sørensen, C. S. (2016). Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat. Rev. Cancer* 16, 599–612. doi: 10.1038/nrc.2016.72
- Petrucelli, N., Daly, M. B., and Feldman, G. L. (2013). *BRCA1 and BRCA2 Hereditary Breast and Ovarian Cancer*. Seattle, WA: University of Washington.
- Pettigrew, C. A., Wayne, N., Wronski, A., Lovelock, P. K., Spurdle, A. B., and Brown, M. A. (2008). Colocalisation of predicted exonic splicing enhancers in *BRCA2* with reported sequence variants. *Breast Cancer Res. Treat.* 110, 227–234. doi: 10.1007/s10549-007-9714-5
- Reese, M. G., Eeckman, F. H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. *J. Comput. Biol.* 4, 311–323. doi: 10.1089/cmb.1997.4.311
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (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, 405–423. doi: 10.1038/gim.2015.30
- Ricks, T. K., Chiu, H.-J., Ison, G., Kim, G., McKee, A. E., Kluetz, P., et al. (2015). Successes and challenges of PARP inhibitors in cancer therapy. *Front. Oncol.* 5:222. doi: 10.3389/fonc.2015.00222
- Roy, R., Chun, J., and Powell, S. N. (2012). *BRCA1* and *BRCA2*: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* 12, 68–78. doi: 10.1038/nrc3181
- Sanz, D. J., Acedo, A., Infante, M., Durán, M., Pérez-Cabornero, L., Esteban-Cardesa, E., et al. (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, 1957–1967. doi: 10.1158/1078-0432.CCR-09-2564
- Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G. P., Bresolin, N., et al. (2004). Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res.* 32, 1783–1791. doi: 10.1093/nar/gkh341

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00188/full#supplementary-material>

- Slavin, T. P., Niell-Swiler, M., Solomon, I., Nehoray, B., Rybak, C., Blazer, K. R., et al. (2015). Clinical application of multigene panels: challenges of next-generation counseling and cancer risk management. *Front. Oncol.* 5:208. doi: 10.3389/fonc.2015.00208
- Soemedi, R., Cygan, K. J., Rhine, C. L., Glidden, D. T., Taggart, A. J., Lin, C.-L., et al. (2017). The effects of structure on pre-mRNA processing and stability. *Methods* 125, 36–44. doi: 10.1016/j.ymeth.2017.06.001
- Villate, O., Ibarluzea, N., Fraile-bethencourt, E., and Valenzuela, A. (2018). Functional analyses of a novel splice variant in the CHD7 gene, found by next generation sequencing, confirm its pathogenicity in a Spanish patient and diagnose him with CHARGE syndrome. 9, 26–31. doi: 10.3389/fgene.2018.00007
- Wang, G.-S., and Cooper, T. A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* 8, 749–761. doi: 10.1038/nrg2164
- Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., and Burge, C. B. (2004). Systematic identification and analysis of exonic splicing silencers. *Cell* 119, 831–845. doi: 10.1016/j.cell.2004.11.010
- Whiley, P. J., Guidugli, L., Walker, L. C., Healey, S., Thompson, B. A., Lakhani, S. R., et al. (2011). Splicing and multifactorial analysis of intronic BRCA1 and BRCA2 sequence variants identifies clinically significant splicing aberrations up to 12 nucleotides from the intron/exon boundary. *Hum. Mutat.* 32, 678–687. doi: 10.1002/humu.21495
- Will, C. L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3:a003707. doi: 10.1101/cshperspect.a003707
- Yeo, G., and Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* 11, 377–394. doi: 10.1089/1066527041410418
- Zhang, M. (1998). Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.* 7, 919–932. doi: 10.1093/hmg/7.5.919
- Zhang, X. H., and Chasin, L. A. (2004). Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* 18, 1241–1250. doi: 10.1101/gad.1195304

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Fraile-Bethencourt, Valenzuela-Palomo, Díez-Gómez, Acedo and Velasco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

**Supplementary Table S1. Bioinformatics analysis of the 72 variants from the BIC and BRCA Share databases.**

| BIC/UMD variants <sup>1</sup> | NNSplice <sup>2</sup>                      | MaxEnt <sup>2</sup> | ESE Finder <sup>3</sup>                                 | ESE HSF <sup>3</sup>                         | hnRNPA1 <sup>3</sup> |
|-------------------------------|--|---------------------|---|--|----------------------|
| c.7618-114_7618-112delAGT     |  |                     | [-]SRp40 (80.78)  |  | [-]hnRNPA1 (73.10)   |
| c.7618-59G>T                  |  |                     | [-]SRp55 (74.82)  |  |                      |
| c.7618[-]27T>A                |  |                     |   |  |                      |
| c.7618[-]16T>G                |  |                     | [-]SRp55 (74.82)  |  |                      |
| <b>c.7618[-]2A&gt;T</b>       | [-] 3'SS: 0,69 → <0,4                      | 3'ss (7.11→[-]1.24) | [-]SRp55 (83.78)  |  | [-]hnRNPA1 (70.24)   |
| <b>c.7618[-]2A&gt;G</b>       | [-] 3'SS: 0,69 → <0,4                      | 3'ss (7.11→[-]0.83) |   |  | [-]hnRNPA1 (70.24)   |
| <b>c.7618[-]1G&gt;A</b>       | [-] 3'SS: 0,69 → <0,4                      | 3'ss (7.11→[-]1.63) | [-]SRp55 (77.06)  |  | [-]hnRNPA1 (70.24)   |
| <b>c.7618[-]1G&gt;C</b>       | [-] 3'SS: 0,69 → <0,4                      | 3'ss (7.11→[-]0.94) | [-]SRp55 (77.06)  |  | [-]hnRNPA1 (70.24)   |
| c.7621T>C                     |  |                     | [+]SF2/ASF (IgM[-]BRCA1) (80.08);<br>[+]SF2/ASF (84.51) |  |                      |
| <b>c.7625C&gt;G</b>           | [+] 3'SS (<0,4→0,94); [+] 5'SS (<0,4→0,91) | +5'ss (2.23→7.37)   | [-]SRp55 (89.80)  | [-]9G8 (62.28)                               | [+]hnRNPA1 (88.33)   |
| c.7626G>A                     |  |                     |   |  | [-]hnRNPA1 (69.05)   |
| c.7627_7628insT               | 3'SS (0,69→0.59)                           |                     | [-]SRp55 (77.06)  |  |                      |
| c.7630G>C                     |  |                     | [+]SRp55 (74.37); [+]SC35 (78.93)                       |  | [-]hnRNPA1 (78.81)   |
| c.7631delG                    |  |                     | [+]SC35 (75.17)   |  |                      |
| c.7632C>T                     |  |                     | [-]SRp55 (75.17)  |  |                      |
| c.7633G>A                     | 3'SS (0,69→0.75)                           |                     | [-]SC35 (75.17)   |  |                      |
| c.7636_7645del                |  |                     | [-]SC35 (75.17)   | [-]Tra2B (81.02)                             |                      |
| c.7639A>G                     |  |                     | [+]SC35(78.12)  | [+]9G8 (65.17)                               |                      |
| c.7643A>G                     |  |                     |   | [-]Tra2B (81.02)                             |                      |
| c.7647C>A                     |  |                     |   | [+]Tra2B (98.53); [+]9G8 (67.38)             |                      |
| c.7648A>G                     |  |                     |   |  |                      |
| c.7651A>C                     |  |                     |   | [-]2 Tra2B (94.14, 81.02);<br>[+]9G8 (62.28) |                      |
| c.7654dup                     |  |                     |   | [+] Tra2B (81.02)                            |                      |
| c.7655_7658delTTAA            |  |                     | [-]SRp40 (81.38)  |  |                      |
| c.7658_7660delACA             |  |                     | [-]SRp40 (81.38)  | [-] Tra2B (83.21)                            | [+]hnRNPA1 (76.43)   |
| c.7659delC                    |  |                     |   | [-] Tra2B (83.21)                            | [+]hnRNPA1 (69.53)   |

|                     |  |  |   |                                   |                    |
|---------------------|--|--|---|-----------------------------------|--------------------|
| c.7661G>A           |  |  | [-]SRp40 (81.38)  | [+] Tra2B (94.14)                 |                    |
| c.7669G>A           |  |  |   | [+] Tra2B (85.42); [-]9G8 (67.72) | [+]hnRNPA1 (69.29) |
| c.7671delA          |  |  |   | [-] 9G8 (67.72)                   | [-]hnRNPA1 (72.38) |
| c.7673_7674delAG    |  |  | [-]SRp55 (84.62)  |                                   | [-]hnRNPA1 (72.38) |
| c.7673delA          |  |  |   | [-] 9G8 (67.72)                   |                    |
| c.7679_7680delTT    |  |  | [+]SC35 (83.65)   |                                   |                    |
| c.7680delT          |  |  |   |                                   |                    |
| c.7680dup           |  |  |   |                                   |                    |
| c.7681C>T           |  |  |   |                                   | [+]hnRNPA1 (72.14) |
| c.7681dup           |  |  |   |                                   |                    |
| c.7683G>A           |  |  | [-]SC35 (76.58)   |                                   |                    |
| c.7684T>C           |  |  | [+] SRp40 (86.77); [-]SC35 (76.58)                                      |                                   |                    |
| c.7689delC          |  |  | [-]SC35 (76.58); [-]SF2/ASF (IgM[-]BRCA1) (82.84); [-]SF2/ASF (83.40)   |                                   |                    |
| c.7696G>T           |  |  | [-] SF2/ASF (IgM[-]BRCA1) (76.85); [-]SF2/ASF (77.34); [+]SRp55 (76.10) | [-]9G8 (62.15)                    |                    |
| c.7704delT          |  | 5'ss ([-]29.63→8.92); 5'ss (8.92 → [-]11.56) |   |                                   |                    |
| c.7712A>G           |  | 5'ss (2.81 → 3.35)                           |   |                                   |                    |
| c.7715G>A           |  |  |   | [+]9G8 (66.38); [+]Tra2B (81.02)  |                    |
| c.7719_7720insA     |  |  | [+]SC35 (80.39); [-]SRp55 (74.69)                                       |                                   |                    |
| c.7721G>A           |  |  | [-]SRp40 (80.96); [-]SRp55 (74.69)                                      | [+]9G8 (61.07)                    | [+]hnRNPA1 (77.62) |
| c.7731delA          |  |  | [+]SF2/ASF (72.79)  | [-]Tra2B (83.21); [-]9G8(61.34)   |                    |
| <b>c.7738C&gt;T</b> |  |  | [-]SC35 (81.01)   |                                   | [+]hnRNPA1 (70.24) |
| c.7742T>G           |  |  | [+]SRp40 (81.92); [+]SF2/ASF (IgM[-]BRCA1) (77.15); [+]SF2/ASF (76.47)  |                                   | [+]hnRNPA1 (66.67) |
| c.7751G>A           |  |  |   |                                   |                    |
| <b>c.7753G&gt;A</b> |  | 5'ss ([+]-1.33 → 3.06)                       |   |                                   | [+]hnRNPA1 (74.05) |
| c.7757G>A           |  |  |   |                                   |                    |
| c.7758G>A           |  |  |   |                                   | [-]hnRNPA1 (67.38) |

|                       |  |                     |                                   |                                   |                             |
|-----------------------|--|---------------------|-----------------------------------|-----------------------------------|-----------------------------|
| c.7759C>T             |  |                     | [+]SRp55 (75.08)                  |                                   |                             |
| c.7761delC            |  |                     | [-]SC35 (88.01); [-]SRp40 (78.80) |                                   |                             |
| c.7762delA            |  |                     | [-]SC35 (88.01); [-]SRp40 (78.80) |                                   |                             |
| c.7764A>T             |  |                     | [-]SRp40 (78.80)                  |                                   |                             |
| c.7766C>A             |  |                     | [+]SF2/ASF (IgM[-]BRCA1) (72.31)  |                                   |                             |
| <b>c.7772A&gt;G</b>   |  |                     | [+]SF2/ASF (IgM[-]BRCA1) (74.85)  | [-]9G8 (60.94)                    | [+]hnRNPA1 (69.52)          |
| c.7781A>G             |  |                     |                                   |                                   |                             |
| c.7783G>C             |  |                     | [-]SC35 (78.49)                   |                                   | [-]2 hnRNPA1 (65.72; 73.81) |
| c.7783G>T             |  |                     | [-]SC35 (78.49); [+]SRp40 (81.02) | [-]9G8 (62.28)                    | [-]2 hnRNPA1 (65.72; 73.81) |
| c.7786G>A             |  |                     |                                   |                                   | [+]hnRNPA1 (80.24)          |
| c.7791_7792insA       |  |                     |                                   | [+]Tra2B (100.00)                 |                             |
| c.7795_7797delGAA     |  |                     |                                   | [-]Tra2B (100.00); [-]9G8 (87.92) |                             |
| c.7795G>T             |  |                     |                                   |                                   |                             |
| <b>c.7802A&gt;G</b>   |  |                     | [-]SRp40 (83.89)                  |                                   | [-]hnRNPA1 (66.43)          |
| <b>c.7805G&gt;C</b>   |  | 5'ss (4.68→[-]2.79) | [+]SRp55 (89.80)                  |                                   | [-]hnRNPA1 (69.05)          |
| <b>c.7805+1G&gt;A</b> |  | 5'ss (4.68→[-]3.50) | [-]SRp40 (83.89); [-]SC35 (82.36) |                                   |                             |
| <b>c.7805+3A&gt;C</b> |  | 5'ss (4.68→[-]3.85) |                                   |                                   |                             |
| c.7805+6C>G           |  | 5'ss (4.68→5.90)    |                                   |                                   |                             |
| c.7805+7T>A           |  |                     | [+]SRp40 (84.01)                  |                                   | [-]hnRNPA1 (70.72)          |
| c.7805+8A>G           |  |                     |                                   |                                   | [-]hnRNPA1 (70.72)          |

<sup>1</sup> Selected variants are shown in bold type

<sup>2</sup> Bioinformatics: [+] and [-] symbols indicate creation or disruption of splicing motifs, respectively. Thresholds of the splicing programs:

**Splice Sites** (ss): NNSPLICE (values 0-1): Cut-offs=0.4 for both 5' and 3' splice sites (ss disruption <0.4, ss creation >0.4) ; MaxEnt: 3.0 for 5' and 3' splice sites (variation threshold ±30% according to HSF).

<sup>3</sup> **Enhancers and Silencers** HSF: Human Splicing Finder matrices (default values, <http://www.umd.be/HSF3/technicaltips.html>): ESEfinder cut-offs (HSF scale, normalized to 0-100): SF2/ASF: 72.98 / SF2/ASF (IgM-BRCA1): 70.51 / SRp40: 78.08 / SC35: 75.05 / SRp55: 73.86 ; ESE motifs from HSF, cut-offs values (0-100): Tra2: 75.964 / 9G8: 59.245; hnRNP motifs: hnRNPA1: 65.476 (these values were considered the limits for the disruption or creation of a splicing regulatory element).



## ARTÍCULO 4

Eugenia Fraile-Bethencourt, Alberto Valenzuela-Palomo, Beatriz Díez-Gómez, Elisa Goina, Alberto Acedo, Emanuele Buratti, Eladio A. Velasco.

**Mis-splicing in breast cancer: identification of pathogenic *BRCA2* variants by systematic** minigene assays.

Journal of Pathology 2019; <https://doi.org/10.1002/path.5268>

*Impact factor 2017: 6.253*


**Introducción:** A menudo encontramos variantes germinales que actúan inactivando genes asociados a enfermedad a través de la alteración del *splicing*. Con el objetivo de estudiar el papel de la desregulación del *splicing* de *BRCA2* en el Cáncer de Mama y Ovario Hereditario, se ha realizado un análisis exhaustivo de las variantes comprendidas entre los exones 2 y 9 de *BRCA2*, así como la caracterización inicial de los mecanismos reguladores en el reconocimiento de dichos exones.

**Resultados:** El minigen con los exones 2-9 de *BRCA2* fue construido en pSAD (Patente P201231427, CSIC) y validado funcionalmente. Los resultados en células MCF-7 mostraron un transcrito con el tamaño y la estructura esperada (1.016 nt / V1-BRCA2\_exones\_2-9-V2). Las variantes de DNA recopiladas de las bases de datos fueron analizadas con los programas NNSplice y *Human Splicing Finder*. Con el objetivo de refinar las predicciones bioinformáticas, se mapearon las regiones reguladoras mediante la introducción de 26 microdeleciones exónicas en el minigen y el análisis funcional de las mismas en MCF-7. Así se identificaron nueve intervalos ricos en ESEs (*Exonic Splicing Enhancers*) donde era probable la presencia de variantes espliceogénicas. A través de los datos obtenidos mediante el estudio bioinformático y el ensayo de microdeleciones, se seleccionaron 83 variantes que fueron introducidas en el minigen mediante mutagénesis dirigida. La mayoría (53 variantes) alteraban el *splicing*: Veintiocho variantes afectaban a los sitios canónicos de *splicing*, cuatro creaban nuevos sitios de *splicing*, diez eliminaban *enhancers*, ocho creaban silenciadores y tres causaban un doble efecto. Es importante

señalar que nueve de las variantes espliceogénicas se encontraban en los intervalos con ESEs detectados a través de las microdeleciones. Los datos de secuenciación y electroforesis capilar mostraron más de 23 transcritos aberrantes, siendo la pérdida del exón o *exon skipping* el resultado más frecuente. La variante c.67G>A provocó el uso de un donador GC críptico en el exón 2, situado 4 nucleótidos *upstream* del donador natural. Las 36 variantes que resultaron en una alteración grave del *splicing* (>60% de transcritos aberrantes) fueron analizadas según las normas de ACMG (*American College of Medical Genetics and Genomics*). Veintiocho variantes fueron clasificadas como patogénicas, cinco como probablemente patogénicas y tres como variantes de significado clínico desconocido (VUS). Cabe destacar que trece de las variantes fueron reclasificadas desde VUS a patogénicas o probablemente patogénicas.

**Conclusión:** La mayoría de las variantes de *BRCA2* analizadas provocaban anomalías en el *splicing* (~64%), apoyando la idea de la alteración del *splicing* como un mecanismo común asociado al desarrollo de enfermedades. El mapeo funcional de elementos reguladores mediante microdeleciones representa una estrategia óptima para mejorar la baja precisión de los algoritmos de predicción e identificar regiones susceptibles a contener variantes espliceogénicas. Finalmente, los estudios funcionales mediante minigenes han demostrado ser una herramienta válida para la caracterización inicial anomalías de *splicing* y la interpretación clínica de variantes.

# Mis-splicing in breast cancer: identification of pathogenic *BRCA2* variants by systematic minigene assays

Eugenia Fraile-Bethencourt<sup>1</sup>, Alberto Valenzuela-Palomo<sup>1</sup>, Beatriz Díez-Gómez<sup>1</sup>, Elisa Goina<sup>2</sup>, Alberto Acedo<sup>†</sup>, Emanuele Buratti<sup>2</sup> and Eladio A Velasco<sup>1\*</sup> 

<sup>1</sup> Splicing and Genetic Susceptibility to Cancer, Instituto de Biología y Genética Molecular (CSIC-UVA), Valladolid, Spain

<sup>2</sup> Molecular Pathology Group, International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

\*Correspondence to: EA Velasco, Grupo de Splicing y Cáncer, Instituto de Biología y Genética Molecular (IBGM), Consejo Superior de Investigaciones Científicas (CSIC)-Uva, Sanz y Forés 3, 47003 Valladolid, Spain. E-mail: eavelsam@ibgm.uva.es

<sup>†</sup>Present address: AC-GEN Reading Life SL, Valladolid, Spain.

## Abstract

Splicing disruption is a common mechanism of gene inactivation associated with germline variants of susceptibility genes. To study the role of *BRCA2* mis-splicing in hereditary breast/ovarian cancer (HBOC), we performed a comprehensive analysis of variants from *BRCA2* exons 2–9, as well as the initial characterization of the regulatory mechanisms of such exons. A pSAD-based minigene with exons 2–9 was constructed and validated in MCF-7 cells, producing the expected transcript (1016-nt/V1-*BRCA2*\_exons\_2–9-V2). DNA variants from mutational databases were analyzed by NNSplice and Human Splicing Finder softwares. To refine ESE-variant prediction, we mapped the regulatory regions through a functional strategy whereby 26 exonic microdeletions were introduced into the minigene and tested in MCF-7 cells. Thus, we identified nine spliceogenic ESE-rich intervals where ESE-variants may be located. Combining bioinformatics and microdeletion assays, 83 variants were selected and genetically engineered in the minigene. Fifty-three changes impaired splicing: 28 variants disrupted the canonical sites, four created new ones, 10 abrogated enhancers, eight created silencers and three caused a double-effect. Notably, nine spliceogenic-ESE variants were located within ESE-containing intervals. Capillary electrophoresis and sequencing revealed more than 23 aberrant transcripts, where exon skipping was the most common event. Interestingly, variant c.67G>A triggered the usage of a noncanonical GC-donor 4-nt upstream. Thirty-six variants that induced severe anomalies (>60% aberrant transcripts) were analyzed according to the ACMG guidelines. Thus, 28 variants were classified as pathogenic, five as likely pathogenic and three as variants of uncertain significance. Interestingly, 13 VUS were reclassified as pathogenic or likely pathogenic variants. In conclusion, a large fraction of *BRCA2* variants (~64%) provoked splicing anomalies lending further support to the high prevalence of this disease-mechanism. The low accuracy of ESE-prediction algorithms may be circumvented by functional ESE-mapping that represents an optimal strategy to identify spliceogenic ESE-variants. Finally, systematic functional assays by minigenes depict a valuable tool for the initial characterization of splicing anomalies and the clinical interpretation of variants. Copyright © 2019 Pathological Society of Great Britain and Ireland. Published by John Wiley & Sons, Ltd.

**Keywords:** Breast cancer; *BRCA2*; DNA variants; VUS; splicing; hybrid minigenes

Received 14 November 2018; Revised 21 February 2019; Accepted 11 March 2019

No conflicts of interest were declared.

## Introduction

Hereditary breast and ovarian cancer (HBOC) is a highly heterogeneous genetic disease with more than 20 known or proposed susceptibility genes [1,2], of which *BRCA1* (MIM#113705) and *BRCA2* (MIM#600185) accumulate a relevant fraction of pathogenic mutations [3]. Despite the presence of several recurrent mutations [4], there have been reported thousands of different *BRCA1/2* changes in the databases. However, up to 20% of *BRCA1/2* tests report variants of uncertain clinical significance (VUS) [5]. These pose a challenge for clinical genetics testing, since VUS-carrier families are usually considered as having a negative result,

when risk assessment is solely based on family history [6]. Consequently, identification of pathogenic variants will contribute to breast cancer prevention, therapy, and surveillance as well as to a better knowledge of the breast cancer genetics [7,8].

Variants are usually classified attending to their impact on the protein translated, but disruption of other gene expression steps may occur, including transcription initiation, miRNA and long noncoding RNA regulation, RNA stability or post-transcriptional modifications [9–12]. Splicing is a gene expression step whereby introns are excised and exons are joined sequentially. This process is finely regulated by the spliceosome, which consists of five small nuclear ribonucleoproteins

and more than 150 additional proteins, and a wide array of cis-elements that include the conserved 5' and 3' splice sites, as well as exonic or intronic splicing enhancers and silencers that promote or repress, respectively, exon inclusion [13]. Variants at these motifs may induce abnormal processing events such as exon skipping, alternative 5' or 3' site usage or intron retention events, causing, therefore, the introduction of premature termination codons (PTCs) or the deletion of critical protein domains. However, the lack of a deep knowledge of the splicing code and its regulation, the inaccuracy of *in silico* predictions and the unpredictability of splicing outcomes make functional assays indispensable for checking potential splicing variants [14,15]. Furthermore, the characterization of physiological alternative splicing of any disease-gene is essential to assess the biological significance of the splicing impact of a particular DNA variant. Thus, a total of 63 and 24 alternative splicing events in *BRCA1* and *BRCA2*, respectively, had been previously described [16,17].

So far, different approaches have been used to study splicing variants, where patient RNA analysis is considered as the most suitable method [18,19]. Nevertheless, patient samples are not always available, thus minigenes offer an alternative strategy to evaluate the impact of variants and represent powerful tools to study the regulation of splicing [20,21]. The splicing reporter vector pSAD (Patent P201231427-CSIC) formed the backbone of the largest ever reported *BRCA2* minigenes with exons 19–27 and 14–20 [20,22] and minigenes of other disease-responsible genes, such as *SERPINA1* and *CHD7* [23,24]. To date, we have functionally tested 200 different DNA variants of the *BRCA* genes, 101 of which (~50.5%) impaired splicing.

We have herein studied *BRCA2* exons 2–9 with a new minigene (MGBR2\_ex2–9). We identified regulatory splicing cis-elements by functional mapping of overlapping exonic microdeletions and located ESE-rich intervals. Then, we selected and functionally assayed 83 candidate splicing variants. Finally, we made a tentative classification of 36 spliceogenic variants following the guidelines of the American College of Medical Genetics and Genomics (ACMG) [25] and the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA; <https://enigmaconsortium.org/>).

## Materials and methods

### Ethical approval

Ethical approval for this study was obtained from the Ethics Review Committee of the Hospital Universitario Río Hortega de Valladolid (6 November 2014).

### Databases and bioinformatics

*BRCA2* variants were obtained from the BIC and the UMD mutation databases (date last accessed 1

April 2017; <https://research.nhgri.nih.gov/projects/bic/Member/index.shtml>; <http://www.umd.be/BRCA2/>) [26]. To identify potential splicing variants, mutant and WT sequences were analyzed with NNSPLICE v0.9 [27] and Human Splicing Finder 3.0 [28].

### Minigene construction and mutagenesis

Exons 2–9 and their flanking intronic sequences were amplified with Phusion-High Fidelity polymerase (Fisher Scientific, Waltham, MA, USA) (see supplementary material, Table S1). Inserts were cloned into the splicing vector pSAD [20] in five steps: exons 4–7 + exon 3 + exon 8 + exon 9 + exon 2 (Figure 1). The WT minigene MGBR2\_ex2–9 was genetically engineered to introduce 26 exonic microdeletions and 83 reported DNA variants (see supplementary material, Table S2) with the QuikChange Lightning kit (Agilent, Santa Clara, CA, USA). All constructs were confirmed by sequencing (Macrogen, Madrid, Spain). An outline of the splicing functional assay protocol is shown in supplementary material, Figure S1.

### Transfection of eukaryotic cells

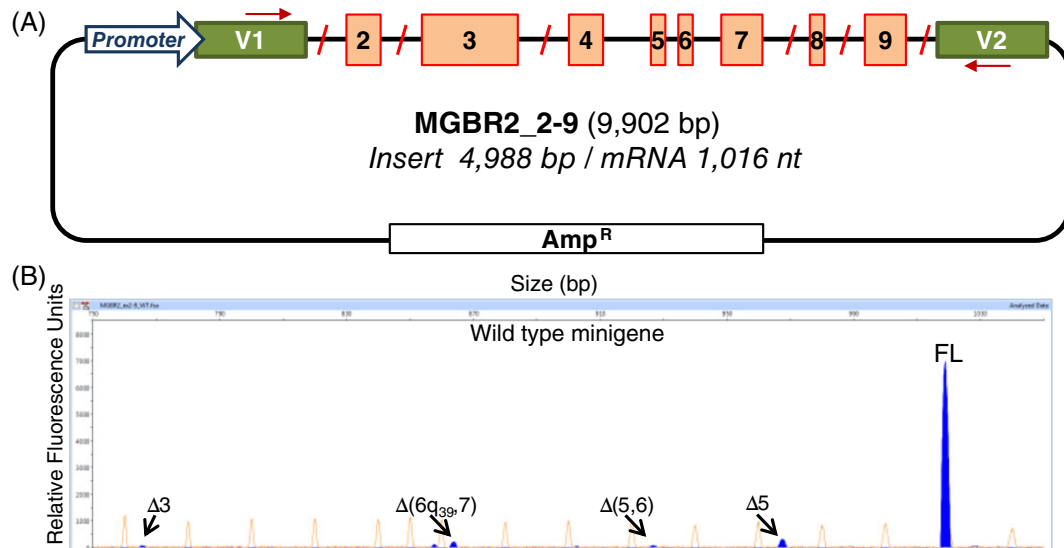
Approximately  $2 \times 10^5$  MCF-7 cells were grown to 90% confluency in 0.5 ml of medium (MEM, 10% FBS, 1% nonessential amino acids, 2 mM glutamine and 1% penicillin/streptomycin stock solution) in 4-well plates (Nunc, Roskilde, Denmark). Cells were transfected with 1 µg of minigene using 2 µl of Lipofectamine LTX (Life Technologies, Carlsbad, CA, USA). To inhibit nonsense mediated decay (NMD), a 4-h incubation with cycloheximide 300 µg/ml (Sigma-Aldrich, St. Louis, MO, USA) was performed. RNA was purified using the Genomatrix Universal RNA Purification Kit (EURx, Gdansk, Poland) including on-column DNase I digestion.

### siRNA experiments

Seven SR proteins (see supplementary material, Table S3) were inhibited by specific siRNAs using anti-Luciferase siRNA as negative control. Approximately  $1.5 \times 10^5$  cells in six-well plates (Nunc; Thermo Fisher Scientific, Waltham, MA, USA) were transfected with siRNA/Oligofectamine in Optimem medium (Life Technologies), and then, the minigene was introduced by Lipofectamine LTX. Silencing was confirmed by qPCR on a 7500 real time PCR system using 10 ng of cDNA (final volume of 25 µl) and SG qPCR Master Mix (EURx) following manufacturer's instructions.

### Reverse transcription polymerase chain reaction

Reverse transcription was carried out with 400 ng of RNA and the RevertAid First Strand cDNA Synthesis Kit (Life Technologies), using the vector-specific primer RTPSPL3-RV (5'-TGAGGAGTGAATTGGTTCGAA-3'). Samples were incubated at 42 °C for 1 h, followed by 5 min



**Figure 1.** Structure and functional analysis of the minigene MGBR2\_2–9. (A) Structure of the minigene MGBR2\_2–9 (slashes indicate shortened introns): / *ivs1* (252 bp) – *ex2* (106 bp) – *ivs2* (357 bp) / cloning tail / *ivs2* (210 bp) – *ex3* (249 bp) – *ivs3* (277 bp) / cloning tail / *ivs3* (494 bp) – *ex4* (109 bp) – *ivs4* (916 bp) – *ex5* (50 bp) – *ivs5* (91 bp) – *ex6* (41 bp) – *ivs6* (216 bp) – *ex7* (115 bp) – *ivs7* (296 bp) / cloning tail / *ivs7* (296 bp) – *ex8* (50 bp) – *ivs8* (190 bp) / cloning tail / *ivs8* (218 bp) – *ex9* (112 bp) – *ivs9* (295 bp). (B) Splicing functional analysis of MGBR2\_2–9 in MCF-7 cells. Transcripts are shown as blue peaks and the Genescan Liz-1200 size standard is shown as orange/faint peaks.

at 70 °C. Then, 40 ng of cDNA (final volume of 50 µl) were amplified with SD6-PSPL3\_RT-FW (5'-TCACCTGGACAACCTCAAAG-3') and RTpSAD-RV (Patent P201231427) (size 1016 nt) using Platinum-Taq DNA polymerase (Life Technologies). Samples were denatured at 94 °C for 2 min, followed by 35 cycles of 94 °C/30 s, 60 °C/30 s, and 72 °C (1 min/kb), and a final extension step at 72 °C for 5 min. RT-PCR products were sequenced as previously indicated. Rare transcripts were subcloned into the pJet1.2 PCR cloning vector (Thermo Fisher Scientific) and sequenced.

In order to relatively quantify all transcripts, semiquantitative fluorescent RT-PCRs were undertaken in triplicate with primers PSPL3\_RT-FW and RTpSAD-RV (FAM-labeled) and Platinum Taq DNA polymerase (Life Technologies) under standard conditions [20] except that 26 cycles were applied here. FAM-labeled products were run with LIZ-1200 Size Standard at the MacroGen facility and analyzed with the Peak Scanner software V1.0 (Life Technologies). Only peak heights  $\geq 50$  RFU (relative fluorescence units) were considered.

#### Availability of data and material

All relevant data is available at <https://figshare.com/s/460c77582d659846e27b>.

## Results

Exons 2–9 and adjacent intron sequences were cloned into the vector pSAD (Figure 1A). The 4988 bp final insert corresponds to a genomic region of 15 163 bp.

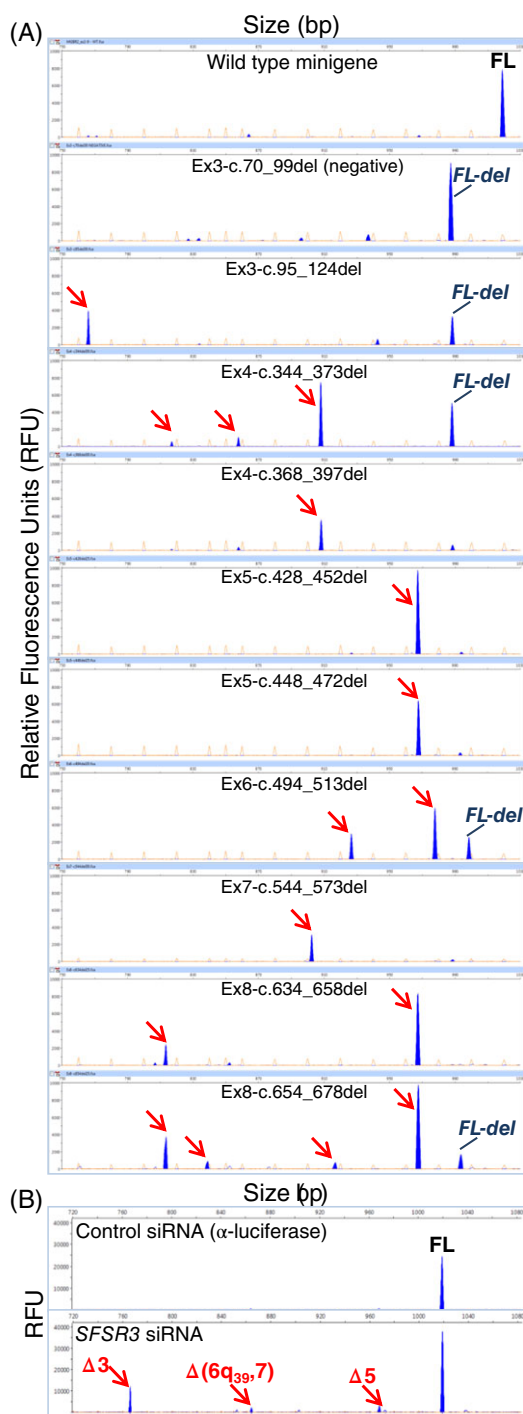
MCF-7 cells were transfected and a major transcript with the expected structure, size and sequence (V1-[BRCA2\_2–9]-V2; 1016 nt) was characterized. The alternative transcripts  $\Delta 3$ ,  $\Delta 5$ ,  $\Delta 5-6$  and  $\Delta(6q_{39},7)$  were also expressed at low levels (0.6–3.4%) by the WT construct (Figure 1B), all of which had been previously reported as physiological alternative events by Fackenthal *et al* [17].

#### Exon splicing enhancer mapping

Efficient exon recognition and inclusion into the mRNA may require the binding of splicing factors to exon splicing enhancer (ESE) sequences. The highest densities of active exonic ESE are found near the splice sites (about 50 nt at the 5' and 3' ends) [29]. To identify ESE-regulated exons, we introduced 26 overlapping microdeletions of 21, 25, or 30 nt (see supplementary material, Tables S2 and S4), which spanned the whole sequences (excluding the conserved two first and three last nucleotides) except for exon 3, which were restricted to the first and last 55 nt due to its size (249 bp).

Mapping assays revealed that nine microdeletions disrupted splicing, (Figure 2A; see supplementary material, Table S4): exon 3-c.95\_124del, exon 4-c.344\_373del and c.368\_397del, exon 5-c.428\_452del, and c.448\_472del, exon 6-c.494\_513del, exon 7-c.544\_573del, exon 8-c.634\_658del, and c.654\_678del. Therefore, they constituted the presumed ESE-rich regions so these intervals constitute hotspots for potential ESE-variants.

To identify potentially involved SRp binding proteins, we performed siRNA experiments to silence seven common SRp (see supplementary material, Table S3) and *Tra2 $\beta$* . Only *SRSF3* (SRp20) had a modest effect



**Figure 2.** Characterization of splicing regulation by functional analyses of minigene microdeletions and siRNA inhibition. (A) Different exonic deletions were functionally tested in MCF-7 cells. FL, full-length transcript, FL-del, FL with a microdeletion. Abnormal transcripts are indicated by red arrows. (B) Inhibition assays with control  $\alpha$ -luciferase siRNA (above) and *SFSR3* (below).

on splicing resulting in a  $\Delta 3$  increase of up to 17.5% (Figure 2B).

### Spliceogenic variants

The main mutation databases (BIC and UMD) were pooled together to collect 302 different variants throughout *BRCA2* ex2–9 region. Intronic variants were

restricted to positions +1 to +6 and –1 to –15, because of their high conservation within the splice sites [30]. Selection was based on the following criteria: splice site disruption or modification, creation of new alternative splice sites, disruption of ESEs within presumed ESE-rich intervals (spliceogenic microdeletions) or creation of exonic splicing silencers (ESS; specifically, hnRNPA1 sites according to the HSF algorithm). After an exhaustive *in silico* analysis, 84 variants (49 VUS) were selected for minigene assays (see supplementary material, Table S5). Thirty-two candidate variants disrupted the canonical splice sites, six created new ones and four strengthened cryptic acceptor sites. One of them, c.517-1G>A simultaneously disrupted the canonical 3' SS and created a new strong one 1-nt downstream (NNSplice: 0.98). On the other hand, 23 variants disrupted ESEs and 19 created hnRNPA1 sites. Several of these variants were predicted to have double-effects.

All variants except one (c.574\_575del) were assayed in MCF-7 cells. Variant c.574\_575del was excluded from this study because its mutagenesis experiments did not work properly. All transcripts were quantified by semiquantitative PCR to assess their impact on gene function. Given that the WT minigene showed 91.9% of the canonical transcript (see supplementary material, Table S6), we considered spliceogenic variants only when the proportion of the canonical transcript was lower than that of the WT minigene. Fragment analysis by capillary electrophoresis showed that 53 out of them (63.8%) impaired splicing (Figure 3; see supplementary material, Figure S2, Table S6). Altogether, they comprised 27 exonic (18 missense, 3 synonymous, 2 nonsense, and 4 frameshift) and 26 intronic changes, confirming that any nucleotide change at any position may, potentially, disrupt pre-mRNA processing. Among the spliceogenic variants, we found that 28 were predicted to disrupt the canonical sites (9 acceptor, 1 Pyr tract and 18 donor variants), four created new alternative splice sites, 10 disrupted putative splicing enhancers, eight presumably created silencers and three had a combined effect (c.517-1G>A, splice site disruption and creation; c.440A>T and c.92G>A – ESE disruption/ESS creation) (see supplementary material, Tables S4–S6). In a previous work, more than 60% or a clear majority of aberrant transcripts from a variant allele was considered as a standard for a severe splicing aberration [31]. Consequently, 36 spliceogenic changes generated severe splicing abnormalities (Table 1).

### Transcripts and splicing events

Fragment analysis of fluorescent RT-PCR products from WT and mutant minigenes revealed more than 23 different transcripts (see supplementary material, Table S6), 19 of them were the main outcomes (Figure 4). Seven transcripts ( $\Delta 2$ ,  $\Delta 3$ ,  $\Delta 4$ ,  $\Delta 4-7$ ,  $\Delta 5$ ,  $\Delta 5,6$ ,  $\Delta 5-7$ ,  $\Delta 6$ ,  $\nabla 5p23$ ,  $\Delta 6q39,7$ ) had been formerly characterized as naturally occurring alternative isoforms, suggesting that alternative transcripts might have a certain predictive value of variant splicing outcomes [17]. On the

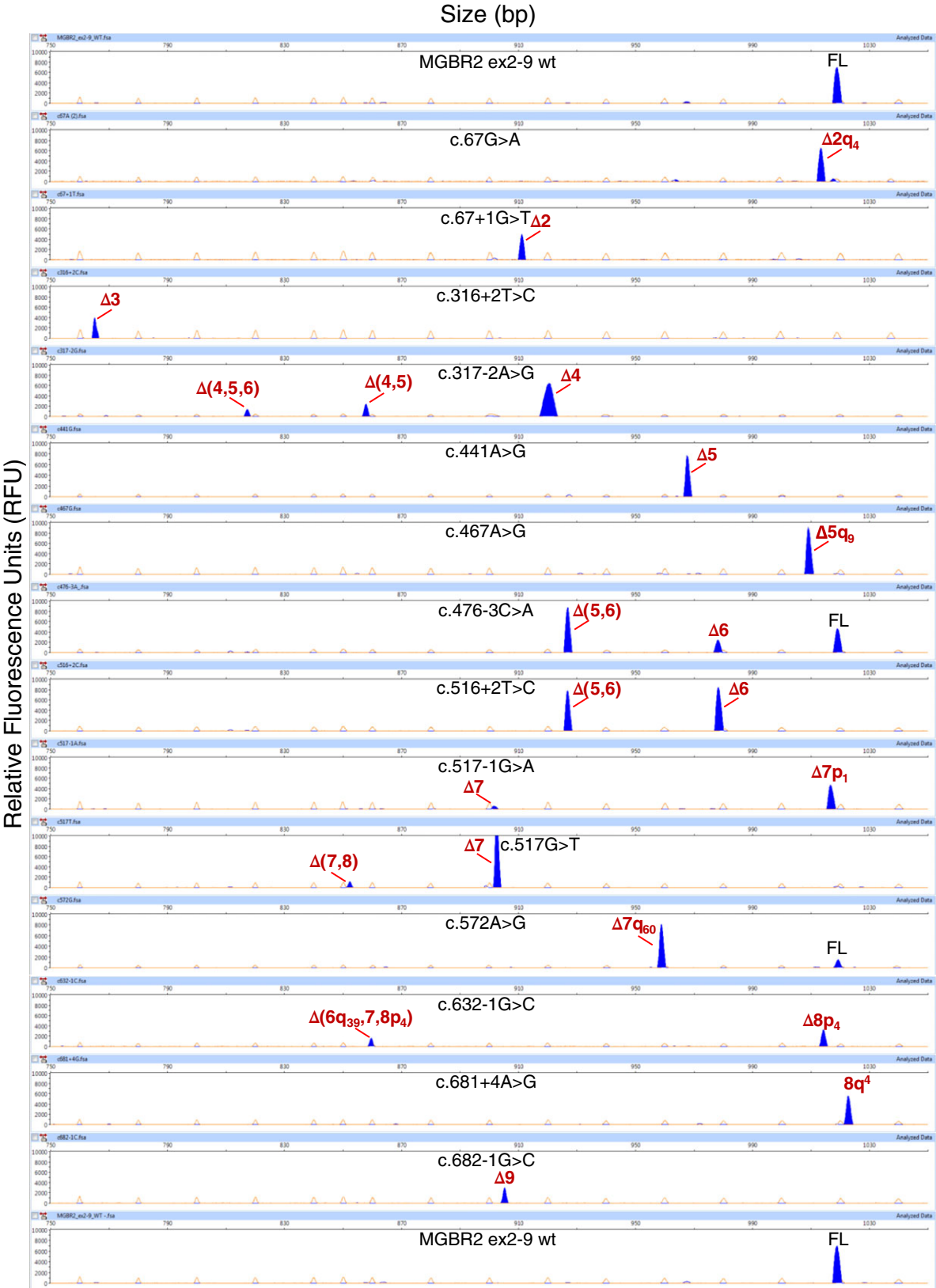


Figure 3. Capillary electrophoresis of RT-PCR products from the WT and selected mutant minigenes. cDNAs were amplified with primers SD6-PSPL3\_RTfW and RTpSAD-RV (full-length transcript V1-*BRCA2* ex2-9-V2 = 1016 nt). FAM-labeled products (blue) were run with LIZ1200 (orange) as size standard. For transcript descriptions see supplementary material, Table S6.

Table 1. Splicing outcomes and classification of spliceogenic variants.

| DNA variant       | Main aberrant transcripts*  | Protein effect                       | ClinVar classification <sup>†</sup> | ACMG classification (criteria) <sup>‡</sup> | Enigma Classification   |
|-------------------|---|--------------------------------------|-------------------------------------|---|-------------------------|
| c.67G>A           | Δ2q <sub>4</sub> (r.64_67del)   | p.A221fs*2                           | VUS/L.Pathog.                       | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.67+1G>T         | Δ2 (r.-39_67del)  | p.M?                                 | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.67+2T>C         | Δ2 (r.-39_67del)  | p.M?                                 | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.67+3A>G         | Δ2 (r.-39_67del)  | p.M?                                 | No data – VUS                       | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.97G>A           | Δ3 (r.68_316del)  | p.D23_L105del                        | No data – VUS                       | L. Pathog. (PM2, PM4, PP3, PP5)             | Class-3                 |
| c.100G>A          | Δ3 (r.68_316del)  | p.D23_L105del                        | No data – VUS                       | L. Pathog. (PM2, PM4, PP3, PP5)             | Class-3                 |
| c.316+2T>C        | Δ3 (r.68_316del)  | p.D23_L105del                        | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.316+3del        | Δ3 (r.68_316del)  | p.D23_L105del                        | No data – VUS                       | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.316+5G>C        | Δ3 (r.68_316del)  | p.D23_L105del                        | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.316+6T>C        | Δ3 (r.68_316del)  | p.D23_L105del                        | VUS                                 | L. Pathog. (PS3, PM2, PP3, PP5)             | Class-4                 |
| c.317-2A>G        | Δ4;Δ(4,5);Δ(4,5,6)<br>(r.[317_425del;317_475del;317_516del])  | p.[G106Vfs*10;G106_Ser158;R107Sfs*9] | L. Pathog.                          | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.426-12_426-8del | Δ5 (r.426_475del)   | p.P143Gfs*23                         | LP – VUS                            | Pathog. (PVS1, PS3, PM2, PP5)               | Class-4                 |
| c.426-2A>T        | Δ5 (r.426_475del)   | p.P143Gfs*23                         | Pathog.                             | Pathog. (PVS1, PS3, PP3, PP5)               | Class-4                 |
| c.441A>G          | Δ5 (r.426_475del)   | p.P143Gfs*23                         | Likely benign                       | VUS (PS3, PP3)                              | Class-3                 |
| c.451G>A          | Δ5;Δ(5,6) (r.[426_475del;426_516del])   | p.[P143Gfs*23;S142Rfs*13]            | VUS                                 | L. Pathog. (PS3, PM2, PP3)                  | Class-3                 |
| c.467A>G          | Δ5q <sub>9</sub> (r.467_475del)   | p.D156_S158del                       | Likely benign – VUS                 | VUS (PM4, PP3, PP5)                         | Class-3                 |
| c.470_474del      | Δ5 (r.426_475del)   | p.P143Gfs*23                         | Pathog. (frameshift)                | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-5<br>(frameshift) |
| c.475+1G>T        | Δ5 (r.426_475del)   | p.P143Gfs*23                         | L. Pathog.– Pathog.                 | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.475+3A>T        | Δ5 (r.426_475del)   | p.P143Gfs*23                         | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.476-2A>G        | Δ6;Δ(5,6) (r.[476_516del;r.426_516del])   | p.[V159Gfs*10;S142Rfs*13]            | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.476-3C>A        | Δ(5,6);Δ6 (r.[476_516del;r.426_516del])   | p.[V159Gfs*10;S142Rfs*13]            | Pathog.                             | L. Pathog. (PS3, PP3, PP5)                  | Class-4                 |
| c.516+1G>T        | Δ6;Δ(5,6) (r.[476_516del;426_516del])   | p.[V159Gfs*10;S142Rfs*13]            | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.516+2T>C        | Δ6;Δ(5,6) (r.[476_516del;426_516del])   | p.[V159Gfs*10;S142Rfs*13]            | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.516+4_516+5del  | Δ(5,6), Δ6 (r.[426_516del;476_516del])  | p.[S142Rfs*13;V159Gfs*10]            | No data – VUS                       | Pathog. (PVS1, PS3, PM2, PP3)               | Class-4                 |
| c.517-2A>G        | Δ7 (r.517_631del)   | p.G173Sfs*19                         | L. Pathog.– Pathog.                 | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.517-1G>A        | Δ7p <sub>1</sub> ;Δ7 (r.[517del;517_631del])  | p.[G173Vfs*12;G173Sfs*19]            | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.517G>T          | Δ7 (r.517_631del)   | p.G173Sfs*19                         | VUS                                 | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.572A>G          | Δ7q <sub>60</sub> (r.572_631del)  | p.D191_Ile210del                     | VUS                                 | VUS (PM2, PM4, PP3)                         | Class-3                 |
| c.631G>A          | Δ7 (r.517_631del)   | p.G173Sfs*19                         | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.631+1G>A        | Δ7 (r.517_631del)   | p.G173Sfs*19                         | L. Pathog.– Pathog.                 | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.631+3A>G        | Δ7 (r.517_631del)   | p.G173Sfs*19                         | VUS – Pathog.                       | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.632-3C>G        | ▼8p <sup>2</sup> ;Δ(6q <sub>39</sub> ,7)▼<br>8p <sup>2</sup> (r.[631_632ins632-1_632-2;478_631delins632-1_632-2]) | p.[V211Efs*20;V160Sfs*4]             | VUS                                 | Pathog. (PVS1, PS3, PP3, PP5)               | Class-4                 |
| c.632-1G>C        | Δ8p <sub>4</sub> ;Δ(6q <sub>39</sub> ,7,8p <sub>4</sub> )<br>r.[632_635del;517_635]                               | p.[V211Efs*18;G173Kfs*2]             | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.681+4A>G        | ▼8q <sup>4</sup> (r.681_682ins681+1_681+4)  | p.N228Vfs*11                         | VUS/L.Pathog/Pathog                 | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.682-1G>C        | Δ9 (r.682_793del)   | p.N228Dfs*12                         | L.Pathog.                           | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |
| c.793+1G>T        | Δ9 (r.682_793del)   | p.N228Dfs*12                         | Pathog.                             | Pathog. (PVS1, PS3, PM2, PP3, PP5)          | Class-4                 |

\*Transcripts were annotated according to previous reports of the ENIGMA consortium. The remaining transcripts are described in Supplementary Table S6.

<sup>†</sup>Previous classifications according to the ClinVar database. Pathog., pathogenic; L. Pathog., Likely pathogenic.

<sup>‡</sup>ACMG criteria, evidence of pathogenicity. Very strong: PVS1: null variant (nonsense, frameshift, canonical  $\pm 1$  or 2 splice sites, initiation codon, single, or multiexon deletion) in a gene where LOF is a known mechanism of disease. Strong: PS3, well-established *in vitro* or *in vivo* functional studies (splicing assays) supportive of a damaging effect on the gene or gene product. Moderate: PM2, Absent from controls (all variants were checked at the ClinVar database, <https://www.ncbi.nlm.nih.gov/clinvar/>); PM4, protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants (this item has been used for variants that induce in-frame transcripts such as ex18-ins6). Supporting: PP3: Multiple lines of computational evidence support a deleterious effect on the gene or gene product. In this study, all variants were checked with NNSplice and HSF 3.0 that contains several splicing algorithms; PP5: Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation (only used for missense, nonsense and frameshift variants).

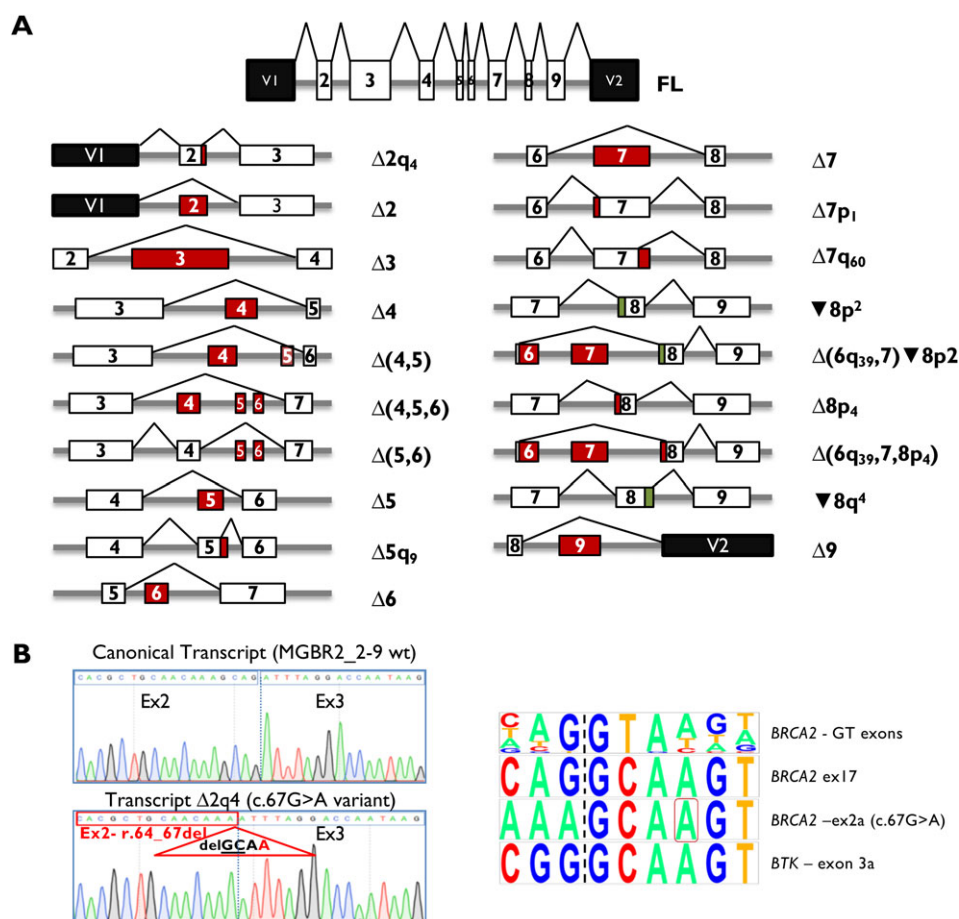


Figure 4. Principal aberrant transcripts. (A) Diagrams of transcripts. Anomalous/skipped exons and insertions are shown in red and green, respectively. (B) Use of an atypical GC-donor site induced by c.67G>A: transcript sequences (left). The consensus sequence of *BRCA2*-5' splice sites and the GC-donor sequences of *BRCA2* exon 17, c.67G>A and *BTK* exon 3a (<http://genes.mit.edu/pictogram.html>) are shown on the right.

other hand, transcripts  $\Delta 3$ ,  $\Delta 5q_9$ ,  $\Delta(4,5)$  and  $\Delta 7q_{60}$  kept the reading frame, while the remaining isoforms introduced PTCs.

Interestingly, variant c.67G>A disrupted exon 2 donor and provoked the use of a noncanonical GC site 4-nt upstream ( $\Delta 2q_4$ , Figure 4B). This sequence was compared with the canonical GT exons of *BRCA2* and other GC exons, such as *BRCA2* exon 17 and the pathological exon 3a of the *BTK* gene (Figure 4B, right) [32]. Remarkably, GC donors shared the same 6-nt intronic sequence GCAAGT. In order to determine the splicing factor involved in the GC recognition, we carried out pull-down assays (see supplementary material, Figure S3). Western-blot with an anti-SRp family antibody showed that SRp55 was apparently bound stronger to the WT sequence than to the c.67G>A mutant. Conversely, SRp40 and, to a less extent SRp30 and SRp20, showed more intense signals in the mutant sequence (see supplementary material, Figure S3).

## Discussion

Thousands of germline variants have been found in *BRCA2* and this number is expected to increase by

virtue of the massive sequencing technology implementation in the clinical setting. Functional studies of variants provide key information for their clinical interpretation and so, for the management of both patients and asymptomatic carriers. In this work, we have performed an exhaustive study of splicing variants of exons 2–9 as well as the preliminary characterization of their regulation.

Our previous studies represent the most comprehensive functional analysis of the splicing impact of *BRCA2* variants in HBOC, supporting the hypothesis of the high prevalence of spliceogenic mutations [20,22,33–35]. Here, we have accomplished an in depth analysis of *BRCA2* exons 2–9 through the functional test of 83 reported variants in a minigene. Fifty-three out of them altered splicing, accounting for 17.5% of all reported variants in these eight exons (see supplementary material, Table S6). Remarkably, this percentage almost quadruples the frequency of splicing variants at the BIC database (4.6%).

In this context, minigenes represent a simple and robust tool to assay all potential splicing variants, since patient RNA samples are not always available. Actually, 260 *BRCA1/2* variants have been so far tested by minigenes [20,22,33,34,36]. Moreover, minigenes present

some important advantages such as: (1) analysis of one single allele effect; (2) identification and quantification of all transcripts; (3) study of multiple DNA variants using just one multiple-exon minigene; (4) assay in many cell types, essential for tissue-specific alternative splicing; and (5) high reproducibility of physiological and pathological splicing patterns.

Previous results from our group showed that minigene tests of at least 16 different *BRCA2* variants from exons 17 to 27 reproduced patient RNA outcomes [20,22,36]. Moreover, the reproducibility of MGBR2\_2–9 splicing outcomes was supported by previous studies based on patient RNA or minigene data, according to which fourteen variants matched preceding results: c.67+3A>G, c.68-7T>A, c.316+3del, c.316+5G>C, c.426-12\_426-8del, c.467A>G, c.516+1G>T, c.517G>T, c.572A>G, c.617C>G, c.627C>A/T, c.631G>A, c.631+3A>G and c.681+4A>G [37–46]. Conversely, in a recent publication, variant c.68-7T>A induced 20% of transcript  $\Delta 3$  in lymphoblastoid cell lines of carrier patients versus 2.8% of  $\Delta 3$  and 7.2% of  $\Delta(6q_{39},7)$  found in our minigene study in MCF-7 cells [47]. These variations may be due to technical differences in RNA isolation methods, retro-transcriptases, DNA polymerases, primers and product length, thermocycling protocol as well as cell-type specific variations between the patient sample and the minigene assay. Moreover, the WT minigene produced the alternative transcripts  $\Delta 3$ ,  $\Delta 5$ ,  $\Delta(5,6)$  and  $\Delta(6q_{39},7)$  which had been previously characterized in healthy breast tissue and other cell types [17]. In this respect, conservation of the genomic environment in minigenes is essential for mimicking physiological and pathological splicing patterns. Thus, pSAD-based minigenes represent efficient and reliable alternative approaches to test candidate variants [20,22]. Finally, it is also worthy of mention that the high resolution and sensitivity of capillary electrophoresis of minigene transcripts that is able to resolve small variations in sizes (only 1 nt in  $\Delta 7p_1$ ) or even detect rare isoforms. Taken together, these data suggest that splicing reporter outcomes provide valuable information for the clinical interpretation of DNA variants.

### Splicing regulatory elements

Splice site mutations comprised most of the variants with impact on splicing (28/53) since they were precisely computed by NNSplice or MaxEntScan (see supplementary material, Table S5). Although ESE/ESS predictions are inaccurate [48], we had previously showed that microdeletion mapping is a valuable method to locate ESEs. Thus, specificity of bioinformatics estimates was increased by selecting only ESE-mutations located at these spliceogenic segments [20,22]. Interestingly, exons 3–8 showed nine critical regulatory segments (Figure 2), which constitute hotspots for spliceogenic ESE-mutations. Indeed, we found that 12 of the 21 spliceogenic ESE/ESS variants (9 ESE-disrupting, 2 ESS-creating and 1 double ESS/ESE

variants) were located within spliceogenic microdeletions: c.97G>A, c.100G>A (exon 3), c.352C>T, c.368\_372del5, c.378A>G, c.379G>T, c.383A>G (exon 4), c.440A>T, c.441A>G, c.451G>A, c.467A>G and c.470\_474del (exon 5). Only two spliceogenic ESE-variants (c.617G>C and c.620C>T) were placed within a microdeletion (c.599\_628del) that did not alter splicing. Both variants disrupted one SC35 motif and c.620C>T also abrogated one SRp40 motif that were deleted (among other motifs) in c.599\_628del. Given the low accuracy of the ESE-finder algorithm, it is plausible that other splicing elements (either enhancers or silencers), not detected by HSF, might be involved. These findings support the value of this strategy at choosing candidate ESE-mutations as an initial step before bioinformatics selection.

Remarkably, variants c.97G>A, and c.100G>A were within interval c.95\_124. The siRNA assays revealed that exon 3 is partially regulated by SRp20 (Figure 2B). Precisely, an SRp20 binding site (TCTTC, c.106–110 and its complementary sequence 9-nt upstream (GAAGA, c.97–101) were detected by RegRNA software (<http://regna.mbc.nctu.edu.tw/html/prediction.html>). Furthermore, according to Mfold (<http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form>), both sequences are part of a hairpin-like secondary structure (see supplementary material, Figure S4) that is destabilized by variants c.97G>A and c.100G>A and eliminated by microdeletion c.95\_124del. Our data illustrate the high-complexity of splicing regulation of *BRCA2* exon 3 where either SRp20, the RNA secondary structure or a combination of both factors play a definitive role in exon 3 recognition. We also found that small exons 5 and 8 (c.426–475 and c.632–681, respectively) were strongly regulated as all microdeletions yielded almost complete splicing anomalies (see supplementary material, Table S4), pointing at the presence of ESEs within these exons.

Conversely, seven spliceogenic ESE/ESS variants were located in exon 3, outside the spliceogenic microdeletion c.95\_124del. They were predicted by HSF to create one or two putative hnRNPA1 motifs: c.92G>A (1x hnRNPA1), c.140C>T (1 x hnRNPA1), c.143A>G (2 x hnRNPA1), c.144delA (2 x hnRNPA1), c.215A>G (2 x hnRNPA1), c.217C>T (1 x hnRNPA1) and c.222G>A (1 x hnRNPA1). However, other postulated hnRNPA1-variants, which even created strong (c.71T>A) or several sites (c.506A>G), did not alter splicing. It is therefore plausible that recognition of certain highly-regulated exons, such as the alternative *BRCA2* exon 3, might be particularly sensitive to ESS generation. Curiously, these 6 hnRNPA1 variants were placed relatively far from the nearest splice site (72–94 bp), just outside the 50-nt intervals of exon ends, where ESE/ESS are assumed to exert their function [29]. Finally, the contribution of splicing enhancers or silencers only can be determined by further experiments, such as RNA-protein binding assays, which will unveil the involved splicing regulatory factors.

### Clinical interpretation of variants

The clinical interpretation of splicing patterns is a particularly complex and arduous task. Indeed, we have described a wide repertoire of splicing rearrangements and transcripts that require a careful analysis. Exon skipping was the most common event (12 events) that involved one (e.g.  $\Delta 4$ ) or several exons (e.g.  $\Delta(4,5)$ ) (Figure 4 and supplementary material, Table S6). Remarkably, seven isoforms ( $\Delta 2$ ,  $\Delta 3$ ,  $\Delta 4$ ,  $\Delta 5$ ,  $\Delta 5,6$ ,  $\Delta 6$ , and  $\Delta 6q_{39,7}$ ) originated by variants had previously been characterized as physiological alternative splicing events [17]. While  $\Delta 3$  deletes the PALB2 binding domain for BRCA2 function, the rest of alternative transcripts introduces PTCs, so their expression at high levels from a variant allele might potentially inactivate BRCA2.  $\Delta 3$  transcript keeps the reading frame but it lacks the PALB2 binding motif (amino acids 10–40) that is essential for BRCA2 function [49]. Hence, changes in this region may cause severe alterations in protein structure and functionality. Likewise, variants c.316+2T>C, c.316+3delA, c.316+5G>C and c.316+6T>C, which also induced  $\Delta 3$ , had been already catalogued as pathogenic [38]. Consequently, all the  $\Delta 3$ -inducing variants should be considered deleterious.

Alternate site usage of cryptic or new splice sites was the second main outcome. Unexpectedly, variant c.67G>A disrupted the canonical donor and promoted the recognition of a cryptic noncanonical GC donor 4-nt upstream. This phenomenon is extremely unusual, since <1% of human 5'SS are GC [32]. Its usage produced a frameshift deletion (r.64\_67del) that truncated BRCA2 (p.A221fs\*2). This finding is similar to that of the use of an anomalous GC site at the *BTB* gene 107-nt downstream of the canonical donor site of exon 3, whose recognition involves 9G8, Tra2 $\beta$  and SC35 factors [32]. Both GC donors, together with that of BRCA2 exon 17, share the same 6-nt intronic sequence GCAAGT (Figure 4B) that is highly conserved in human constitutive versus alternative GC-exons [50], suggesting a common regulatory mechanism.

Additionally, we observed the use of alternative 3'SS 2-nt upstream ( $\nabla 8p^2$ ,  $\Delta(6q_{39,7})\nabla 8p^2$ ) and 4-nt downstream ( $\Delta 8p_4$ ,  $\Delta(6q_{39,7,8p_4})$ ) of exon 8 when the canonical acceptor was broken by c.632-3C>G and c.632-1G>C, respectively. Variant c.681+4A>G improved an intronic cryptic donor which inserted 4 nt (p.N228Vfs\*11), while c.467A>G, c.517-1G>A and c.572A>G generated new splice sites that were recognized by the spliceosome. On one hand, c.517-1G>A created a new 3'SS, producing the truncating transcript  $\Delta 7p_1$  (p.G173Vfs\*12). On the other hand, c.467A>G and c.572A>G created new exonic 5'SS sites that caused in-frame deletions of 9 ( $\Delta 5q_9$ ) and 60-nt ( $\Delta 7q_{60}$ ), respectively. Isoform  $\Delta 5q_9$  implied the loss of Asp-Lys-Ser (p. 156D\_158Sdel), which are poorly conserved (<http://agvgd.hci.utah.edu>) while  $\Delta 7q_{60}$  deleted 20 amino acids (p.D191\_210del), nine of which are conserved from human to sea urchin. Therefore,

c.467A>G and c.572A>G could not be classified as deleterious.

According to the guidelines of the ACMG [25], we evaluated 36 variants with severe splicing anomalies (>60% of aberrant transcripts) as defined by Vallée *et al* [31] (Table 1). Twenty-eight of them could be classified as pathogenic (c.67G>A, c.67+1G>T, c.67+2T>C, c.67+3A>G, c.316+2T>C, c.316+3del, c.316+5G>C, c.317-2A>G, c.426-12\_426-8del, c.426-2A>T, c.470\_474del, c.475+1G>T, c.475+3A>T, c.476-2A>G, c.516+1G>T, c.516+2T>C, c.516+4\_516+5del, c.517-2A>G, c.517-1G>A, c.517G>T, c.631G>A, c.631+1G>A, c.631+3A>G, c.632-3C>G, c.632-1G>C, c.681+4A>G, c.682-1G>C and c.793+1G>T), five as likely pathogenic (c.97G>A, c.100G>A, c.316+6T>C, c.451G>A and c.476-3C>A) and three remained as VUS. Moreover, the ENIGMA guidelines suggest that one variant should be classified as Class-5, 29 as Class-4 and 6 as Class-3. These discrepancies rely on the more restrictive ENIGMA criterion that does not consider construct-based mRNA assays alone as sufficiently robust to be used as evidence for variant classification. Nevertheless, the high reproducibility of this minigene might support an upgrade of most variants from Class-4 to Class-5. Interestingly, 13 VUS variants of ClinVar were reclassified as pathogenic or likely pathogenic following the ACMG criteria (Table 1), whereas 17 pathogenic variants (Table 1) match previous classification at the ClinVar database, thus supporting the high capacity of the minigene approach.

It is also worthy of mention that 10 ACMG pathogenic/likely pathogenic variants (four of them VUS) affected nucleotides surrounding the ultra-conserved positions +1 and +2 of the donor site: c.67G>A, c.67+3A>G, c.316+3delA, c.316+5G>C, c.316+6T>C, c.475+3A>T, c.516+4\_516+5del, c.631G>A, c.631+3A>G and c.681+4A>G. Given the high conservation of these sequences (<http://www.life.umd.edu/labs/mount/RNAinfo/consensus.html>; MAGgttaagt), our data confirm the potential 'spliceogenicity' of mutations in any of these nucleotides, spanning from the last 3 nt of each exon to the intron +6 position. Importantly, *in silico* analysis of c.316+5G>C and c.316+6T>C did not detect abrogation of exon 3-5'SS, like the previously reported variant c.7976+5G>T, which by contrary generated major splicing defects (see supplementary material, Tables S5 and S6) [36].

Finally, four ESE/ESS variants (c.97G>A, c.100G>A, c.441A>G, and c.451G>A), previously catalogued as VUS, impaired splicing and were reclassified as pathogenic or likely pathogenic. This underlines the role of these elements in disease susceptibility and highlights, once more, the value of the strategy of functional mapping by exonic microdeletions at locating specific target segments for putative ESE-variants. Interestingly, the variants c.441A>G, which encodes for a synonymous change (p.Gln147Gln), c.439C>T (predicted nonsense p.Q147\*) [33] and c.451G>A

(predicted missense p.Val151Ile) should be reclassified as splicing variants.

In conclusion, *BRCA2* mis-splicing is a relatively frequent mechanism of gene inactivation in HBOC. According to the microdeletion ESE-mapping assays, *BRCA2* exons 3–8 contain splicing regulatory elements which constitute critical targets for candidate spliceogenic ESE-variants. In fact, splicing impact of any exonic variant at the small exons 5 and 8 should be checked given that presumed ESE-rich intervals cover the entire sequences of both exons. Unexpectedly, a large fraction of all tested variants (63.8%) disrupted splicing, and 33 out of them (14 VUS) were classified as pathogenic or likely pathogenic according to the ACMG standards. In our hands, the minigene technology has demonstrated to be a straightforward, reliable and powerful approach to describe a wide range of aberrant splicing phenomena in HBOC genes as well as a means to address the basic regulatory rules that govern splicing. Besides, the pSAD-derived minigenes have also been useful tools to study other genetic diseases, such as  $\alpha$ -1 antitrypsin deficiency [24] or Charge syndrome [23]. Finally, RNA assays supply with essential information for the classification of DNA variants and, in this context, minigenes represent a sensitive tool for the initial characterization of splicing anomalies. So, they should be considered as a supportive test in the clinical interpretation of splicing variants.

## Acknowledgements

EAV's lab was supported by grants from the Spanish Ministry of Economy and Competitiveness, Plan Nacional de I+D+I 2013–2016, ISCIII (PI13/01749 and PI17/00227) co-funded by FEDER from Regional Development European Funds (European Union), and grants CSI090U14 (ORDEN EDU/122/2014) and CSI242P18 (actuación cofinanciada P.O. FEDER 2014–2020 de Castilla y León) from the Consejería de Educación, Junta de Castilla y León. EFB is supported by a predoctoral fellowship from the University of Valladolid and Banco de Santander (2015–2019). AVP is supported by a predoctoral fellowship from Consejería de Educación, Junta de Castilla y León (2018–2022). We are grateful to Susana Gómez-Barrero for the critical reading of the manuscript.

## Author contributions statement

EF-B contributed to the bioinformatics analysis, participated in all the experiments of this study and wrote together with EAV the first draft of the manuscript. AV-P participated in mutagenesis, functional assays, and semiquantitative RT-PCR experiments. BD-G performed minigene construction, mutagenesis and functional assays. AA participated in minigene construction and its functional validation. EG and EB contributed to all the experiments to define the splicing

regulatory mechanisms of this *BRCA2* minigene. EAV conceived the study and the experimental design, performed bioinformatics analysis, supervised all the experiments and wrote the manuscript. All authors contributed to data interpretation and were involved in writing the manuscript and had final approval of the submitted and published versions.

## References

- Slavin TP, Maxwell KN, Lilyquist J, et al. The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. *NPJ Breast Cancer* 2017; **3**: 22.
- Nielsen FC, van Overeem Hansen T, Sørensen CS. Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat Rev Cancer* 2016; **16**: 599–612.
- Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet* 2008; **40**: 17–22.
- Infante M, Durán M, Acedo A, et al. The highly prevalent *BRCA2* mutation c.2808\_2811del (3036delACAA) is located in a mutational hotspot and has multiple origins. *Carcinogenesis* 2013; **34**: 2505–2511.
- Eccles DM, Mitchell G, Monteiro ANA, et al. *BRCA1* and *BRCA2* genetic testing-pitfalls and recommendations for managing variants of uncertain clinical significance. *Ann Oncol* 2015; **26**: 2057–2065.
- Radice P, De Summa S, Caleca L, et al. Unclassified variants in *BRCA* genes: guidelines for interpretation. *Ann Oncol* 2011; **22**(suppl 1): i18–i23.
- Smith SA, French T, Hollingsworth SJ. The impact of germline mutations on targeted therapy. *J Pathol* 2014; **232**: 230–243.
- Padamsee TJ, Wills CE, Yee LD, et al. Decision making for breast cancer prevention among women at elevated risk. *Breast Cancer Res* 2017; **19**: 34.
- Brewster BL, Rossiello F, French JD, et al. Identification of fifteen novel germline variants in the *BRCA1* 3'UTR reveals a variant in a breast cancer case that introduces a functional miR-103 target site. *Hum Mutat* 2012; **33**: 1665–1675.
- Dos Santos ES, Caputo SM, Castera L, et al. Assessment of the functional impact of germline *BRCA1/2* variants located in non-coding regions in families with breast and/or ovarian cancer predisposition. *Breast Cancer Res Treat* 2018; **168**: 311–325.
- Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, et al. Genetic dissection of the *BRCA2* promoter and transcriptional impact of DNA variants. *Breast Cancer Res Treat* 2018; **171**: 53–63.
- Rhine CL, Cygan KJ, Soemedi R, et al. Hereditary cancer genes are highly susceptible to splicing mutations. *PLoS Genet* 2018; **14**: e1007231.
- Wang G-S, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 2007; **8**: 749–761.
- Ward AJ, Cooper TA. The pathobiology of splicing. *J Pathol* 2010; **220**: 152–163.
- Pedrotti S, Cooper TA. In Brief: (mis)splicing in disease. *J Pathol* 2014; **233**: 1–3.
- Colombo M, Blok MJ, Whaley P, et al. Comprehensive annotation of splice junctions supports pervasive alternative splicing at the *BRCA1* locus: a report from the ENIGMA consortium. *Hum Mol Genet* 2014; **23**: 3666–3680.
- Fackenthal JD, Yoshimatsu T, Zhang B, et al. Naturally occurring *BRCA2* alternative mRNA splicing events in clinically relevant samples. *J Med Genet* 2016; **53**: 548–558.
- Whaley PJ, de la Hoya M, Thomassen M, et al. Comparison of mRNA splicing assay protocols across multiple laboratories:

- recommendations for best practice in standardized clinical testing. *Clin Chem* 2014; **60**: 341–352.
19. Baralle D, Lucassen A, Buratti E. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep* 2009; **10**: 810–816.
  20. Acedo A, Hernández-Moro C, Curiel-García Á, *et al.* Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum Mutat* 2015; **36**: 210–221.
  21. Cooper TA. Use of minigene systems to dissect alternative splicing elements. *Methods* 2005; **37**: 331–340.
  22. Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V, *et al.* Functional classification of DNA variants by hybrid minigenes: identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. *PLoS Genet* 2017; **13**: e1006691.
  23. Villate O, Ibarluzea N, Fraile-Bethencourt E, *et al.* Functional analyses of a novel splice variant in the CHD7 gene, found by next generation sequencing, confirm its pathogenicity in a spanish patient and diagnose him with CHARGE syndrome. *Front Genet* 2018; **9**: 26–31.
  24. Lara B, Martínez MT, Blanco I, *et al.* Severe alpha-1 antitrypsin deficiency in composite heterozygotes inheriting a new splicing mutation QOMadrid. *Respir Res* 2014; **15**: 125.
  25. Richards S, Aziz N, Bale S, *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; **17**: 405–424.
  26. Beroud C, Letovsky SI, Braastad CD, *et al.* BRCA share: a collection of clinical BRCA gene variants. *Hum Mutat* 2016; **37**: 1318–1328.
  27. Reese MG, Eckman FH, Kulp D, *et al.* Improved splice site detection in Genie. *J Comput Biol* 1997; **4**: 311–323.
  28. Desmet F-O, Hamroun D, Lalonde M, *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: e67.
  29. Fairbrother WG, Holste D, Burge CB, *et al.* Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol* 2004; **2**: E268.
  30. Zhang MQ. Statistical features of human exons and their flanking regions. *Hum Mol Genet* 1998; **7**: 919–932.
  31. Vallée MP, Di Sera TL, Nix DA, *et al.* Adding in silico assessment of potential splice aberration to the integrated evaluation of BRCA gene unclassified variants. *Hum Mutat* 2016; **37**: 627–639.
  32. Kralovicova J, Hwang G, Asplund C, *et al.* Compensatory signals associated with the activation of human GC 5' splice sites. *Nucleic Acids Res* 2011; **39**: 7077–7091.
  33. Sanz DJ, Acedo A, Infante M, *et al.* A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin Cancer Res* 2010; **16**: 1957–1967.
  34. Acedo A, Sanz DJ, Durán M, *et al.* Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res* 2012; **14**: R87.
  35. Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, *et al.* Identification of eight spliceogenic variants in BRCA2 exon 16 by minigene assays. *Front Genet* 2018; **9**: 188.
  36. Montalbán G, Fraile-Bethencourt E, López-Perolio I, *et al.* 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* 2018; **39**: 1155–1160.
  37. Parsons MT, Whiley PJ, Beesley J, *et al.* Consequences of germline variation disrupting the constitutional translational initiation codon start sites of MLH1 and BRCA2: use of potential alternative start sites and implications for predicting variant pathogenicity. *Mol Carcinog* 2015; **54**: 513–522.
  38. Muller D, Rouleau E, Schultz I, *et al.* An entire exon 3 germ-line rearrangement in the BRCA2 gene: pathogenic relevance of exon 3 deletion in breast cancer predisposition. *BMC Med Genet* 2011; **12**: 121.
  39. Bonnet C, Krieger S, Vezain M, *et al.* Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *J Med Genet* 2008; **45**: 438–446.
  40. Zhang L, Bacaes R, Boyar S, *et al.* cDNA analysis demonstrates that the BRCA2 intronic variant IVS4-12del5 is a deleterious mutation. *Mutat Res* 2009; **663**: 84–89.
  41. Whiley PJ, Guidugli L, Walker LC, *et al.* Splicing and multifactorial analysis of intronic BRCA1 and BRCA2 sequence variants identifies clinically significant splicing aberrations up to 12 nucleotides from the intron/exon boundary. *Hum Mutat* 2011; **32**: 678–687.
  42. Théry JC, Krieger S, Gaildrat P, *et al.* Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. *Eur J Hum Genet* 2011; **19**: 1052–1058.
  43. Gaildrat P, Krieger S, Di Giacomo D, *et al.* Multiple sequence variants of BRCA2 exon 7 alter splicing regulation. *J Med Genet* 2012; **49**: 609–617.
  44. Pensabene M, Spagnoletti I, Capuano I, *et al.* Two mutations of BRCA2 gene at exon and splicing site in a woman who underwent oncogenetic counseling. *Ann Oncol* 2009; **20**: 874–878.
  45. Houdayer C, Caux-Moncoutier V, Krieger S, *et al.* Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* 2012; **33**: 1228–1238.
  46. Di Giacomo D, Gaildrat P, Abuli A, *et al.* Functional analysis of a large set of brca2 exon 7 variants highlights the predictive value of hexamer scores in detecting alterations of exonic splicing regulatory elements. *Hum Mutat* 2013; **34**: 1547–1557.
  47. Colombo M, López-Perolio I, Meeks HD, *et al.* The BRCA2 c.68-7T>A variant is not pathogenic: a model for clinical calibration of spliceogenicity. *Hum Mutat* 2018; **39**: 729–741.
  48. Grodecká L, Buratti E, Freiburger T. Mutations of pre-mRNA splicing regulatory elements: are predictions moving forward to clinical diagnostics? *Int J Mol Sci* 2017; **18**: E1668.
  49. Xia B, Sheng Q, Nakanishi K, *et al.* Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol Cell* 2006; **22**: 719–729.
  50. Thanaraj T, Clark F. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Res* 2001; **29**: 2581–2593.
  - \*51. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; **11**: 377–394.
  - \*52. Cartegni L, Wang J, Zhu Z, *et al.* ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003; **31**: 3568–3571.

\*Cited only in supplementary materials.

**SUPPLEMENTARY MATERIAL ONLINE**

**Figure S1.** Pipeline of splicing functional assays

**Figure S2.** Capillary electrophoresis of RT-PCR products from the WT and 22 mutant minigenes

**Figure S3.** Pull-down assays of the WT and c.67A mutant sequences with an anti-SRp antibody

**Figure S4.** Analysis of the predicted RNA secondary structure of the *BRCA2* exon 3 interval between c.90 and c.125

**Table S1.** Primers used for the construction of the MGBR2\_ex2–9 minigene

**Table S2.** Mutagenesis primers of 84 DNA variants and 26 microdeletions

**Table S3.** siRNAs against SR proteins, and qPCR primer sequences

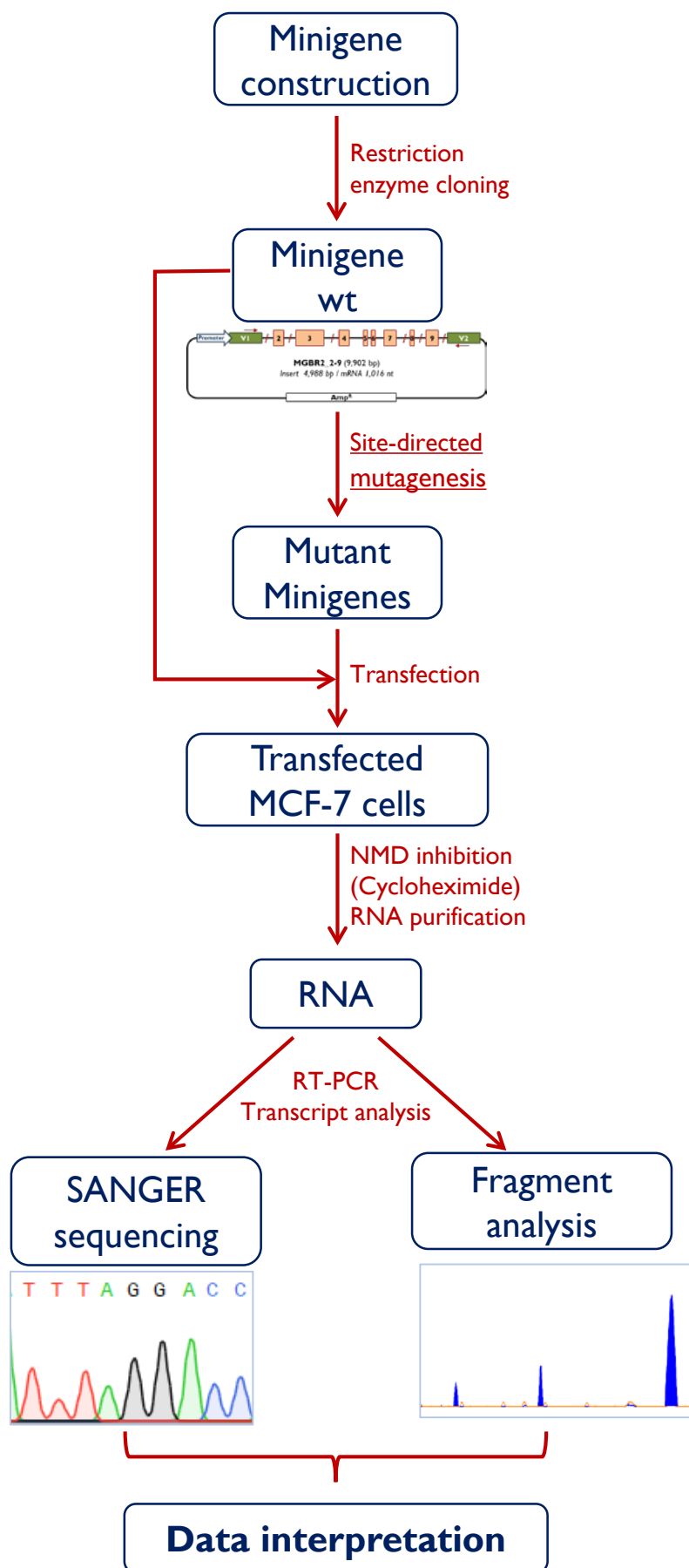
**Table S4.** Quantification of transcripts from spliceogenic microdeletion assays

**Table S5.** Bioinformatics analysis of reported variants of *BRCA2* exons 2–9

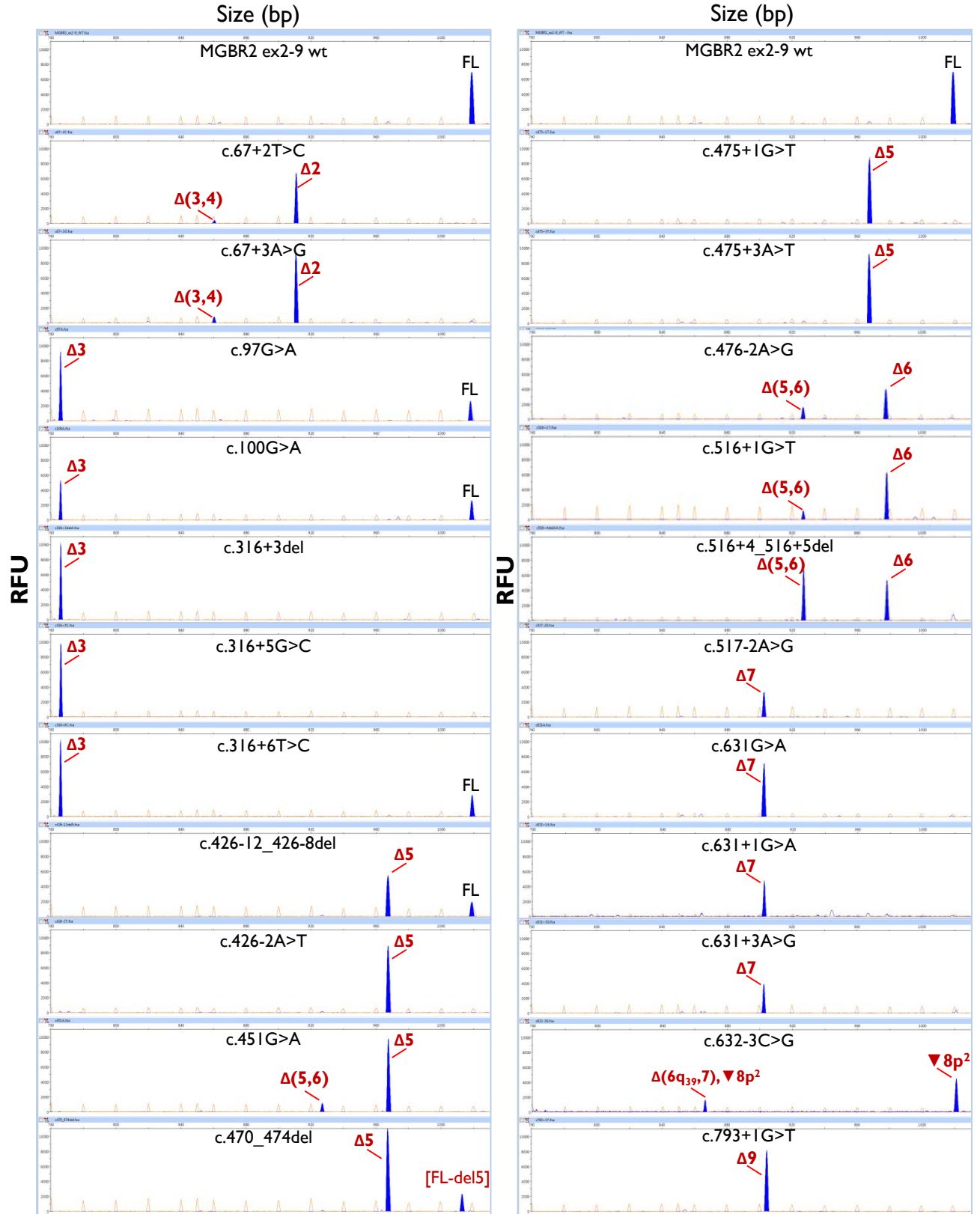
**Table S6.** Splicing outcome and quantification of transcripts of the WT and 83 mutant minigenes

**Mis-splicing in breast cancer: identification of pathogenic BRCA2 variants by systematic minigene assays**

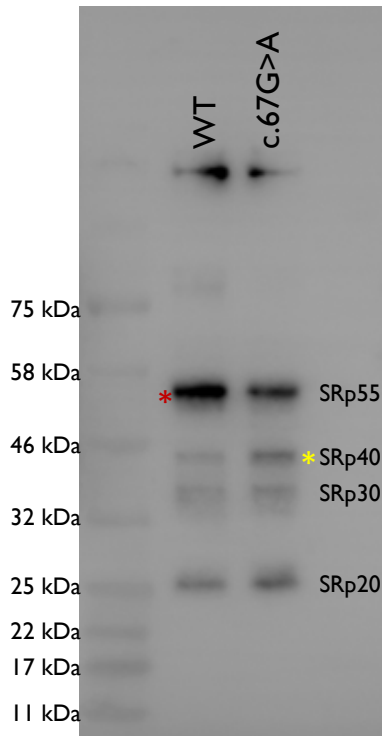
**Fraile-Bethencourt E *et al. J Pathol* DOI: 10.1002/path.5268**



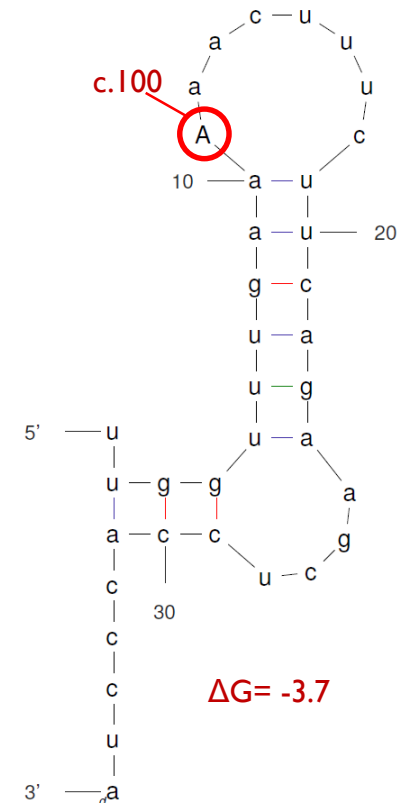
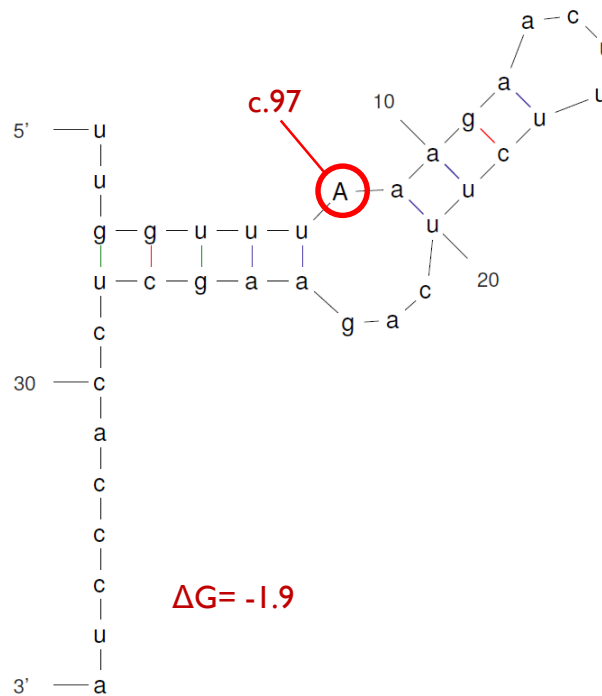
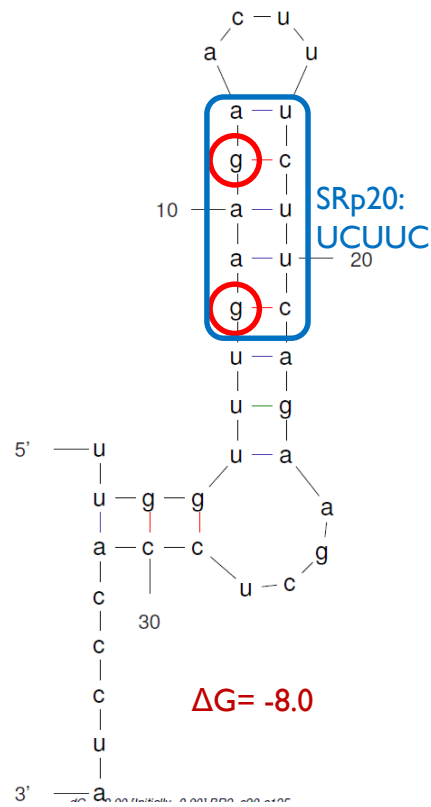
**Figure S1. Pipeline of Splicing Functional Assays.** The basic protocol comprises the following steps: 1) Minigene construction; 2) Site-directed mutagenesis; 3) Transfection of the wild type and mutant minigenes; 3) Inhibition of Nonsense-mediated decay and RNA purification; 4) Transcript sequencing and fragment analysis by fluorescence capillary electrophoresis; 5) Data interpretation.



**Figure S2. Capillary electrophoresis of RT-PCR products from the wild type and 22 mutant minigenes.** cDNAs were amplified with primers SD6-PSPL3\_RTFW and RTpSAD-RV (full-length transcript V1-BRCA2 ex2-9-V2= 1,016 nt). FAM-labeled products (blue peaks) were run with LIZ1200 (orange/faint peaks) as size standard. Transcripts: FL, full length transcript; Δ2, ex 2 skipping; Δ(3,4), ex3+4 skipping; Δ3, ex 3 skipping; Δ5, ex 5 skipping; Δ(5,6), ex5+6 skipping; Δ6, ex 6 skipping; FL-del5, full-length transcript with deletion c.470\_474del; Δ7, ex7 skipping; Δ(6q<sub>39</sub>,7), ▼8p<sup>2</sup>, alternative ex6 donor site 39-nt upstream+ex7 skipping +alternative ex8 acceptor site 2-nt upstream; ▼8p<sup>2</sup>, alternative ex8 acceptor site 2-nt upstream; Δ9, ex9 skipping.



**Figure S3. Pull-down assays of the wild type and c.67A mutant sequences with an anti-SRp antibody.** Asterisks indicate significant differences of bound proteins in the wild type (red) and mutant sequences (yellow). The wild type (MGBR2\_ex2-9) and mutant (c.67G>A) minigenes were amplified with primers PD\_T7\_EX2-FW (5'-TAATACGACTCACTATAGGGGATCCAAAGAGAGGCCAACA-3') and PD\_UG\_ivs2-RV (5'-C ACACACACACATTCTCGGTGTAATTTATAAAGT-3') using Taq polymerase from NEB (#M0267L) (New England Biolabs, Ipswich, MA). Approximately, 3 µg of PCR product were *in vitro* transcribed during 2 h at 37 °C with T7 RNA polymerase (Agilent Technologies, Santa Clara, CA, USA). Purified RNA was activated using Na-m-periodate 5 mM to bind the adipic acid dihydrazide-agarose (Sigma Aldrich, St Louis, MO, USA). Activated RNA was mixed with HeLa nuclear extract 10 mg/ml (NE) and then washed with 20 mM HEPES, 100 mM KCl, 0.2 mM EDTA, 0.5 mM and 6% glycerol solution. To collect the bound proteins, samples were denaturated at 85 °C during 5 min. Proteins were then electrophoresed in NuPAGE 4–12% Bis-Tris Gel (Thermo Fisher Scientific, Waltham, MA, USA) and Coomassie stained or western-blotted. Western-blots were onto Immobilon-P transfer membrane (Merck KGaA, Darmstadt, Germany) and used mouse anti-SRp Proteins antibody (1H4)(Invitrogen, Carlsbad, CA, USA).



**Canonical Transcript: 91.9%**

**23.3%**

**33.7%**

**Figure S4. Analysis of the predicted RNA secondary structure of the *BRCA2* exon 3 interval between c.90 and c.125.** From left to right: wild type sequence, variant c.97G>A and variant c.100G>A. The  $\Delta G$  of each predicted conformation is indicated. The SRp20 motif UCUUC (c.106-110) and its complementary counterpart GAAGA (c.97-101) (both blue-boxed) are within the main double-stranded structure. Variants c.97G>A and c.100G>A are red-circled. The relative proportion of the canonical transcript is shown below, which, curiously, is inversely correlated with  $\Delta G$  values.



**Table S1.** Primers used for the construction of the MGBR2\_ex2-9 minigene.

| Primer               | Sequence (5'- 3')   | Size (bp) |
|----------------------|---|-----------|
| MGBR2_ex4-7-EcoRI-FW | CACACAGAATTCTCTGGTACATTGGTAGGTTTTCA                         | 2352      |
| MGBR2_ex4-7-XhoI-RV  | CACACACTCGAGCCAAATACTGCCTCTTAAACGC                          |           |
| MGBR2_ex3-EcoRI-FW   | CACACAGAATTCCACACATCCGGACACACATTCTGAAGCCATCTTGTAACTTTTGTGAA | 808       |
| MGBR2_ex3-EcoRI-RV   | CACACAGAATTCTTGGCATAAGAAAGACGTAATT                          |           |
| MGBR2_ex8-XhoI-FW    | CACACACTCGAGGTCATAACCTCACAGCATCATCT                         | 572       |
| MGBR2_ex8-KpnI-RV    | CACACAGGTACCCACACAAAGCTTCAACTTTAACAGAGAGACAGCA              |           |
| MGBR2_ex9-HindIII-FW | CACACAAAGCTTTCCTAGTGGTGCAAGATTTTCAT                         | 649       |
| MGBR2_ex9-KpnI-RV    | CACACAGGTACCTGGTGCCAATTAAGAGAGTAG                           |           |
| MGBR2_ex2-EagI-FW    | CACACACGGCCGAAACTAAATCGTATGAAAATCCT                         | 739       |
| MGBR2_ex2-BspEI-RV   | CACACATCCGGATAAGCTACATAAACAAAAGCAC                          |           |

**Table S2.** Mutagenesis primers of 84 DNA variants and 26 microdeletions

| DNA variants (HGVS) | Primers for Site-directed Mutagenesis (5'- 3')  |
|---------------------|---|
| c.48dup             | ACATTTTTTGAATTTTTAAGGACACGCTGCAACAAAGCA<br>CTGCTTTGTTGCAGCGTGTCTTAAAAATTTCAAAAAATG                  |
| c.62A>G             | TTTTAAGACACGCTGCAACAGAGCAGGTATTGACAAATTT<br>AAATTTGTCAATACCTGCTCTGTTGCAGCGTGTCTTAAAA                |
| c.67G>A             | AGACACGCTGCAACAAAGCAAGTATTGACAAATTTTATATA<br>TATATAAAATTTGTCAATACTTGCTTTGTTGCAGCGTGTCT              |
| c.67+1G>T           | CACGCTGCAACAAAGCAGTTATTGACAAATTTTAT<br>ATAAAATTTGTCAATAACTGCTTTGTTGCAGCGTG                          |
| c.67+2T>C           | CACGCTGCAACAAAGCAGGCATTGACAAATTTTATATAACT<br>AGTTATATAAAATTTGTCAATGCCTGCTTTGTTGCAGCGTG              |
| c.67+3A>G           | CACGCTGCAACAAAGCAGGTGTTGACAAATTTTATATAACT<br>AGTTATATAAAATTTGTCAACACCTGCTTTGTTGCAGCGTG              |
| c.68-7T>A           | AAGGTGGGATTTTTTTTTTAAATAGATTAGGACCAA<br>TTGGTCCTAAATCTATTTTAAAAAATCCCACCTT                          |
| c.71T>A             | ATTTTTTTTTTAAATAGATTAAGGACCAATAAGTCTTAATT<br>AATTAAGACTTATTGGTCCTTAATCTATTTAAAAAAT                  |
| c.79A>G             | TTTAAATAGATTTAGGACCAGTAAGTCTTAATTGGTTTGAA<br>TTCAAACCAATTAAGACTTACTGGTCCTAAATCTATTTAAA              |
| c.92G>A             | TTAGGACCAATAAGTCTTAATTAGTTTGAAGAAGCTTTCT<br>AGAAAGTTCTTCAAACCTAATTAAGACTTATTGGTCCTAA                |
| c.92G>C             | TTAGGACCAATAAGTCTTAATTCGTTTGAAGAAGCTTTCT<br>AGAAAGTTCTTCAAACGAATTAAGACTTATTGGTCCTAA                 |
| c.97G>A             | CAATAAGTCTTAATTGGTTTAAAGAAGCTTTCTCAGAAGCT<br>AGCTTCTGAAGAAAGTTCTTTAAACCAATTAAGACTTATTG              |
| c.100G>A            | AAGTCTTAATTGGTTTGAAGAACTTTCTCAGAAGCTCCAC<br>GTGGAGCTTCTGAAGAAAGTTTTCAAACCAATTAAGACTT                |
| c.121C>T            | AACTTTCTCAGAAGCTCCATCCTATAATTCTGAACCTGCA<br>TGCAGGTTTCAAGATTATAGGATGGAGCTTCTGAAGAAAGTT              |
| c.140C>T            | ACCCTATAATTCTGAACCTGTAGAAGAATCTGAACATAAAA<br>TTTTATGTTTCTGAGTTCTTCTACAGGTTTCAAGATTATAGGGT           |
| c.143A>G            | CTATAATTCTGAACCTGCAGGAGAATCTGAACATAAAAAACA<br>TGTTTTTATGTTTCTGAGTTCTCTGCAGGTTTCAAGATTATAG           |
| c.144del            | TATAATTCTGAACCTGCAGAGAATCTGAACATAAAAAACA<br>TTGTTTTTATGTTTCTGAGTTCTCTGCAGGTTTCAAGATTATA             |
| c.215A>G            | ACAAAGGAAACCATCTTATAGTCAGCTGGCTTCAACTCCAA<br>TTGGAGTTGAAGCCAGCTGACTATAAGATGGTTTCCTTTGT              |
| c.217C>T            | AAAGGAAACCATCTTATAATTAGCTGGCTTCAACTCCAATA<br>TATTGGAGTTGAAGCCAGCTAATTATAAGATGGTTTCCTTT              |
| c.222G>A            | AAACCATCTTATAATCAGCTAGCTTCAACTCCAATAATATT<br>AATATTATTGGAGTTGAAGCTAGCTGATTATAAGATGGTTT              |
| c.244A>T            | CTTCAACTCCAATAATATTCTAAGAGCAAGGGCTGACTCTG<br>CAGAGTCAGCCCTTGCTCTTAGAATATTATTGGAGTTGAAG              |
| c.250C>T            | CTCCAATAATATTCAAAGAGTAAGGGCTGACTCTGCCGCTG<br>CAGCGGCAGAGTCAGCCCTTACTCTTTGAATATTATTGGAG              |
| c.316+2T>C          | TTCAAATTAGACTTAGGCAAGTAATGCAATATGGTA<br>TACCATATTGCATTACTTGCCTAAGTCTAATTTGAA                        |
| c.316+3del          | TAAATTCAAATTAGACTTAGGTAGTAATGCAATATGGTAGAC<br>GTCTACCATATTGCATTACTACCTAAGTCTAATTTGAATTTA            |
| c.316+5G>C          | TTCAAATTAGACTTAGGTAACTAATGCAATATGGTAGACTG<br>CAGTCTACCATATTGCATTAGTTACCTAAGTCTAATTTGAA              |
| c.316+6T>C          | TCAAATTAGACTTAGGTAAAGCAATGCAATATGGTAGACTGG<br>CCAGTCTACCATATTGCATTGCTTACCTAAGTCTAATTTGA             |
| c.317-2A>G          | TGAATTATTGTAAGTTTCGGGAAGGAATGTTCCCAATAGT<br>ACTATTGGGAACATTCTTCCCGAAACAGTACAATAATTCA                |
| c.352C>T            | ATAGTAGACATAAAAGTCTTTGCACAGTGAAAACTAAAAATG<br>CATTTTAGTTTCACTGTGCAAGACTTTTATGTCTACTAT               |
| c.356C>G            | TAGACATAAAAGTCTTCGCAGAGTGAAAACTAAAAATGGATC<br>GATCCATTTTAGTTTTCACTCTGCGAAGACTTTTATGTCTA             |
| c.368_372del        | AAAGTCTTCGCACAGTGAAAACTAGATCAAGCAGATGATGTTTCCTGT<br>ACAGGAAACATCATCTGCTTATGCTAGTTTCACTGTGCGAAGACTTT |
| c.375T>A            | ACAGTGAAAACTAAAAATGGAACAAGCAGATGATGTTTCCTG  |

|                   |  |
|-------------------|--|
|                   | CAGGAAACATCATCTGCTTGTTCCATTTTAGTTTTCACTGT  |
| c.378A>G          | GTGAAAACTAAAATGGATCAGGCAGATGATGTTTCCTGTCC<br>GGACAGGAAACATCATCTGCCTGATCCATTTTAGTTTTCAC                   |
| c.379G>T          | TGAAAACTAAAATGGATCAATCAGATGATGTTTCCTGTCCA<br>TGGACAGGAAACATCATCTGATTGATCCATTTTAGTTTTCA                   |
| c.383A>G          | AACTAAAATGGATCAAGCAGGTGATGTTTCCTGTCCACTTC<br>GAAGTGGACAGGAAACATCACCTGCTTGATCCATTTTAGTT                   |
| c.394_398del      | ATGGATCAAGCAGATGATGTTTCCACTTCTAAATTCCTGTCTTAGTGA<br>TCACTAAGACAAGAATTTAGAAGTGGAAACATCATCTGCTTGATCCAT     |
| c.407del          | TGTTTCCTGTCCACTTCTAATTCTTGCTTAGTGAAAGGT<br>ACCTTTCACATAAGACAAGAATTAGAAGTGGACAGGAAACA                     |
| c.426-12_426-8del | TAAAATAACCTAAGGGATTTGCTTTATTTTAGTCCTGTTGTTCTACAATG<br>CATTGTAGAACAACAGGACTAAAATAAAGCAAATCCCTTAGGTTATTTTA |
| c.426-2A>T        | GATTTGCTTTGTTTTATTTTTGTCCTGTTGTTCTACAATGT<br>ACATTGTAGAACAACAGGACAAAAATAAAACAAAGCAAATC                   |
| c.438A>G          | TATTTTAGTCCTGTTGTTCTGCAATGTACACATGTAACACC<br>GGTGTTACATGTGTACATTGCAGAACAACAGGACTAAAATA                   |
| c.440A>G          | TTTTAGTCCTGTTGTTCTACGATGTACACATGTAACACCAC<br>GTGGTGTTACATGTGTACATCGTAGAACAACAGGACTAAAA                   |
| c.440A>T          | TTTTAGTCCTGTTGTTCTACTATGTACACATGTAACACCAC<br>GTGGTGTTACATGTGTACATAGTAGAACAACAGGACTAAAA                   |
| c.441A>G          | TTTAGTCCTGTTGTTCTACAGTGTACACATGTAACACCACA<br>TGTGGTGTTACATGTGTACACTGTAGAACAACAGGACTAAA                   |
| c.447A>G          | CCTGTTGTTCTACAATGTACGCATGTAACACCACAAAGAGA<br>TCTCTTTGTTGGTGTTACATGCGTACATTGTAGAACAACAGG                  |
| c.451G>A          | TGTTCTACAATGTACACATATAACACCACAAAGAGATAAGT<br>ACTTATCTCTTTGTGGTGTTATATGTGTACATTGTAGAACA                   |
| c.467A>G          | CATGTAACACCACAAAGAGGTAAGTCAGGTATGATTAAA<br>TTTAATCATACCTGACTTACCTCTTTGTGGTGTTACATG                       |
| c.470_474del      | ACACATGTAACACCACAAAGAGATAGGTATGATTAAAAACAATGCTTTTT<br>AAAAAGCATTGTTTTTAATCATACCTATCTCTTTGTGGTGTTACATGTGT |
| c.475+1G>T        | CACAAAGAGATAAGTCAGTTATGATTAAAAACAATGC<br>GCATTGTTTTTAATCATAACTGACTTATCTCTTTGTG                           |
| c.475+3A>T        | CACAAAGAGATAAGTCAGGTTTGATTAAAAACAATGCTTTT<br>AAAAGCATTGTTTTTAATCAAACCTGACTTATCTCTTTGTG                   |
| c.476-2A>G        | TCCCCTTTTTTTACCCCCGGTGGTATGTGGGAGTT<br>AACTCCACATACCACCGGGGTAAAAAAGGGGA                                  |
| c.476-3C>A        | TTTTCCCCTTTTTTACCCCCAAGTGGTATGTGGGAGTTTGT<br>ACAAACTCCACATACCACCTTGGGGTAAAAAAGGGGAAAA                    |
| c.478G>A          | CCCTTTTTTTTACCCCCAGTGATATGTGGGAGTTTGTTCAT<br>ATGAAACAAACTCCACATATCACTGGGGGTAAAAAAGGG                     |
| c.506A>G          | GAGTTTGTTCATACACCAAGGTTTGTGAAGGTAAATATTC<br>GAATATTTACCTTCACAAACCTTGGTGTATGAAACAAACTC                    |
| c.508T>G          | GTTTGTTCATACACCAAGGTTTGTGAAGGTAAATATTCTA<br>TAGAATATTTACCTTCACAACTTTGGTGTATGAAACAAAC                     |
| c.516+1G>T        | CACCAAAGTTTGTGAAGTTAAATATTCTACCTGGT<br>ACCAGGTAGAATATTTAACTTCACAACTTTGGTG                                |
| c.516+2T>C        | TACACCAAAGTTTGTGAAGGCAAATATTCTACCTGGTTTAT<br>ATAAACAGGTAGAATATTTGCCTTCACAACTTTGGTGTA                     |
| c.516+4_516+5del  | ATACCAAAGTTTGTGAAGGTATATTCTACCTGGTTTATTTT<br>AAAATAAACAGGTAGAATATACCTTCACAACTTTGGTGT                     |
| c.517-2A>G        | AAACTATTTTCTTCTCCCGGGGTCGTCAGACACCAAAAC<br>GTTTTGGTGTCTGACGACCCCGGGAGGAAAGAAAATAGTTT                     |
| c.517-1G>A        | AACTATTTTCTTCTCCCAAGGTCGTCAGACACCAAAACA<br>TGTTTTGGTGTCTGACGACCTTGGGAGGAAAGAAAATAGTT                     |
| c.517G>T          | ACTATTTTCTTCTCCCAAGTGTCGTCAGACACCAAAACAT<br>ATGTTTTGGTGTCTGACGACACTGGGAGGAAAGAAAATAGT                    |
| c.572A>G          | AGGAGCTGAGGTGGATCCTGGTATGTCTTGGTCAAGTTCTT<br>AAGAAGTTGACCAAGACATACCAGGATCCACCTCAGCTCCT                   |
| c.574_575del      | GGAGCTGAGGTGGATCCTGATGTCTTGGTCAAGTTCTTTAG<br>CTAAAGAAGTTGACCAAGACATCAGGATCCACCTCAGCTCC                   |
| c.599C>T          | TTGGTCAAGTTCTTTAGCTATACCACCCACCCTTAGTTCTA<br>TAGAACTAAGGGTGGGTGGTATAGCTAAAGAAGTTGACCAA                   |
| c.606C>T          | AGTTCTTTAGCTACACCACCTACCCTTAGTTCTACTGTGCT<br>AGCACAGTAGAACTAAGGGTAGGTGGTGTAGCTAAAGAAGT                   |
| c.617C>G          | TACACCACCCACCCTTAGTTGTACTGTGCTCATAGGTAATA  |

|                                   |  |
|-----------------------------------|--|
|                                   | TATTACCTATGAGCACAGTACAACCTAAGGGTGGGTGGTGTGA  |
| c.620C>T                          | ACCACCCACCCCTTAGTTCTATTGTGCTCATAGGTAATAATA<br>TATTATTACCTATGAGCACAACTAGAACTAAGGGTGGGTGGT   |
| c.625C>T                          | CCACCCCTTAGTTCTACTGTGTTTCATAGGTAATAATAGCAAA<br>TTTGCTATTATTACCTATGAACACAGTAGAACTAAGGGTGG   |
| c.627C>A                          | ACCCTTAGTTCTACTGTGCTAATAGGTAATAATAGCAAATG<br>CATTTGCTATTATTACCTATTAGCACAGTAGAACTAAGGGT   |
| c.627C>G                          | ACCCTTAGTTCTACTGTGCTGATAGGTAATAATAGCAAATG<br>CATTTGCTATTATTACCTATCAGCACAGTAGAACTAAGGGT   |
| c.627C>T                          | ACCCTTAGTTCTACTGTGCTTATAGGTAATAATAGCAAATG<br>CATTTGCTATTATTACCTATAAGCACAGTAGAACTAAGGGT   |
| c.631G>A                          | TTAGTTCTACTGTGCTCATAAGTAATAATAGCAAATGTGTA<br>TACACATTGCTATTATTACTTATGAGCACAGTAGAACTAA  |
| c.631+1G>A                        | GTTCTACTGTGCTCATAGATAATAATAGCAAATGTG<br>CACATTTGCTATTATTATCTATGAGCACAGTAGAAC   |
| c.631+3A>G                        | GTTCTACTGTGCTCATAGGTGATAATAGCAAATGTGTATT<br>AAATACACATTGCTATTATCACCTATGAGCACAGTAGAAC   |
| c.632-3C>G                        | ACACATAAATTTTTATCTTAGAGTCAGAAATGAAGAAGCA<br>TGCTTCTTCATTTCTGACTCTAAGATAAAAAATTTATGTGT  |
| c.632-1G>C                        | AAATTTTTATCTTACACTCAGAAATGAAGAAGCATC<br>GATGCTTCTTCATTTCTGAGTGTAAGATAAAAAATTT  |
| c.681+4A>G                        | CTCATGATACTACTGCTGTAGGTAAATATGACATTGATTAG<br>CTAATCAATGTCATATTTACCTACAGCAGTAGTATCATGAG   |
| c.682-1G>C                        | ACTATAATTTTTGCACAAATGTGAAAAGCTATTTTTCC<br>GGAAAAATAGCTTTTCACATTGTGCAAAAATTATAGT  |
| c.709G>T                          | AAAGCTATTTTTCCAATCATTATGAAAGTCTGAAGAAAAAT<br>ATTTTCTTCAGACTTTTCATAATGATTGGAAAAATAGCTTT   |
| c.741C>T                          | AAGAAAAATGATAGATTTATTGCTTCTGTGACAGACAGTGA<br>TCACTGTCTGTCACAGAAGCAATAAATCTATCATTTTTCTT   |
| c.746C>T                          | AATGATAGATTTATCGCTTTTGTGACAGACAGTGAAAAACA<br>TGTTTTCACTGTCTGTCACAAAAGCGATAAATCTATCATT  |
| c.750G>A                          | GATAGATTTATCGCTTCTGTAACAGACAGTGAAAAACACAAA<br>TTTGTGTTTTCACTGTCTGTTACAGAAGCGATAAATCTATC  |
| c.752C>G                          | TAGATTTATCGCTTCTGTGAGAGACAGTGAAAAACACAAAT<br>ATTTGTGTTTTCACTGTCTCTCACAGAAGCGATAAATCTA  |
| c.756_757delCA                    | ATTTATCGCTTCTGTGACAGAGTGAAAAACACAAATCAAAGAG<br>CTCTTTGATTTGTGTTTTCACTCTGTCACAGAAGCGATAAAT  |
| c.772C>T                          | CAGACAGTGAAAAACACAAATTAAGAGAAGCTGCAAGTCAT<br>ATGACTTGCAGCTTCTCTTTAATTTGTGTTTTCACTGTCTG   |
| c.793+1G>T                        | AGAAGCTGCAAGTCATGTTAAGTCCTCTGTTTAGT<br>ACTAAACAGAGGACTTAACATGACTTGCAGCTTCT   |
| <b>MICRODELETIONS<sup>1</sup></b> |  |
| c.-36_-5del [ex2]                 | AAAGTCATTTTGGTCTTCTGTTTTGCAGACGGTAAAAATGCCTATTGGATCCAAAGAGAG<br>CTCTCTTTGGATCCAATAGGCATTTTTACCGTCTGCAAAACAGAAGACCAAAATGCACCT     |
| c.-11_18del [ex2]                 | TTATTTACCAAGCATTGGAGGAATAGAGAGGCCAACATTTTTGAAATTT<br>AAATTTCAAAAAATGTTGGCCTCTCTATTCTCCAATGCTTGGTAAATAA                           |
| c.14_43del [ex2]                  | TCGTAGGTAAAAATGCCTATTGGATATTTTAAAGACACGCTGCAACAAAG<br>CTTTGTTGCAGCGTGTCTTAAAAATATCCAATAGGCATTTTACCTACGA                          |
| c.35_64del [ex2]                  | GGATCCAAAGAGAGGCCAACATTTTCAGGTATTGACAAATTTTATATAAC<br>GTTATATAAAATTTGTCAATACCTGAAAATGTTGGCCTCTCTTTGGATCC                         |
| c.70_99del [ex3]                  | ACTAAGGTGGGATTTTTTTTTTAAATAGATGAACCTTCTTCAGAAGCTCCACCCTATAAT<br>ATTATAGGGTGGAGCTTCTGAAGAAAGTTTCATCTATTTAAAAAATCCACCTTAGT         |
| c.95_124del [ex3]                 | TAGATTTAGGACCAATAAGTCTTAATTTGGTATAATTTCTGAACCTGCAGAAGAACTCTGAAC<br>GTTTCAGATTCTTCTGCAGGTTGAGAATTATACCAATTAAGACTTATTGGTCTTAAATCTA |
| c.259_288del [ex3]                | ACTCCAATAATATTCAAAGAGCAAGGGCTGGAATTAGATAAATTCAAATTAGACTTAGGT<br>ACCTAAGTCTAATTTGAATTTATCTAATTCAGCCCTTGCTCTTTGAATATTATTGGAGT      |
| c.284_313del [ex3]                | GGCTGACTCTGCCGCTGTACCAATCTCTGTAGGTAAGTAATGCAATATGGTAGACTGGG<br>CCCAGTCTACCATATTGCATTACTTACCTACAGGAGATTGGTACAGCGGCAGAGTCAGCC      |
| c.319_348del [ex4]                | TTCTACTGAATTTATTGTACTGTTTCAGGACTTCGCACAGTGAAAACTAAAATGGATCAA<br>TTGATCCATTTTAGTTTTCACTGTGCGAAGTCTGAAACAGTACAATAATTCAGTGAGAA      |
| c.344_373del [ex4]                | CAGGAAGGAATGTTCCCAATAGTAGACATAATCAAGCAGATGATGTTTCCTGTCCACTTC<br>GAAGTGACAGGAAACATCATCTGCTTGATTATGTCTACTATTGGGAACATTCTTCTCTG      |
| c.368_397del [ex4]                | GACATAAAAGTCTTCGCACAGTGAAAACTACACTTCTAAATTTCTGTCTTAGTGAAAGGT<br>ACCTTTCACTAAGACAAAGAATTTGAAGGTGAGTTTTCACTGTGCGAAGACTTTTATGTC     |
| c.393_422del [ex4]                | AACTAAAATGGATCAAGCAGATGATGTTTCAAGGTATGATGAAGCTATTATATTAAATA<br>TATTTTAATATAATAGCTTCATCATACCTTGAAACATCATCTGCTTGATCCATTTTAGTT      |
| c.428_452del [ex5]                | CTAAGGGATTTGCTTTGTTTTATTTTAGTCAACACCACAAAGAGATAAGTCAGGTATGAT<br>ATCATACCTGACTTATCTTTGTGGTGTGACTAAAAATAAAACAAAGCAAATCCCTTAG       |

|                    |  |
|--------------------|--|
| c.448_472del [ex5] | TATTTTAGTCCTGTTGTTCTACAATGTACACAGGTATGATTA AAAACAATGCTTTTATT<br>AATAAAAAGCATTGTTTTTAATCATACCTGTGTACATTGTAGAACAACAGGACTAAAATA |
| c.478_497del [ex6] | TTTTCCCTTTTTTTACCCCCAGTGACACCAAAGTTTGTGAAGGTAAATA<br>TATTACCTTCACAAACTTTGGTGTCACTGGGGGTAAAAAAGGGGAAAA                        |
| c.494_513del [ex6] | CCCCCAGTGGTATGTGGGAGTTTGTAAGGTAAATATTCTACCTGGTTTAT<br>ATAAACCAGGTAGAAATATTTACCTTACAACTCCACATACCACTGGGGG                      |
| c.519_548del [ex7] | AAAAATAAACTATTTTCTTCTCCAGGGTCTAGGAGCTGAGGTGGATCCTGATATGTC<br>GACATATCAGGATCCACCTCAGCTCCTAGACCCTGGGAGGAAAGAAAATAGTTTATTTTT    |
| c.544_573del [ex7] | CAGGGTCGTCAGACACCAAACATATTTCTATGTCTTGGTCAAGTTCTTTAGCTACACCA<br>TGGTGTAGCTAAAGAACTTGACCAAGACATAGAAATATGTTTTGGTGTCTGACGACCCTG  |
| c.574_603del [ex7] | GAAAGTCTAGGAGCTGAGGTGGATCCTGATCCACCCCTAGTTCTACTGTGCTCATAGGT<br>ACCTATGAGCACAGTAGAACTAAGGGTGGGATCAGGATCCACCTCAGCTCCTAGACTTTC  |
| c.599_628del [ex7] | TCTAGGAGCTGAGGTGGATCCTGATCCACCCCTAGTTCTACTGTGCTCA<br>TGAGCACAGTAGAACTAAGGGTGGGATCAGGATCCACCTCAGCTCCTAGA                      |
| c.634_658del [ex8] | ACACATAAAATTTTTATCTTACAGTCTATTTCTCATGATACTACTGCTGT<br>ACAGCAGTAGTATCATGAGGAAATAGACTGTAAGATAAAAAATTTATGTGT                    |
| c.654_678del [ex8] | CAGTCAGAAATGAAGAAGCATCTGAGCTGTAAGTAAATATGACATTGATT<br>AATCAATGTCATATTTACTTACAGCTCAGATGCTTCTTCATTTCTGACTG                     |
| c.684_713del [ex9] | TAGTGATTTTAAACTATAATTTTTGCAGAAAAGTCTGAAGAAAAATGATAGATTTATCGC<br>GCGATAAATCTATCATTTTTCTTCAGACTTTTCTGCAAAAATTATAGTTTAAATCACTA  |
| c.709_738del [ex9] | CAGAAATGTGAAAAGCTATTTTTCCAATCATATCGCTTCTGTGACAGACAGTGAAAACACA<br>TGTGTTTTCACTGTCTGTACAGAAGCGATATGATTGGAAAAATAGCTTTTCACATTCTG |
| c.736_765del [ex9] | TGAAAGTCTGAAGAAAAATGATAGAACAATCAAAGAGAAGCTGCAAGTC<br>GACTTGCAGCTTCTCTTTGATTTGTTCTATCATTTTTCTTCAGACTTTCA                      |
| c.761_790del [ex9] | ATAGATTTATCGCTTCTGTGACAGACAGTGATGGTAAGTCCTCTGTTTAGTTGAACTACA<br>TG TAGTTCAACTAAACAGAGGACTTACCATCACTGTCTGTACAGAAGCGATAAATCTAT |

<sup>1</sup> The first two and the last three nucleotides of each exon were always preserved

**Table S3.** siRNAs against SR proteins, and qPCR primer sequences.

| siRNA anti-SRprotein   | Sequence (5'→3')         |
|------------------------|--------------------------|
| <b>SRSF1 (SF2/ASF)</b> | ACGAUUGCCGCAUCUACGU      |
|                        | ACGUAGAUGCGGCAAUCGU      |
| <b>SRSF2 (SC35)</b>    | AAUCCAGGUCGCGAUCGAA      |
|                        | UUCGAUCGCGACCUGGAUU      |
| <b>SRSF3 (SRp20)</b>   | GAGUGGAACUGUCGAAUUGG     |
|                        | CCAUUCGACAGUUCCACUC      |
| <b>SRSF5 (SRp40)</b>   | CCACCUGUAAGAACAGAAA      |
|                        | UUUCUGUUCUACAGGUGG       |
| <b>SRSF7 (9G8)</b>     | GGAGAGUUAGAAAGGGCUU      |
|                        | AAGCCCUUUCUACUCUCC       |
| <b>SRSF9 (SRp30c)</b>  | GGAAUAUGCCCUGCGUAAA      |
|                        | UUUACGCAGGGCAUAUUC       |
| <b>Tra2β</b>           | GGAGGAUACAGAUACGUU       |
|                        | AACGUGAUCUGUAUCCUCC      |
| <b>LUCIFERASE</b>      | UAAGGCUAUGAAGAGAUAC      |
|                        | GUAUCUCUUAUAGCCUUA       |
| qPCR primers           | Sequence (5'→3')         |
| <b>SRSF1 (SF2)</b>     | CATCGACCTCAAGAATCGCC     |
|                        | CCCATCGTAATCATAGCCGT     |
| <b>SRSF2 (SC35)</b>    | TTCGCCTTCGTTGCTTTC       |
|                        | CCGTAGCGCGCCATTTGCA      |
| <b>SRSF3 (SRp20)</b>   | TGCTTTTGTGAATTTGAAGA     |
|                        | CATTCGACAGTTCCACTCTT     |
| <b>SRSF5 (SRp40)</b>   | AATGATAGACGAAATGCTCC     |
|                        | CCGCAAACGTTACTTCCCC      |
| <b>SRSF7 (9G8)</b>     | GGGCTTTTCAGTTATTATGGT    |
|                        | GGGCAGGTGGTCTATCAAAA     |
| <b>SRSF9 (SRp30c)</b>  | CTGGGGATGTCTGTTATGCT     |
|                        | TGAGAGCGGAATTTGGTGT      |
| <b>Tra2β</b>           | TTGATGGGCGTAGGATCAGAGTTG |
|                        | TCCTCTGTCATAGTAATCCCGACG |
| <b>GAPDH</b>           | TGCCAAATATGATGACATCAAGAA |
|                        | GGAGTGGGTGTCGCTGTTG      |

**Table S4.** Quantification of transcripts from spliceogenic microdeletion assays.

|              | <b>Canonical Transcript</b> | <b>Δ3</b>         | <b>Δ4</b>         | <b>Δ5</b>         | <b>Δ6</b>         | <b>Δ7</b>         | <b>Δ8</b>         | <b>Other aberrant transcripts</b>                                       |
|--------------|-----------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|---|
| <b>Exon3</b> |                             |                   |                   |                   |                   |                   |                   |   |
| c.95_124del  | <b>45.3%±0.1%</b>           | <b>46.3%±0.2%</b> |                   |                   |                   |                   |                   | <b>8.3%±0.1%</b> (942-nt transcript, Δ3p75?)                            |
| <b>Exon4</b> |                             |                   |                   |                   |                   |                   |                   |   |
| c.344_373del | 35.9%±0.5%                  |                   | 53.1%±0.4%        |                   |                   |                   |                   | <b>7.3%±0.1%</b> -Δ(4,5)<br>3.7%±0.1%-Δ(4,5,6)                          |
| c.368_397del | <b>13.9%±0.1%</b>           |                   | <b>75.8%±0.4%</b> |                   |                   |                   |                   | <b>7.7%±0.2%</b> -Δ(4,5)<br>2.6%±0.1%-Δ(4,5,6)                          |
| <b>Exon5</b> |                             |                   |                   |                   |                   |                   |                   |   |
| c.428_452del |                             |                   |                   | <b>100%</b>       |                   |                   |                   |   |
| c.448_472del | <b>3.4%±1.6%</b>            |                   |                   | <b>96.6%±1.6%</b> |                   |                   |                   |   |
| <b>Exon6</b> |                             |                   |                   |                   |                   |                   |                   |   |
| c.494_513del | <b>22.4%±0.1%</b>           |                   |                   |                   | <b>53.0%±0.3%</b> |                   |                   | <b>24.6%±0.2%</b> (Δ5,6)  |
| <b>Exon7</b> |                             |                   |                   |                   |                   |                   |                   |   |
| c.544_573del | <b>3.7%±0.8%</b>            |                   |                   |                   |                   | <b>96.3%±0.7%</b> |                   |   |
| <b>Exon8</b> |                             |                   |                   |                   |                   |                   |                   |   |
| c.634_658del |                             |                   |                   |                   |                   |                   | <b>72.6%±1.0%</b> | <b>18.9%±1.0%</b> -Δ(6q39,7,8)<br>2.4%±0.1%(852nt)*<br>~6.0% Others**   |
| c.654_678del | <b>9.4%±0.2%</b>            |                   |                   |                   |                   |                   | <b>85.2%±0.8%</b> | <b>22.7%±0.7%</b> -Δ(6q39,7,8)<br>5.5%±0.5%-Δ(6q7)<br>4.2%±0.1%(917nt)* |

\*Not characterized peak, size from PeakScanner Software v1.0 (Applied Biosystem™)

\*\*Other minority transcripts that represent <2.0%.

\*\*\* Transcripts were annotated according to previous reports of the ENIGMA consortium (Fackenthal *et al* [17]): Δ, skipping or deletion; ▼, insertion; p, alternative acceptor site; q, alternative donor site; subscript number, number of deleted nt; superscript number, number of inserted nt. For example, Δ2 indicates exon 2 skipping whereas Δ2q<sub>4</sub> indicates loss of 4 nt at the exon 2 donor site.

**Table S5.** Bioinformatics analysis of reported variants of *BRCA2* exons 2 to 9.

| DNA variant <sup>1</sup> | Type                          | Previous Classification | Bioinformatics <sup>2,3</sup>       |  |                              |
|--------------------------|-------------------------------|-------------------------|-------------------------------------|--|------------------------------|
|                          |                               |                         | Acceptor/Donor                      | ESEs   | ESSs                         |
| EXON 2 [c.-39-c.67]      |                               |                         |                                     |  |                              |
| c.48dup                  | Frameshift<br>p.Thr17Aspfs*14 | Causal                  |                                     |  | [+] hnRNPAI (74.53)          |
| c.62A>G                  | Missense<br>p.Lys21Arg        | VUS                     |                                     |  | [+] hnRNPAI (71.43)          |
| c.67G>A                  | Missense<br>p.Asp23Asn        | Causal                  | [-] 5'SS (NNS: 0.81 →<0.4)          |  |                              |
| c.67+1G>T                | Intronic                      | VUS                     | [-] 5'SS (NNS: 0.81 →<0.4)          |  |                              |
| c.67+2T>C                | Intronic                      | Causal                  | [-] 5'SS (NNS: 0.81 →<0.4)          |  |                              |
| c.67+3A>G                | Intronic                      | VUS                     | [-] 5'SS (NNS: 0.81 →<0.4)          |  |                              |
| EXON 3 [c.68-c.316]      |                               |                         |                                     |  |                              |
| c.68-7T>A                | Intronic                      | VUS                     | [↓] 3'SS (NNS: 0,94 → 0,68)         |  |                              |
| c.71T>A                  | Nonsense<br>p.Leu24*          | Causal                  |                                     |  | [+] hnRNPAI (82.86)          |
| c.79A>G                  | Missense<br>p.Ile27Val        | VUS                     | [+]5'SS 238-nt upstream (NNS: 0.96) |  |                              |
| c.92G>A                  | Nonsense<br>p.Trp31*          | Causal                  |                                     | [-]SC35 (75.78)  | [+] hnRNPAI (72.14)          |
| c.92G>C                  | Missense<br>p.Trp31Ser        | VUS                     |                                     | [-]SC35 (75.78)  |                              |
| c.97G>A                  | Missense<br>p.Glu33Lys        | VUS                     |                                     | [-]SC35 (75.78) – μDEL+  |                              |
| c.100G>A                 | Missense<br>p.Glu34Lys        | VUS                     |                                     | [-] SRp40 (78.14) – μDEL+  |                              |
| c.121C>T                 | Missense<br>p.Pro41Ser        | VUS                     |                                     | [+] SRp55 (77.45); [-] SF2/ASF (IgM-BRCA1) (82.69); [-]SF2/ASF (81.89) – μDEL+ |                              |
| c.140C>T                 | Missense<br>p.Ala47Val        | VUS                     |                                     |  | [+] hnRNPAI (77.38)          |
| c.143A>G                 | Missense<br>p.Glu48Gly        | VUS                     |                                     |  | 2 [+] hnRNPAI (73.57, 67.86) |
| c.144del                 | Frameshift<br>p.Glu49Asnfs*31 | Likely causal           |                                     |  | 2 [+] hnRNPAI (73.57, 67.86) |
| c.215A>G                 | Missense<br>p.Asn72Ser        | VUS                     |                                     |  | 2 [+] hnRNPAI (68.33, 74.76) |
| c.217C>T                 | Nonsense<br>p.Gln73*          | Causal                  |                                     |  | [+] hnRNPAI (70.24)          |
| c.222G>A                 | Synonymous<br>p.Leu74Leu      | VUS                     |                                     |  | [+] hnRNPAI (72.14)          |
| c.244A>T                 | Nonsense<br>p.Lys82*          | Causal                  |                                     |  | [+] hnRNPAI (71.67)          |
| c.250C>T                 | Nonsense<br>p.Gln84*          | Causal                  | [+]5'SS 68-nt upstream (NNS: 0.95)  |  | [+] hnRNPAI (80.00)          |
| c.316+2T>C               | Intronic                      | Causal                  | [-] 5'SS (NNS: 1.00 →<0.4)          |  |                              |
| c.316+3del               | Intronic                      | Causal                  | [-] 5'SS (NNS: 1.00 →<0.4)          |  |                              |
| c.316+5G>C               | Intronic                      | Causal                  | [↓] 5'SS (MaxEnt: 9.66 → 6.43)      |  |                              |

|                             |                                |                |  |   |   |
|-----------------------------|--------------------------------|----------------|--|---|---|
| c.316+6T>C                  | Intronic                       | Causal         | +6T highly conserved (>60%)            | [-] SRp55 (76.10)   |   |
| <b>EXON 4 [c.317-c.425]</b> |                                |                |  |   |   |
| c.317-2A>G                  | Intronic                       | Causal         | [-] 3'SS (NNS: 0.66 → <0.4)            |   |   |
| c.352C>T                    | Missense<br>p.Arg118Cys        | VUS            |  | [-] SF2/ASF (IgM-BRCA1) (83.00);<br>[-] SF2/ASF (78.16) – <b>μDEL+</b>                    |   |
| c.356C>G                    | Missense<br>p.Thr119Arg        | VUS            |  |   | [+] hnRNPAI (72.38) – <b>μDEL+</b>          |
| c.368_372del                | Frameshift<br>p.Lys123Argfs*5  | Causal         |  | [-] SC35 (78.30) – <b>μDEL+</b>   |   |
| c.375T>A                    | Missense<br>p.Asp125Glu        | VUS            |  | [-] SC35 (78.30) – <b>μDEL+</b>   |   |
| c.378A>G                    | Synonymous<br>p.Gln126Gln      | VUS            |  | [-] SC35 (78.30) – <b>μDEL+</b>   |   |
| c.379G>T                    | Missense<br>p.Ala127Ser        | VUS            |  | [-] SC35 (78.30) – <b>μDEL+</b>   |   |
| c.383A>G                    | Missense<br>p.Asp128Gly        | VUS            | [+] 5'SS 43-nt upstream (MaxEnt: 5.99) |   | [+] hnRNPAI (70.24) – <b>μDEL+</b>          |
| c.394_398del                | Frameshift<br>p.Cys132Thrfs*7  | VUS            |  | [-] 2 SC35 (91.09, 79.23) – <b>μDEL+</b>  |   |
| c.407del                    | Frameshift<br>p.Asn136Ilefs*16 | Causal         | [+] 3'SS (NNS: <0.4 → 0.61)            |   |   |
| <b>EXON 5 [c.426-c.475]</b> |                                |                |  |   |   |
| c.426-12_426-8del           | Intronic                       | VUS            | [↓] 3'SS (NNS: 0.99 → 0.79)            |   |   |
| c.426-2A>T                  | Intronic                       | Causal         | [-] 3'SS (NNS: 0.99 → <0.4)            |   |   |
| c.438A>G                    | Synonymous<br>p.Leu146Leu      | VUS            |  | [-] SRp40 (78.32) – <b>μDEL+</b>  |   |
| c.440A>G                    | Missense<br>p.Gln147Arg        | VUS            |  | [-] SRp40 (78.32); [+] SC35 (76.09) – <b>μDEL+</b>  | [+] hnRNPAI (71.43)                         |
| c.440A>T                    | Missense<br>p.Gln147Leu        | VUS            |  | [-] SRp40 (78.32); [+] SC35 (76.34); [+] SRp55 (79.88) – <b>μDEL+</b>                     | [+] hnRNPAI (69.05)                         |
| c.441A>G                    | Synonymous<br>p.Gln147Gln      | VUS            |  |   | [+] 2 hnRNPAI (68.33, 68.57) – <b>μDEL+</b> |
| c.447A>G                    | Synonymous<br>p.Thr149Thr      | Likely Neutral |  | [-] SRp40 (81.62); [-] 9G8 (61.07) – <b>μDEL+</b>   | [+] hnRNPAI (70.72)                         |
| c.451G>A                    | Missense<br>p.Val151Ile        | VUS            |  | [-] SRp40 (81.62); [-] SRp55 (74.56) – <b>μDEL+</b>                                       |   |
| c.467A>G                    | Missense<br>p.Asp156Gly        | VUS            | [+] 5'SS, 9-nt upstream (NNS: 1.0)     |   | [+] hnRNPAI (76.90) – <b>μDEL+</b>          |
| c.470_474del                | Frameshift<br>p.Lys157Serfs*24 | Causal         | [↓] 5'SS (NNS: 0.95 → 0.77)            | [-] SRp55 (78.47); [-] SRp40 (82.51); [-] SF2/ASF (78.68); [-] 9G8 (60.14) – <b>μDEL+</b> |   |
| c.475+1G>T                  | Intronic                       | Causal         | [-] 5'SS (NNS: 0.95 → <0.4)            |   |   |
| c.475+3A>T                  | Intronic                       | VUS            | [↓] 5'SS (NNS: 0.95 → 0.49)            |   |   |
| <b>EXON 6 [c.476-c.516]</b> |                                |                |  |   |   |
| c.476-2A>G                  | Intronic                       | Causal         | [-] 3'SS (NNS: 0.91 → <0.4)            |   |   |
| c.476-3C>A                  | Intronic                       | VUS            | [-] 3'SS (NNS: 0.91 → <0.4)            | [-] 2 SF2/ASF (75.69, 78.68); [-] SF2/ASF (IgM-BRCA1) (78.69); [-] SRp40 (86.05)          |   |
| c.478G>A                    | Missense<br>p.Val160Ile        | VUS            |  | [-] SRp40 (86.05); [-] SF2/ASF (IgM-BRCA1) (78.69); [-] SF2/ASF (78.68)                   |   |
| c.506A>G                    | Missense<br>p.Lys169Arg        | VUS            |  |   | [+] 2 hnRNPAI (66.67, 72.14) – <b>μDEL+</b> |

|                             |                                |                |   |  |   |
|-----------------------------|--------------------------------|----------------|---|--|---|
| c.508T>G                    | Missense<br>p.Phe170Val        | VUS            |   |  | [+] 2 hnRNPAI (66.67, 72.14) – $\mu$ DEL+ |
| c.516+1G>T                  | Intronic                       | Causal         | [-] 5'SS (NNS: 0.98 $\rightarrow$ <0.4)   |  |   |
| c.516+2T>C                  | Intronic                       | Causal         | [-] 5'SS (NNS: 0.98 $\rightarrow$ <0.4)   |  |   |
| c.516+4_516+5del            | Intronic                       | VUS            | [↓] 5'SS (NNS: 0.98 $\rightarrow$ 0.64)   |  |   |
| <b>EXON 7 [c.517-c.631]</b> |                                |                |   |  |   |
| c.517-2A>G                  | Intronic                       | Causal         | [-] 3'SS (NNS: 0.98 $\rightarrow$ <0.4)   |  |   |
| c.517-1G>A                  | Intronic                       | VUS            | [-] 3'SS (NNS: 0.98 $\rightarrow$ <0.4)<br>[+] 3'SS 1-nt downstream (NNS: 0.98) |  |   |
| c.517G>T                    | Missense<br>p.Gly173Cys        | VUS            | [↓] 3' SS (NNS: 0.98 $\rightarrow$ 0.93)  |  |   |
| c.572A>G                    | Missense<br>p.Asp191Gly        | VUS            | [+] 5'SS 60-nt upstream (NNS: 0.99) – $\mu$ DEL+                                |  |   |
| c.574_575del                | Frameshift<br>p.Met192Valfs*13 | Causal         |   |  | [-] hnRNPAI (67.14)                       |
| c.599C>T                    | Missense<br>p.Thr200Ile        | VUS            |   | [-]SRp40 (89.04)   |   |
| c.606C>T                    | Synonymous<br>p.Pro202Pro      | VUS            | [↑] Cryptic 3' SS (NNS: 0.57 $\rightarrow$ 0.70)                                | [-] SF2/ASF (IgM-BRCA1) (80.2); [-] SF2/ASF (79.85)                    |   |
| c.617C>G                    | Missense<br>p.Ser206Cys        | VUS            |   | [-] SC35 (79.10)   |   |
| c.620C>T                    | Missense<br>p.Thr207Ile        | VUS            |   | [-]SRp40 (80.24); [-] SC35 (79.10)                                     |   |
| c.625C>T                    | Missense<br>p.Leu209Phe        | VUS            | [↓] 5' SS (NNS: 0.44 $\rightarrow$ <0.4)  |  |   |
| c.627C>A                    | Synonymous<br>p.Leu209Leu      | Likely Neutral |   | [-] SC35 (75.41); [-] SF2/ASF (IgM-BRCA1) (78.46); [-] SF2/ASF (80.96) |   |
| c.627C>G                    | Synonymous<br>p.Leu209Leu      | Likely Neutral | [↓] 5' SS (NNS: 0.44 $\rightarrow$ <0.4)  |  |   |
| c.627C>T                    | Synonymous<br>p.Leu209Leu      | Likely Neutral | [↓] 5' SS (NNS: 0.44 $\rightarrow$ <0.4)  |  |   |
| c.631G>A                    | Missense<br>p.Val211Ile        | Causal         | [-] 5' SS (NNS: 0.44 $\rightarrow$ <0.4)  |  |   |
| c.631+1G>A                  | Intronic                       | VUS            | [-] 5' SS (NNS: 0.44 $\rightarrow$ <0.4)  |  |   |
| c.631+3A>G                  | Intronic                       | VUS            | [-] 5' SS (NNS: 0.44 $\rightarrow$ <0.4)  |  |   |
| <b>EXON 8 [c.632-c.681]</b> |                                |                |   |  |   |
| c.632-3C>G                  | Intronic                       | VUS            | [↓] 3' SS (NNS: 0.96 $\rightarrow$ 0.66)  |  |   |
| c.632-1G>C                  | Intronic                       | Causal         | [-] 3' SS (NNS: 0.96 $\rightarrow$ <0.4)  |  |   |
| c.681+4A>G                  | Intronic                       | Causal         | [↓] 5' SS (MaxEnt: 8.56 $\rightarrow$ 3.41)                                     |  |   |
| <b>EXON 9 [c.682-c.793]</b> |                                |                |   |  |   |
| c.682-1G>C                  | Intronic                       | Causal         | [-] 3' SS (MaxEnt: 7.06 $\rightarrow$ -0.99)                                    |  |   |
| c.709G>T                    | Missense<br>Asp237Tyr          | VUS            |   | [-] 9G8 (63.36)  | [+] hnRNPAI (72.38)                       |
| c.741C>T                    | Missense<br>p.Ile247Val        | VUS            | [↑] Cryptic 3' SS (NNS:0.4 $\rightarrow$ 0.45)                                  | [-]SRp40 (79.10)   |   |
| c.746C>T                    | Missense<br>p.Ser249Phe        | VUS            | [↑] Cryptic 3' SS (NNS:0.4 $\rightarrow$ 0.42)                                  | [-] SRp55 (76.04); [-] SC35 (81.44)                                    |   |
| c.750G>A                    | Synonymous<br>p.Val250Val      | VUS            | [↑] Cryptic 3' SS (NNS:0.4 $\rightarrow$ 0.43)                                  | [-]SRp40 (80.84)   |   |

|                      |                               |        |  |   |                                     |
|----------------------|-------------------------------|--------|--|---|-------------------------------------|
| c.752C>G             | Missense<br>p.Thr251Arg       | VUS    |  | <b>[-] SRp40 (80.84); [-] SF2/ASF (IgM-BRCA1) (77.85); [-] SF2/ASF (79.32); [-] 9G8 (61.14)</b> | <b>[+] 2 hnRNPA1 (77.14, 67.14)</b> |
| c.756_757del         | Frameshift<br>p.Asp252Glufs*2 | Causal | [↑] Cryptic 3' SS (NNS:<0.4 → 0.57)    | [-] 9G8 (61.07)   | [+] hnRNPA1 (73.33)                 |
| c.772C>T             | Nonsense<br>p.Gln258*         | Causal |  |   | <b>[+] hnRNPA1 (71.43)</b>          |
| <u>c.793+1G&gt;T</u> | Intronic                      | Causal | <b>[-] 5' SS (NNS: 1.00 → &lt;0.4)</b> |   |                                     |

<sup>1</sup> Spliceogenic variants are underlined.

<sup>2</sup>NNS: NNSplice. [+] and [-] symbols indicate creation or disruption of splicing motifs, respectively. [↑] and [↓] denote increase and decrease of the motif scores, respectively. The most relevant data are indicated in bold-type. **μDEL+** designates variants that are placed in positive microdeletion intervals.

<sup>3</sup> **Cut-offs of the splicing programs:**

**Splice Sites:**

NNSPLICE (values 0-1): Cut-off=0.4 for both 5' and 3' splice sites (Reese *et al*, [27]).

HSF: Human Splicing Finder matrices (default values, cut-offs cannot be defined) (Desmet *et al*, [29]).

MaxEnt: 0 for 5' and 3' splice sites (Yeo and Burge, [51]).

**Enhancers and Silencers:**

ESEfinder cut-offs (HSF scale 0-100): SF2/ASF: 72.98 / SF2/ASF (IgM-BRCA1): 70.51 / SRp40: 78.08 / SC35: 75.05 / SRp55: 73.86 (Cartegni *et al*, [52]).

ESE motifs from HSF, cut-offs values (0-100): Tra2: 75.964 / 9G8: 59.245

hnRNP motifs: hnRNPA1: 65.476

**Table S6.** Splicing outcome and quantification of transcripts of the wild type and 83 mutant minigenes.

|                             | <b>Canonical Transcript (CT)</b> | <b><math>\Delta 3</math></b>    | <b><math>\Delta 5</math></b> | <b><math>\Delta(5,6)</math></b> | <b><math>\Delta(6q_{39},7)</math></b>  |
|-----------------------------|----------------------------------|---------------------------------|------------------------------|---------------------------------|--|
| <b>MGBR2_2-9 WT</b>         | 91.9%±0.7%                       | 0.8%±0.7%                       | 3.2%±1.0%                    | 0.6%±0.4%                       | 3.4%±0.5%  |
| <b>Exon 2 [c.-39-c.67]</b>  |                                  |                                 |                              |                                 |  |
| <b>VARIANTS<sup>1</sup></b> | <b>CT</b>                        | <b><math>\Delta 2q_4</math></b> | <b><math>\Delta 2</math></b> | <b><math>\Delta 3</math></b>    | <b>Other transcripts</b>   |
| c.48dup                     | 96.2%±0.1%                       |                                 |                              | 1.3%±0.1%                       | 2.5%±0.1%( $\Delta 5$ )  |
| c.62A>G                     | 93.9%±0.7%                       |                                 |                              |                                 | 2.7%±0.1( $\Delta(6q_{39},7)$ );<br>3.4%±0.7% (857-nt)*  |
| <b><u>c.67G&gt;A</u></b>    | 7.7%±1.2%                        | 89.0%±1.7%                      |                              |                                 | 3.3%±1% (963-nt)*  |
| <b><u>c.67+1G&gt;T</u></b>  |                                  |                                 | 100%                         |                                 |  |
| <b><u>c.67+2T&gt;C</u></b>  | -                                |                                 | 83.7%±0.5%                   |                                 | 0.9%±0.1%(609-nt)*<br>8.7%±0.2%( $\Delta 3,4$ )<br>2.0%±0.1%(757-nt)*<br>4.7%±0.1%(860-nt)*                |
| <b><u>c.67+3A&gt;G</u></b>  | 2.4%±0.2%                        |                                 | 77.6%±3.4%                   |                                 | 1.1%±0.1%(609-nt)*<br>10.5%±1.8%( $\Delta 3,4$ )<br>2.4%±0.5%(757-nt)*<br>6.0%±0.8%(860-nt)*               |
| <b>Exon 3 [c.68-c.316]</b>  |                                  |                                 |                              |                                 |  |
| <b>VARIANTS<sup>1</sup></b> | <b>CT</b>                        | <b><math>\Delta 3</math></b>    |                              |                                 | <b>Other transcripts</b>   |
| c.68-7T>A                   | 90.0%±1.6%                       | 2.8%±0.6%                       |                              |                                 | 7.2%±1.1% ( $\Delta(6q_{39},7)$ )  |
| c.71T>A                     | 97.2%±0.1%                       |                                 |                              |                                 | 2.8%±0.1% ( $\Delta(6q_{39},7)$ )  |
| c.79A>G                     | 97.3%±0.2%                       |                                 |                              |                                 | 2.7%±0.2% ( $\Delta(6q_{39},7)$ )  |
| <b><u>c.92G&gt;A</u></b>    | 52.0%±0.3%                       | 35.5%±0.2%                      |                              |                                 | 3.3%±0.1% ( $\Delta 5$ )<br>1.5%±0.1% ( $\Delta 5,6$ )<br>4.8%±0.3%(715-nt)*<br>2.8%±0.1%(675-nt)*         |
| c.92G>C                     | 94.6%±0.1%                       | 1.6%±0.04%                      |                              |                                 | 1.7%±0.1% ( $\Delta 5$ )<br>2.2%±0.02%( $\Delta(6q_{39},7)$ )  |
| <b><u>c.97G&gt;A</u></b>    | 23.3%±0.5%                       | 71.1%±0.5%                      |                              |                                 | 1.4%±0.1% ( $\Delta(6q_{39},7)$ )<br>4.2%±0.1%(715-nt)*  |
| <b><u>c.100G&gt;A</u></b>   | 33.7%±0.8%                       | 57.0%±0.7%                      |                              |                                 | 5.1%±0.01%(973-nt)*<br>4.2%±0.1%(715-nt)*  |
| c.121C>T                    | 94.5%±0.2%                       |                                 |                              |                                 | 5.5%±0.2%( $\Delta 5$ )  |
| <b><u>c.140C&gt;T</u></b>   | 82.7%±0.1%                       | 13.4%±0.1%                      |                              |                                 | 3.9%±0.1%( $\Delta 5$ )  |
| <b><u>c.143A&gt;G</u></b>   | 70.3%±0.2%                       | 26.9%±0.3%                      |                              |                                 | 2.9%±0.1%(715-nt)*   |
| <b><u>c.144del</u></b>      | 78.7%±0.3%                       | 15.8%±0.3%                      |                              |                                 | 3.4%±0.1%( $\Delta 5$ );<br>2.1%±0.1%( $\Delta(6q_{39},7)$ )   |
| <b><u>c.215A&gt;G</u></b>   | 79.0%±0.3%                       | 8.3%±0.01%                      |                              |                                 | 6.6%±0.2%( $\Delta 5$ )<br>1.8%±0.1% ( $\Delta(6q_{39},7)$ )<br>2.6%±0.1% (927-nt)*<br>1.8%±0.1% (715-nt)* |
| <b><u>c.217C&gt;T</u></b>   | 78.9%±0.3%                       | 21.1%±0.3%                      |                              |                                 |  |
| <b><u>c.222G&gt;A</u></b>   | 62.3%±0.5%                       | 31.5%±0.3%                      |                              |                                 | 2.1%±0.1%( $\Delta 5$ )<br>1.7%±0.1%( $\Delta(6q_{39},7)$ )<br>2.4%±0.1%(715-nt)*                          |
| c.244A>T                    | 91.3%±0.2%                       | 3.9%±0.03%                      |                              |                                 | 2.4%±0.1%( $\Delta 5$ )<br>2.4%±0.1%( $\Delta(6q_{39},7)$ )  |
| c.250C>T                    | 96.5%±0.6%                       |                                 |                              |                                 | 3.5%±0.6%( $\Delta(6q_{39},7)$ )   |
| <b><u>c.316+2T&gt;C</u></b> | -                                | 96.2%±0.2%                      |                              |                                 | 3.8%±0.2%(715-nt)*   |
| <b><u>c.316+3del</u></b>    | -                                | 88.6%±0.2%                      |                              |                                 | 7.0%±0.1% (715-nt)*<br>2.3%±0.1%(675-nt)*  |

|                             |                      |             |               |                 |   |
|-----------------------------|----------------------|-------------|---------------|-----------------|---|
|                             |                      |             |               |                 | 2.1%±0.1% (612-nt)*   |
| <u>c.316+5G&gt;C</u>        | -                    | 90.3%±0.8%  |               |                 | 7.3%±0.5(715-nt)*<br>2.4±0.3%(675-nt)*                                    |
| <u>c.316+6T&gt;C</u>        | 21.9%±0.4%           | 68.7%±0.4%  |               |                 | 7.0%±0.2%(715-nt)*<br>2.4%±0.4%(675-nt)*                                  |
| <b>Exon 4 c.317-c.425</b>   |                      |             |               |                 |   |
| <b>VARIANTS<sup>1</sup></b> | <b>CT</b>            | <b>Δ4</b>   | <b>Δ(4,5)</b> | <b>Δ(4,5,6)</b> | <b>Other transcripts</b>  |
| <u>c.317-2A&gt;G</u>        | -                    | 62.9%±1.9%  | 23.8%±1.4%    | 13.3%±0.5%      |   |
| <u>c.352C&gt;T</u>          | 73.7%±0.6%           | 11.6%±0.2%  | 6.3%±0.1%     | 8.4%±0.3%       |   |
| <u>c.356C&gt;G</u>          | 87.9%±0.4%           |             |               |                 | 6.5%±0.6% (Δ5)<br>2.9%±0.3%(Δ5,6)<br>2.7%±0.1%( Δ(6q <sub>39</sub> ,7))   |
| <u>c.368_372del</u>         | 70.6%±0.1%           | 13.7%±0.1%  | 6.8%±0.1%     | 8.9%±0.1%       | Δ4,5,6) Δ4,5)   |
| <u>c.375T&gt;A</u>          | 93.1%±0.1%           |             |               |                 | 6.9%±0.1% (Δ5)  |
| <u>c.378A&gt;G</u>          | 54.0%±0.4%           | 25.6%±0.2%  | 10.9%±0.1%    | 9.4%±0.3        |   |
| <u>c.379G&gt;T</u>          | 84.1%±0.04%          | 4.0%±0.1%   |               | 4.3%±0.2%       | 5.8%±0.1%(Δ5)<br>1.9%±0.1%(Δ5,6)  |
| <u>c.383A&gt;G</u>          | 78.3%±0.8%           | 3.6%±0.2%   |               | 3.9%±0.2%       | <b>12.1%±0.2% (Δ4q<sub>43</sub>)</b><br>2.1%±0.2% (Δ 5)                   |
| <u>c.394_398del</u>         | 92.0%±1.0%           |             |               |                 | 3.3 %±0.4%(Δ5)<br>3.1%±0.3%(Δ(6q <sub>39</sub> ,7))<br>1.6%±0.2%(817-nt)* |
| <u>c.407del</u>             | 87.8%±0.1%           |             |               |                 | 1.5%±0.03%(Δ3)<br>8.0%±0.1% (Δ5)<br>2.7%±0.1%(927-nt)*                    |
| <b>Exon 5 [c.426-c.475]</b> |                      |             |               |                 |   |
| <b>VARIANTS<sup>1</sup></b> | <b>CT</b>            | <b>Δ5q9</b> | <b>Δ5</b>     | <b>Δ(5,6)</b>   | <b>Other transcripts</b>  |
| <u>c.426-12 426-8del</u>    | 24.7%±0.6%           |             | 72.7%±0.9%    | 1.9%±0.8%       | 0.8%±0.6% (806-nt)*   |
| <u>c.426-2A&gt;T</u>        | -                    |             | 100%          |                 |   |
| <u>c.438A&gt;G</u>          | 95.0%±0.2%           |             |               |                 | 1.8%±0.1%(Δ4)<br>3.2%±0.3%(Δ(6q <sub>39</sub> ,7))                        |
| <u>c.440A&gt;G</u>          | 93.7%±0.4%           |             |               |                 | 1.7%±0.1%(Δ3)<br>4.6%±0.3%(Δ(6q <sub>39</sub> ,7))                        |
| <u>c.440A&gt;T</u>          | 80.1%±0.6%           |             | 15.7%±0.5%    |                 | 2.5%±0.1% (Δ(6q <sub>39</sub> ,7))<br>1.7%±0.1%(817-nt)*                  |
| <u>c.441A&gt;G</u>          | -                    |             | 96.0%±0.3%    | 4.0%±0.3%       |   |
| <u>c.447A&gt;G</u>          | 92.9%±0.6%           |             |               |                 | 4.3%±0.3%(Δ(6q <sub>39</sub> ,7))<br>2.8%±0.4%(Δ3)                        |
| <u>c.451G&gt;A</u>          | -                    |             | 89.2%±0.3%    | 9.6%±0.4%       | 1.3%±0.1% (Δ7,8)  |
| <u>c.467A&gt;G</u>          | -                    | 96.6%±0.1%  |               |                 | 1.5%±0.1%(Δ4,5);<br>1.9%±0.1% (930-nt)*                                   |
| <u>c.470_474del</u>         | 16.7%±0.2%<br>(del5) |             | 83.2%±0.2%    |                 |   |
| <u>c.475+1G&gt;T</u>        | -                    |             | 100%          |                 |   |
| <u>c.475+3A&gt;T</u>        | -                    |             | 96.8%±0.3%    | 3.2%±0.3%       |   |
| <b>Exon 6 [c.476-c.516]</b> |                      |             |               |                 |   |
| <b>VARIANTS<sup>1</sup></b> | <b>CT</b>            | <b>Δ6</b>   | <b>Δ(5,6)</b> |                 | <b>Other transcripts</b>  |
| <u>c.476-2A&gt;G</u>        | -                    | 72.7%±0.6%  | 27.3%±0.6%    |                 |   |
| <u>c.476-3C&gt;A</u>        | 29.8%±0.3%           | 15.0%±0.5%  | 55.3%±0.9%    |                 |   |
| <u>c.478G&gt;A</u>          | 100%                 |             |               |                 |   |
| <u>c.506A&gt;G</u>          | 94.1%±0.3%           |             |               |                 | 1.0%±0.1%(Δ3)<br>1.9%±0.1% (Δ5)<br>3.0%±0.2% (Δ(6q <sub>39</sub> ,7))     |
| <u>c.508T&gt;G</u>          | 95.7%±0.1%           |             |               |                 | 1.7%±0.01%(Δ5)  |

|                             |            |                                  |  |                          |   |
|-----------------------------|------------|----------------------------------|--|--------------------------|---|
|                             |            |                                  |  |                          | 2.6%±0.04% (Δ(6q <sub>39</sub> ,7))   |
| <u>c.516+1G&gt;T</u>        | -          | 85.2%±0.1%                       | 14.8±0.1%  |                          |   |
| <u>c.516+2T&gt;C</u>        | -          | 52.3%±1.8%                       | 47.7%±1.8%   |                          |   |
| <u>c.516+4 516+5del</u>     | 6.4%±0.5%  | 40.9%±1.1%                       | 52.7%±0.9%   |                          |   |
| <b>Exon 7 [c.517-c.631]</b> |            |                                  |  |                          |   |
| <b>VARIANTS<sup>1</sup></b> | <b>CT</b>  | <b>Δ7p<sub>1</sub></b>           | <b>Δ7q<sub>60</sub></b>                                    | <b>Δ7</b>                | <b>Other transcripts</b>  |
| <u>c.517-2A&gt;G</u>        | -          |                                  |  | 93.8%±0.5%               | 3.8%±0.3% (Δ7,8)<br>2.4%±0.2% (651-nt)*                                       |
| <u>c.517-1G&gt;A</u>        | -          | 84.6%±0.2%                       |  | 11.1%±0.3%               | 2.1%±0.1% (Δ5)<br>2.2%±0.3% (856-nt)*   |
| <u>c.517G&gt;T</u>          | -          |                                  |  | 90.8nt±0.2%              | 6.4%±0.1% (Δ7,8)<br>2.8%±0.1% (651-nt)*                                       |
| <u>c.572A&gt;G</u>          | 14.9%±0.2% |                                  | 83.6%±0.3%   |                          | 1.4%±0.1% (Δ(6q <sub>39</sub> ,7))  |
| c.599C>T                    | 90.4%±0.7% |                                  |  |                          | 6.1%±0.4% (872-nt)*<br>3.5%±0.7% (834-nt)*                                    |
| c.606C>T                    | 90.0%±0.6% |                                  |  |                          | 1.3%±0.1% (Δ3)<br>5.7%±0.2% (Δ5)<br>3.1%±0.2% (Δ(6q <sub>39</sub> ,7))        |
| <u>c.617C&gt;G</u>          | 72.4%±0.1% |                                  |  | 23.1%±0.1%               | 4.5%±0.1% (Δ(6q <sub>39</sub> ,7))  |
| <u>c.620C&gt;T</u>          | 78.1%±1.3% |                                  |  | 21.9%±1.3%               |   |
| c.625C>T                    | 92.5%±0.2% |                                  |  |                          | 1.7%±0.1% (Δ3)<br>1.9%±0.1% (Δ5)<br>4.0%±0.1% (Δ(6q <sub>39</sub> ,7))        |
| c.627C>A                    | 95.9%±0.1% |                                  |  |                          | 1.2%±0.1% (Δ5)<br>2.9%±0.1% (Δ(6q <sub>39</sub> ,7))                          |
| c.627C>G                    | 97.1%±1.5% |                                  |  |                          | 2.9%±1.5% (Δ(6q <sub>39</sub> ,7))  |
| c.627C>T                    | 96.5%±0.1% |                                  |  |                          | 3.5%±0.1% (Δ(6q <sub>39</sub> ,7))  |
| <u>c.631G&gt;A</u>          | -          |                                  |  | 91.0%±0.1%               | 4.5%±0.1% (Δ(6q <sub>39</sub> ,7))<br>2.8%±0.1% (Δ7,8)<br>1.7%±0.1% (742-nt)* |
| <u>c.631+1G&gt;A</u>        |            |                                  |  | 91.3%±0.7%               | 8.7%±0.7% (Δ(6q <sub>39</sub> ,7))  |
| <u>c.631+3A&gt;G</u>        | 7.8%±0.5%  |                                  |  | 87.5%±0.5%               | 4.7%±0.1% (Δ7,8)  |
| <b>Exon 8 [c.632-c.681]</b> |            |                                  |  |                          |   |
| <b>VARIANTS<sup>1</sup></b> | <b>CT</b>  | <b>Main aberrant transcripts</b> |  | <b>Other transcripts</b> |   |
| <u>c.632-3C&gt;G</u>        | -          | ▼8p <sup>2</sup> (74.9%±0.6%)    | [Δ(6q <sub>39</sub> ,7) ▼8p <sup>2</sup> ]<br>(25.1%±0.6%) |                          |   |
| <u>c.632-1G&gt;C</u>        | -          | Δ8p <sub>4</sub> (67.9%±0.5%)    | Δ(6q <sub>39</sub> ,7,8p <sub>4</sub> )<br>(30.9%±0.4%)    |                          | 1.2%±0.1% (963-nt)*   |
| <u>c.681+4A&gt;G</u>        | -          | ▼8q <sup>4</sup> (92.7%±0.1%)    |  |                          | 1.8%±0.1% (Δ3)<br>3.5%±0.03% (971-nt)*<br>2.0%±0.1% (868-nt)*                 |
| <b>Exon 9 [c.682-c.793]</b> |            |                                  |  |                          |   |
| <b>VARIANTS<sup>1</sup></b> | <b>CT</b>  | <b>Δ9</b>                        |  |                          | <b>Other transcripts</b>  |
| <u>c.682-1G&gt;C</u>        | -          | 100%                             |  |                          |   |
| c.709G>T                    | 97.3%±0.7% |                                  |  |                          | 2.7%±0.7% (Δ5)  |
| c.741C>T                    | 92.8%±0.2% |                                  |  |                          | 4.7%±0.2% (Δ5)<br>2.5%±0.02% (Δ(6q <sub>39</sub> ,7))                         |
| c.746C>T                    | 100%       |                                  |  |                          |   |
| c.750G>A                    | 100%       |                                  |  |                          |   |
| c.752C>G                    | 97.8%±0.1% |                                  |  |                          | 2.2%±0.1% (Δ(6q <sub>39</sub> ,7))  |
| c.756_757del                | 95.6%±1.0% |                                  |  |                          | 2.2%±0.8% (Δ5);<br>2.2%±0.6% (Δ(6q <sub>39</sub> ,7))                         |
| c.772C>T                    | 97.8%±0.2% |                                  |  |                          | 2.2%±0.2% (Δ(6q <sub>39</sub> ,7))  |
| <u>c.793+1G&gt;T</u>        | -          | 95.1%±0.6%                       |  |                          | 2.8%±0.1% (854-nt)*<br>2.2%±0.5% (744-nt)*                                    |

<sup>1</sup> Spliceogenic variants are underlined. Pathogenic/Likely Pathogenic variants are shown in bold type.

\* Uncharacterized transcript (peak scanner size).

---

**List of Transcripts** (in red those transcripts that introduce PTCs):

**Δ2q<sub>4</sub> [r.64\_67del]**: Alternative ex2 donor site 4-nt upstream (ex2 4-nt deletion).

Δ2 [r.-39\_67del]: Exon 2 skipping (elimination of the Met start codon).

**Δ3 [r.68\_316del]**: Exon 3 skipping.

**Δ3,4 [r.68\_425del]**: Exons 3+4 skipping.

**Δ4[r.317\_425del]**: Exon 4 skipping.

**Δ(4,5,6) [r. 317\_516del]**: Exons 4+5+6 skipping.

**Δ(4,5) [r.317\_475del]**: Exons 4+5 skipping.

**Δ4q<sub>43</sub> [r.383\_425del]**: Alternative ex4 donor site 43-nt upstream (ex4 43-nt deletion).

**Δ5q<sub>9</sub> [r.467\_475 del]**: Alternative ex5 donor site 9-nt upstream (ex5 9-nt deletion).

**Δ5 [r.426\_475del]**: Exon 5 skipping.

**Δ6 [r.476\_516del]**: Exon 6 skipping.

**Δ(6q<sub>39</sub>,7) [r.478\_631del]**: Alternative ex6 donor site 39-nt upstream + ex7 skipping.

**Δ(5,6) [r. 426\_516del]**: Exons 5+6 skipping.

**Δ7p<sub>1</sub> [r.517del]**: Alternative acceptor site 1-nt downstream (ex7 1-nt deletion)

**Δ7q<sub>60</sub> [r.572\_c.631del]**: Alternative ex7 donor site 60-nt upstream (ex7 60-nt deletion)

**Δ7 [r.517\_631del]**: Exon 7 skipping.

**Δ7,8 [r.517\_681del]**: Exons 7+8 skipping.

**▼8p<sup>2</sup> [r.631\_632ins632-1\_632-2]**: Alternative ex8 acceptor site 2-nt upstream (ex8 2-nt insertion).

**Δ(6q<sub>39</sub>,7), ▼8p<sup>2</sup> [r.478\_631delins632-1\_632-2]**: Alternative ex6 donor site 39-nt upstream+ex7 skipping + alternative ex8 acceptor site 2-nt upstream.

**Δ8p<sub>4</sub> [r.632\_635del]**: alternative ex8 acceptor site 4-nt downstream (ex8 4-nt deletion).

**Δ(6q<sub>39</sub>,7,8p<sub>4</sub>) [r.478\_635del]**: Alternative ex6 donor site 39-nt upstream+ex7 skipping+alternative ex8 acceptor site 4-nt downstream.

**▼8q<sup>4</sup> [r.681\_682ins681+1\_681+4]**: Alternative ex8 donor site 4-nt downstream (ex8 4-nt insertion).

**Δ9 [r.682\_793del]**: Exon 9 skipping.

Other minor uncharacterized transcripts (21) by size: 609-nt, 612-nt, 651-nt, 675-nt, 715-nt, 742-nt, 744-nt, 757-nt, 806-nt, 817-nt, 834-nt, 854-nt, 856-nt, 857-nt, 860-nt, 868-nt, 872-nt, 930-nt, 963-nt, 971-nt, 973-nt.

---



# RESULTADOS

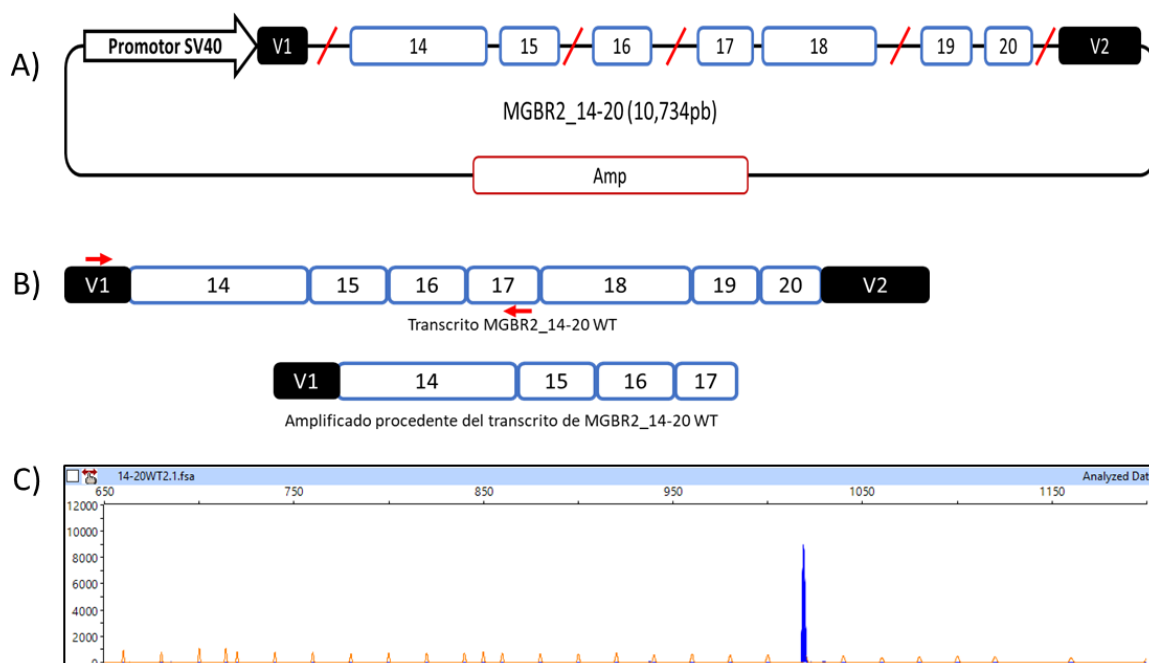


## RESULTADOS

### ANÁLISIS DE LAS VARIANTES DE *SPLICING* DE LOS EXONES 14 Y 15 DE *BRCA2*.

#### A. MINIGEN MGBR2\_14-20 Y ESTUDIO DE LAS SECUENCIAS REGULADORAS EN LOS EXONES 14 Y 15.

El minigen MGBR2 14-20 se construyó como se ha descrito previamente (Fraile-Bethencourt et al., 2017). Este minigen ha sido validado en estudios anteriores de los exones 16, 17 y 18 de *BRCA2* (Fraile-Bethencourt et al., 2017, 2018; Montalban et al., 2018). El minigen MGBR2\_14-20 wt (10,7 Kb) produce un transcrito estable de estructura V1-*BRCA2*(EX14-20)-V2, con un tamaño de 1806 nt. En este trabajo, el análisis de las isoformas resultantes de los ensayos en MCF-7 se realizó amplificando el cDNA procedente del minigen con un *primer* en V1 (pMAD\_607FW; Patente P201231427, CSIC) en *forward* y un *primer* en el exón 17 (RTBR2\_ex17RV2; 5'-GGCTTAGGCATCTATTAGCA-3') en *reverse*, obteniendo un único transcrito estable (Figura 8 C).



**Figura 8.** Minigen MGBR2\_14-20. A) Representación esquemática del MGBR2\_14-20. Los exones de *BRCA2* aparecen representados en cajas con el borde azul. Las líneas rojas indican los intrones acortados. B) Representación del transcrito generado después de la reacción de *splicing* y el

amplificado obtenido tras la PCR utilizando los *primers* indicados con flechas rojas. C) Resultado de la electroforesis capilar tras el ensayo funcional del MGBR2\_14-20 en células MCF-7. El transcrito obtenido aparece representado en un pico azul de 1019 nt. Los picos naranjas pertenecen al estándar de tamaño LIZ1200. Los ejes "x" e "y" indican el tamaño en pb y las unidades de fluorescencia relativa, respectivamente.

---

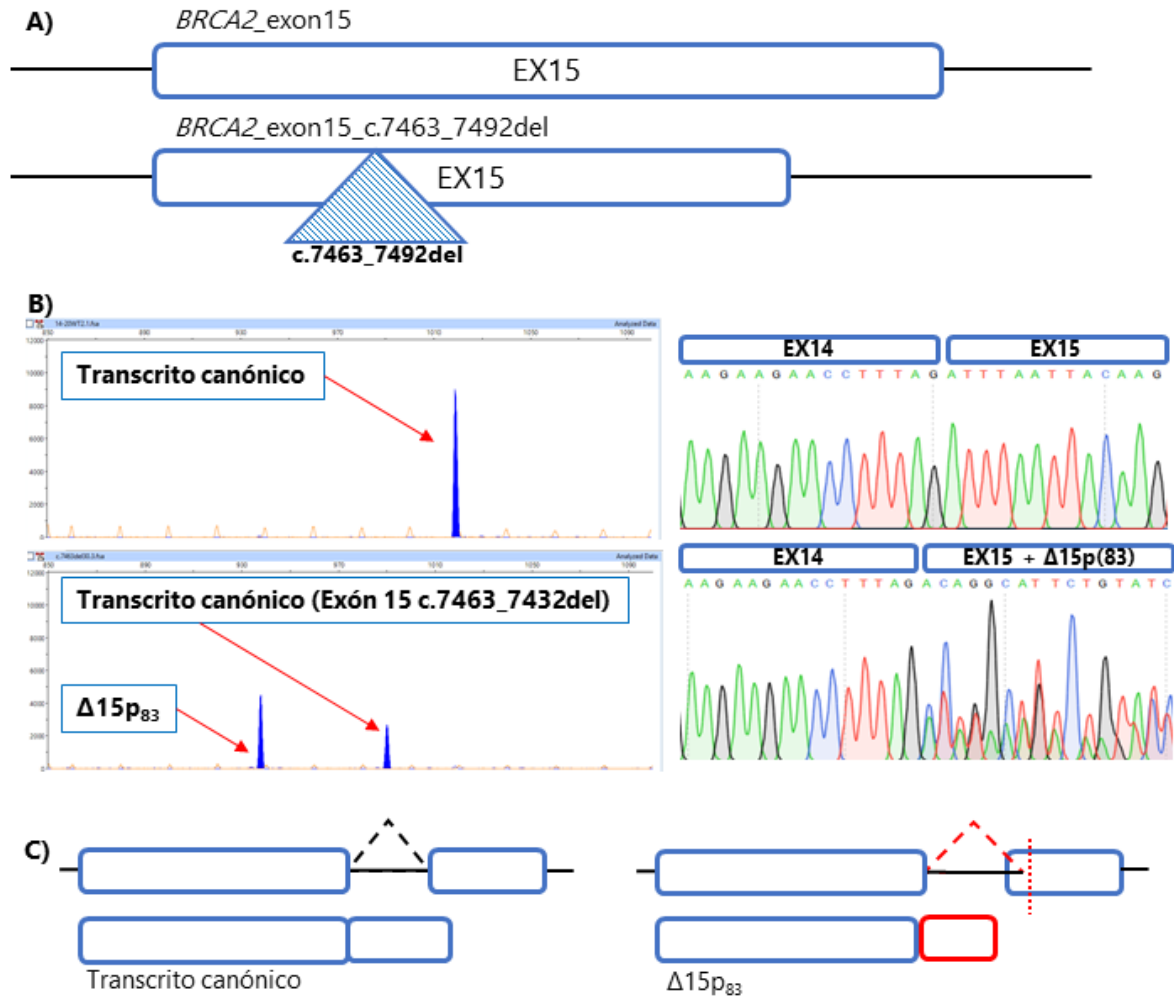
Con el objetivo de mapear los ESEs, se realizaron ocho microdeleciones solapantes mediante mutagénesis dirigida utilizando *QuikChange Lightning Kit* (Agilent, Santa Clara, CA, Estados Unidos) y los oligonucleótidos indicados en la Tabla Anexa 1. Las microdeleciones de 30 pb se realizaron en los extremos 5' y 3' de los exones 14 y 15, respetando los nucleótidos conservados de las secuencias consenso de *splicing* (los primeros 2 nt y los últimos 3 nt de cada exón), cubriendo así los primeros y últimos 55 nt nucleótidos de cada exón, donde se sitúan la mayoría de secuencias reguladoras de *splicing* (De Conti et al., 2012).

Una vez obtenidos, los vectores fueron transfectados en células MCF-7, a las que, después de 48 horas, se les trató con 300 µg/ml de cicloheximida durante 4 horas para inhibir el NMD, y se les extrajo el RNA utilizando *Genematrix Universal RNA Purification Kit* (EURx, Gdansk, Polonia). Se realizó RT-PCR con *RevertAid H Minus First Strand cDNA Synthesis Kit* (Life Technologies, Carlsbad, CA, Estados Unidos), utilizando RTPSPL3-RV (5'-TGAGGAGTGAATTGGTCGAA-3'). El cDNA fue amplificado con *Platinum Taq DNA polymerase* (Life Technologies, Carlsbad, CA, Estados Unidos) mediante el siguiente protocolo: 2 min a 94°C seguido por 35 ciclos que consistían en 30s a 94°C, 30s a 62°C y 1min a 72°C; y finalmente una extensión de 5 min a 72°C. Se utilizaron los *primers* pMAD\_607FW y RTBR2\_ex17RV2 (mencionados anteriormente). Los transcritos fueron analizados mediante gel de agarosa, secuenciación SANGER y electroforesis capilar (Macrogen, Madrid, España).

Los estudios funcionales mostraron que sólo una (c.7463\_7492del) de las ocho microdeleciones ensayadas provocaba modificaciones de los patrones de *splicing* esperados. La microdelección c.7463\_7492del en el exón 15 alteraba el reconocimiento del

aceptor de dicho exón. Ésta inducía un 62% de transcrito aberrante ( $\Delta 15p_{83}$ ) generado por el uso de un sitio aceptor alternativo 83 nt *downstream* del comienzo del exón 15 (Figura 9). Los datos muestran la presencia de un aceptor críptico en el exón 15, además sugieren que este exón contiene elementos reguladores en *cis* en la región c.7463\_7492 que se encuentran favoreciendo el reconocimiento del sitio aceptor canónico.

Con el objetivo de conocer los factores de *splicing* implicados en el reconocimiento del sitio aceptor del exón 15 de *BRCA2*, se realizó un estudio bioinformático sobre la secuencia c.7463\_7492 (5'-GAGATATACAGGATATGCGAATTAAGAAGA-3') utilizando HSF 3.1. Los resultados obtenidos mostraban secuencias de unión para las siguientes proteínas: SRSF1 (SF2), SRSF5 (SRp40), SRSF6 (SRp55), SRF7 (9G8) y Tra2 $\beta$ . Para conocer la implicación de estos factores en la regulación del *splicing* se llevaron ensayos de pérdida de función a través de RNA de interferencia (siRNA). Mediante esta técnica fueron inhibidos: SRSF1 (SF2), SRSF2 (SC35), SRSF3 (SRp20), SRSF5 (SRp40), SRSF7 (9G8), SRSF9 (SRp30c) y Tra2 $\beta$ . Las secuencias de los siRNA (Sigma-Aldrich, St. Louis, MO, Estados Unidos) aparecen en Tabla Anexa 2. El silenciamiento se confirmó a través de qPCR usando SG qPCR Master Mix (Eurx, Gdansk, Polonia), siguiendo las indicaciones del fabricante. Los oligonucleótidos utilizados para la amplificación aparecen en la Tabla Anexa 2. Los resultados fueron negativos, ninguno de los factores de *splicing* mencionados anteriormente parecía ser esencial para el reconocimiento del aceptor del exón 15 (Figura Suplementaria 2). Nuestra hipótesis para explicar este resultado se basa en la cooperación y sinergia de dichos factores para cumplir su función. Además, debido a la ubicuidad y el solapamiento de las funciones de estas proteínas, es fácil pensar que cuando se produce la inactivación de una de ellas, otras podrían sustituirla.



**Figura 9.** Ensayo funcional de la microdeleción c.7463\_7492. A) Esquema explicativo de la microdeleción c.7463\_7492 del exón 15 de *BRCA2*. B) Resultados de electroforesis capilar (izquierda) y secuenciación SANGER (derecha) después del ensayo funcional en MCF-7. La amplificación se realizó utilizando *Platinum Taq DNA polymerase* (Life Technologies) con pMAD\_607FW marcado con FAM para electroforesis capilar o sin marcar para secuenciación, y RTBR2\_Ex17RV2. Los transcritos marcados con FAM aparecen representados en picos azules, mientras que en naranja se muestra el estándar de tamaño (LIZ1200). Los resultados se analizaron con PeakScanner v1.0. C) Diagrama de cajas representativo de los transcritos encontrados. Las reacciones de *splicing* están representadas como líneas diagonales discontinuas.

## B. RECOPIACIÓN DE VARIANTES Y ANÁLISIS BIOINFORMÁTICO.

En total, 294 variantes situadas desde el final del intrón 13 hasta el comienzo del intrón 15 de *BRCA2*, fueron recopiladas de las principales bases de datos (ClinVar, UMD y BIC) (Tabla Anexa 3). Éstas fueron ensayadas *in silico* con los programas de predicción de sitios de *splicing* MaxEnt y NNSplice. Además, para conocer las alteraciones que podrían causar en las secuencias consenso de unión a elementos reguladores, las variantes se analizaron con las siguientes matrices incluidas en HSF: ESE-Finder, ESE-matrices y hnRNPA matrix. El punto de ramificación también fue analizado con HSF. Todos los análisis se hicieron con los valores predeterminados por el *software*. Tras el estudio bioinformático, se seleccionaron 53 variantes (18%) candidatas de *splicing* (Tabla 1). Los criterios de selección fueron: i) Creación o eliminación de sitios de *splicing*; ii) Eliminación del punto de ramificación; iii) Alteración del tracto de polipirimidinas; iv) Eliminación de *enhancers*; o v) Creación de silenciadores. Algunas de las variantes seleccionadas tenían un efecto mixto, cumpliendo varios de los criterios de selección. De acuerdo con la clasificación clínica reportada en las bases de datos, se seleccionaron ocho variantes benignas o probablemente benignas, 30 VUS y 15 variantes patogénicas o probablemente patogénicas. Atendiendo a su efecto en la proteína eran 19 variantes intrónicas, 18 *missense*, cinco *nonsense*, ocho sinónimas y tres *frameshift*.

Según el estudio bioinformático, 12 de las 53 variantes alteraban los sitios naturales de *splicing*, una alteraba el punto de ramificación, otra el tracto de polipirimidinas, tres alteraban ISE o ISS, 23 modificaban ESE o ESS y 11 creaban nuevos sitios de *splicing* (Tabla 1). Excepcionalmente se seleccionaron las variantes c.7008-5T>C (IVS13-5T>C) y c.7435+5T>C (IVS14+5T>C) porque, a pesar de que los resultados bioinformáticos fueron negativos (Tabla 1), las variantes se encuentran en posiciones conservadas de las secuencias consenso de los sitios aceptor y donador. Además, nuestro grupo y otros, han asociado previamente la posición +5 al desarrollo de enfermedades (Farber-Katz et al., 2018; Montalban et al., 2018; Sharma et al., 2014).

**Tabla 1. Estudio bioinformático de las variantes candidatas de alterar el *splicing* de los exones 14 y 15 de *BRCA2*. Anotación de las variantes según los criterios de HGVS.**

| Variantes<br>(HGVS)   | Sitio de splicing |                     | Punto de<br>ramificación | ESEs  |                   | ESS                     |
|-----------------------|-------------------|---------------------|--------------------------|---|-------------------|-------------------------|
|                       | <i>NNSplice</i>   | <i>MaxEnt</i>       |                          | <i>ESE Finder</i>   | <i>ESE-motif</i>  | hnRNP-A1                |
| Exón 14 (c.7008_7435) |                   |                     |                          |   |                   |                         |
| c.7008-5T>C           | 3'SS 0.56→0.53    | 3'SS 0.56→0.53      |                          |   |                   |                         |
| c.7008-3C>G           | 3'SS 0.56→<0.4    | 3'SS<br>10.37→1.64  |                          |   |                   |                         |
| c.7008-2A>T           | 3'SS 0.56→<0.4    | 3'SS 10.37→2.0      |                          |   |                   |                         |
| c.7008-1G>A           | 3'SS 0.56→<0.4    | 3'SS<br>10.37→1.62  |                          |   |                   |                         |
| c.7009A>G             |                   |                     |                          | [+] 2 SF2/ASF (IgM-BRCA1) (77.36) (70.85) [+] SF2/ASF (74.26) | [-] 9G8 (59.53)   | [+] 67.62               |
| c.7010C>T             |                   |                     |                          | [+] SRp55 (77.77)   | [-] 9G8 (59.53)   |                         |
| c.7024C>T             |                   |                     |                          | [-] SRp55 (80.13) [-] SRp40 (83.23)                           |                   | [+] 71.67               |
| c.7030A>G             | 5'SS <0.4→0.71    | 5'SS -<br>1.55→6.62 |                          | [+] SF2/ASF (78.33)   | [+] 9G8 (63.69)   | [+] 76.90               |
| c.7037A>G             |                   |                     |                          | [-] SC35 (81.13)  | [-] 9G8 (60.07)   | [+] 72.38               |
| c.7157C>A             |                   |                     |                          | [-] SC35 (79.66) [-] SRp40 (85.27)                            |                   | [+] 69.76               |
| c.7170T>G             |                   |                     |                          | [+] SRp40 (79.40) [+] SF2/ASF (73.79)                         |                   | [+]2 (67.62)<br>(76.43) |
| c.7177A>G             |                   |                     |                          |   | [-] Tra2B (83.21) | [+] (69.52)             |

|             |                |                     |  |   |                                   |             |
|-------------|----------------|---------------------|--|---|-----------------------------------|-------------|
| c.7180A>T   | 3'SS <0.4→0.49 |                     |  | [+] SRp40 (78.98)   |                                   | [-] (67.14) |
| c.7182A>G   | 3'SS <0.4→0.45 |                     |  |   |                                   |             |
| c.7203A>G   | 3'SS <0.4→0.46 |                     |  | [-] SF2/ASF (IgM-BRCA1) (82.15) [-] SF2/ASF (83.58)                             |                                   | [+] (72.86) |
| c.7261C>G   |                |                     |  | [-] SC35 (83.10)  |                                   | [+] (72.38) |
| c.7266T>A   | 5'SS <0.4→0.76 | 5'SS 2.7→8.34       |  |   |                                   |             |
| c.7294A>G   |                |                     |  |   | [-] Tra2B (83.21) [-] 9G8 (61.07) | [+] (65.48) |
| c.7296A>G   |                |                     |  | [-] SF2/ASF (IgM-BRCA1) (81.15) [-] SF2/ASF (80.90)                             | [-] 2 9G8 (61.07) (66.78)         | [+] (74.76) |
| c.7330G>T   |                |                     |  | [-] SF2/ASF (IgM-BRCA1) (77.62) [-] SF2/ASF (77.34)                             |                                   | [+] (68.33) |
| c.7339A>G   |                |                     |  |   |                                   | [+] (71.43) |
| c.7397C>T   |                |                     |  | [-] 2 SRp55 (74.95) (74.95) [-] SF2/ASF (IgM-BRCA1) (76.00) [-] SF2/ASF (76.06) |                                   | [+] (70.24) |
| c.7418G>A   | 5'SS <0.4→0.58 |                     |  | [+] SF2/ASF (IgM-BRCA1) (70.69)   | [+] 9G8 (61.95)                   |             |
| c.7428A>G   |                |                     |  |   |                                   | [+] (71.43) |
| c.7435+1G>C |                | 5'SS 5.64→-<br>2.62 |  |   |                                   |             |
| c.7435+3A>G |                | 5'SS 5.64→-<br>1.06 |  |   |                                   |             |
| c.7435+5T>C |                | 5'SS 5.64→5.56      |  |   |                                   |             |
| c.7435+6G>A |                |                     |  | [-] SC35 (75.35)  |                                   |             |

|                              |                 |                 |            |   |                          |                       |
|------------------------------|-----------------|-----------------|------------|---|--------------------------|-----------------------|
| c.7435+7T>G                  |                 |                 |            | [-] SC35 (75.35)  |                          | [-] (70.48)           |
| c.7435+10G>A                 |                 |                 |            | [+] SRp55 (79.88)   | [-] 9G8 (74.03)          | [-] (70.48)           |
| <b>Exón 15 (c.7436_7617)</b> |                 |                 |            |   |                          |                       |
| c.7436-22C>T                 |                 |                 | 79.34→55.5 | [+]SF2/ASF (75.19)  |                          | [+](73.81)            |
| c.7436-14T>G                 | 3'SS 0.9 → 0.81 | 3'SS 5.1 → 2.0  |            |   |                          | [+] (69.05)           |
| c.7436-4A>G                  | 3'SS 0.9→<0.4   | 3'SS 5.1 → 4.5  |            |   |                          |                       |
| c.7436-4A>T                  | 3'SS 0.9→<0.4   | 3'SS 5.1 → 4.7  |            |   |                          |                       |
| c.7436-2A>T                  | 3'SS 0.9→<0.4   | 3'SS 5.1 → -3.2 |            |   |                          |                       |
| c.7436-1G>A                  | 3'SS 0.9→<0.4   | 3'SS 5.1 → -3.6 |            |   |                          |                       |
| c.7447A>G                    | 3'SS <0.4→0.73  |                 |            | [+] SF2/ASF (IgM-BRCA1) (70.54) [+] SF2/ASF (77.23)                                   | [-] 9G8 (61.28)          | [+] 2 (66.43; (71.19) |
| c.7466A>G                    | 5'SS <0.4→0.54  | 5'SS 1.53→6.6   |            | [+]SRp40 (87.54); [+] SC35 (79.23)  | [+] 2 9G8 (63.69; 62.28) | [+] (76.90)           |
| c.7467T>C                    |                 |                 |            | [-] SF2/ASF (IgM-BRCA1) (77.77) [-] SF2/ASF (77.98)                                   | [+] 9G8 (65.37)          | [+] (67.14)           |
| c.7471C>T                    |                 |                 |            | [-]SRp40 (91.86) [-]SF2/ASF (IgM-BRCA1) (72.31) [-] SF2/ASF (79.29) [+] SRp55 (75.97) |                          |                       |
| c.7471delC                   |                 |                 |            | [-] SRp40 (91.86) [-] SF2/ASF (IgM-BRCA1) (72.31) [-] SF2/ASF (79.27)                 |                          |                       |
| c.7472A>G                    |                 |                 |            | [-] SRp40 (91.86)   | [-] 9G8 (62.75)          | [-] (75.48)           |
| c.7474_7475del               |                 |                 |            | [-] SRp40 (91.86) [-] SF2/ASF (IgM-BRCA1) (72.31) [-] SF2/ASF (79.27)                 | [-] 9G8 (62.75)          | [-] (75.48)           |

|                |                |                    |  |   |                   |                         |
|----------------|----------------|--------------------|--|---|-------------------|-------------------------|
| c.7492A>G      |                |                    |  |   | [+] 9G8 (66.18)   | [+] 2 (71.43;<br>76.43) |
| c.7501C>T      | 5'SS <0.4→0.96 | 5'SS<br>2.44→10.19 |  |   | [-] 9G8 (62.28)   | [+] (65.72)             |
| c.7544C>T      |                |                    |  | [-] SRp55 (76.23)   |                   |                         |
| c.7598C>G      |                | 5'SS -4.0→4.27     |  | [-] SC35 (78.61) [-] SRp40 (84.97) [-] SF2/ASF<br>(IgM-BRCA1) (71.54) |                   |                         |
| c.7601C>T      |                |                    |  | [-] SRp40 (84.97) [-] SF2/ASF (IgM-BRCA1)<br>(71.54)                  |                   |                         |
| c.7611_7615del | 3'SS <0.4→0.86 |                    |  | [+]SC35 (78.80) [+] SF2/ASF (IgM-BRCA1)<br>(76.46)                    | [-] Tra2B (83.21) | [+] (66.67)             |
| c.7617G>A      | 5'SS 0.99→0.8  |                    |  | [-] SRp40 (80.18) [-] SF2/ASF (79.79)                                 |                   | [-] (73.10)             |
| c.7617+1G>A    | 5'SS 0.99→<0.4 | 5'SS 9.8→1.62      |  | [-]SRp40 (80.18) [-]SF2/ASF (79.79)                                   |                   | [-] (73.10)             |
| c.7617+1G>T    | 5'SS 0.99→<0.4 | 5'SS 9.8→1.62      |  | [-]SRp40 (80.18) [-]SF2/ASF (79.79)                                   |                   | [-] (73.10)             |
| c.7617+2T>G    | 5'SS 0.99→<0.4 | 5'SS 9.8→2.15      |  |   | [+] 9G8 (61.28)   | [-] (66.19)             |

### C. ENSAYO FUNCIONAL DE VARIANTES DE *SPLICING* EN LOS EXONES 14 Y 15 CON EL MINIGEN MGBR2\_14-20

Cincuenta y dos variantes candidatas de alterar el *splicing* de los exones 14 y 15 fueron introducidas en el minigen MGBR2\_14-20 mediante mutagénesis dirigida. Los oligonucleótidos utilizados aparecen en la Tabla Anexa 4. A pesar de que fueron seleccionadas 53, no fuimos capaces de introducir la variante c.7598C>G del exón 15. Los resultados del ensayo funcional en células MCF-7 mostraron que 12 de las variantes (23%) alteraban el *splicing* (Tabla 2). Entre las variantes espliceogénicas, encontramos nueve intrónicas (c.7008-3C>G, c.7008-2A>T, c.7008-1G>A, c.7435+1G>C, c.7436-2A>T, c.7436-1G>A, c.7617+1G>A, c.7617+1G>T y c.7617+2T>G), dos *missense* (c.7177A>G y c.7447A>G) y una *nonsense* (c.7501C>T). Funcionalmente, las variantes intrónicas alteraban los sitios naturales de *splicing*, las *missense* provocaban el uso de nuevos sitios de *splicing*, y la *nonsense*, a pesar de haber sido seleccionada por la creación de un nuevo sitio de *splicing*, los resultados sugieren que actuaba alterando SREs.

#### Variantes que alteran el reconocimiento del sitio aceptor

Según los resultados obtenidos, el reconocimiento del aceptor del exón 14 se ve afectado por las variantes c.7008-3C>G, c.7008-2A>T y c.7008-1G>A. Sin embargo, la c.7008-5T>C dio lugar a un transcrito canónico. Curiosamente, mientras la c.7008-3C>G resultaba principalmente en la pérdida del exón 14 ( $\Delta 14$ ), las variantes c.7008-2A>T y c.7008-1G>A producían mayoritariamente la isoforma aberrante  $\Delta 14p_{10}$  (Tabla 2). Este transcrito, generado debido al reconocimiento de un aceptor críptico situado 10 nt *downstream* del aceptor natural, portaría un PTC 27 codones después de la alteración (p.Thr2337Asnfs\*27). Cabe destacar que el aceptor críptico encontrado no fue reconocido por ninguna de las dos matrices utilizadas en el estudio bioinformático (NNSplice y MaxEnt).

Nuestros resultados muestran también la utilización de un aceptor críptico localizado a 246 nt del aceptor natural del exón 14. Este aceptor alternativo fue reconocido en el minigen portador de la variante c.7008-3C>G, generando un 13,4% de transcrito  $\Delta 14p_{246}$  (Tabla 2). En este caso, el 3' ss críptico era reconocido por MaxEnt, pero no por NNSplice. El

transcrito  $\Delta 14p_{246}$  provocaría una pérdida *in-frame* de 82 aa desde la posición p.2337 hasta p.2418 (p.Thr2337\_Arg2418del).

En el intrón 14 encontramos dos variantes que podrían estar alterando indirectamente el reconocimiento del aceptor del exón 15. Por un lado, la variante c.7436-22C>T parecía alterar la secuencia consenso del punto de ramificación (Tabla 1). Por otro, c.7436-14T>G supondría una alteración en el tracto de polipirimidinas. Sin embargo, los ensayos funcionales revelaron resultados negativos para ambas variantes, pues generaban un transcrito sin alteraciones de *splicing*. Las variantes ensayadas que afectaban directamente a la secuencia consenso del aceptor del exón 15 fueron: c.7436-4A>G, c.7436-4A>T, c.7436-2A>T y c.7436-1G>A. Los datos obtenidos sugieren que, mientras la posición -4 no parece ser relevante en para el reconocimiento del aceptor, las variantes c.7436-2A>T y c.7436-1G>A producen mayoritariamente un transcrito con un aceptor alternativo 13 nt *downstream* del natural ( $\Delta 15p_{13}$ ) (Tabla 2). La pérdida de 13 nt al comienzo del exón 15 provocaría un PTC en la proteína (p.Asp2479Valfs\*41). Al igual que ocurría anteriormente con el aceptor críptico en el exón 14 ( $\Delta 14p_{10}$ ), éste tampoco es reconocido por las matrices utilizadas en el estudio *in silico*. Los resultados de c.7008-1G>A, mostraron también el transcrito  $\Delta 15p_{83}$  de forma minoritaria (3,7%) (Tabla 2). Aquí, es importante recalcar la diferencia entre los resultados obtenidos tras el estudio *in silico* y los ensayos funcionales de las variantes, subrayando la necesidad en la mejora de las herramientas bioinformáticas de predicción de *splicing*.

En resumen, encontramos 5 variantes (c.7008-3C>G, c.7008-2A>T, c.7008-1G>A, c.7436-2A>T y c.7436-1G>A) que alteran el reconocimiento de los 3' ss de los exones 14 y 15 de *BRCA2* (Tabla 2). Las variantes mostraron la ausencia total de transcrito canónico, a excepción de c.7008-3C>G, que generó un 7,2% de la isoforma wt. Además, nuestros resultados muestran el uso de aceptores crípticos en los exones 14 y 15 que pasan desapercibidos por los programas de predicción, pero son utilizados cuando los naturales se encuentran alterados.

**Tabla 2. Cuantificación de los transcritos generados por las 12 variantes que alteran el *splicing* de los exones 14 y 15 de *BRCA2*.**

| Variantes                    | TRANSCRITOS |                    |                    |                     |                     |            |                   |
|------------------------------|-------------|--------------------|--------------------|---------------------|---------------------|------------|-------------------|
|                              | Canónico    | ▼14q <sup>5</sup>  | Δ14p <sub>10</sub> | Δ14p <sub>246</sub> | Δ14q <sub>256</sub> | Δ14        | Otros             |
| <b>Exon 14 (c.7008_7435)</b> |             |                    |                    |                     |                     |            |                   |
| c.7008-3C>G                  | 7,2%±0,1%   |                    |                    | 13,4%±0,1%          |                     | 79,3%±0,2% |                   |
| c.7008-2A>T                  |             |                    | 88,1%±0,4%         |                     |                     | 11,9%±0,4% |                   |
| c.7008-1G>A                  |             |                    | 86,5%±0,7%         |                     |                     | 13,5%±0,7% |                   |
| c.7177A>G                    | 95,2%±0,1%  |                    |                    |                     | 1,5%±0,02%          |            | 3,3%±0,1% (942nt) |
| c.7435+1G>C                  |             | 100%               |                    |                     |                     |            |                   |
| Variantes                    | Canónico    | Δ15p <sub>12</sub> | Δ15p <sub>13</sub> | Δ15p <sub>83</sub>  | Δ15                 |            |                   |
| <b>Exon 15 (c.7436_7617)</b> |             |                    |                    |                     |                     |            |                   |
| c.7436-2A>T                  |             |                    | 100%               |                     |                     |            |                   |
| c.7436-1G>A                  |             |                    | 96,3%±0,1%         | 3,7%±0,1%           |                     |            |                   |
| c.7447A>G                    | 89,9%±0,3%  | 10,1%±0,3%         |                    |                     |                     |            |                   |
| c.7501C>T                    | 84,0%±0,1%  |                    |                    | 16,0%±0,1%          |                     |            |                   |
| c.7617+1G>A                  |             |                    |                    |                     | 100%                |            |                   |
| c.7617+1G>T                  |             |                    |                    |                     | 100%                |            |                   |
| c.7617+2T>G                  |             |                    |                    |                     | 100%                |            |                   |

#### Variantes que alteran el reconocimiento del sitio donador

Se ensayaron siete variantes que supuestamente alteraban el reconocimiento del sitio donador de los exones 14 o 15 de *BRCA2* (Tabla 1). Las variantes que afectaban al donador del exón 14 fueron: c.7435+1G>C, c.7435+3A>G, c.7435+5T>C y c.7435+6G>A; y las que afectaban al exón 15 fueron: c.7617+1G>A, c.7617+1G>T y c.7617+2T>G. Entre las del exón 14, solo c.7435+1G>C resultó tener un efecto en el *splicing*. Ésta generó un único transcrito con una inserción de 5 nt debido al reconocimiento de un sitio críptico en el intrón 14 (▼14q<sup>5</sup>). Este donador críptico no había sido previamente recocido por los programas de predicción. ▼14q<sup>5</sup> es una isoforma aberrante que llevaría a la aparición de un PTC en la proteína (p.Asp2479Glyfs\*4).

En cuanto a las variantes que afectaban al donador del exón 15, los resultados de los ensayos funcionales mostraban que las tres (c.7617+1G>A, c.7617+1G>T y c.7617+2T>G) producían un transcrito único carente del exón 15 ( $\Delta 15$ ) (Tabla 2). La pérdida del exón 15 supondría una delección *frameshift* (p.Asp2479Alafs\*8).

En resumen, encontramos 4 variantes que alteran el reconocimiento de los 5' ss de los exones 14 y 15 de *BRCA2* (Tabla 2). Todas ellas alteraban gravemente el *splicing*, pues resultaron en la ausencia total de transcrito canónico. Además, nuestros resultados muestran la presencia de un donador críptico en el intrón 14.

#### Variantes que alteran secuencias reguladoras de *splicing* (SREs)

Se ensayaron en MCF-7 un total de 26 variantes que alteraban secuencias de unión a elementos reguladores de *splicing* (Tabla 1). Las variantes que afectaban al exón 14 fueron 17 (tres intrónicas y 14 exónicas). Los datos obtenidos mostraron que sólo una de ellas, c.7177A>G, alteraba ligeramente el *splicing* del exón 14. En conjunto, estos resultados negativos refuerzan los datos obtenidos mediante el ensayo de microdelecciones mencionado anteriormente, el cual sugiere que el exón 14 no se encuentra regulado por SREs.

A diferencia del exón 14, el exón 15 si mostró secuencias esenciales para el correcto procesamiento del mRNA. La delección de los nt comprendidos en la región c.7463\_7492 provocaba mayoritariamente el uso de un aceptor críptico (62%  $\Delta 15p_{83}$ ) (Figura 9), sugiriendo la presencia de ESEs activos que actúan en el reconocimiento del aceptor natural. Teniendo en cuenta estos resultados, ensayamos cinco variantes que alteraban ESEs en dicha región (c.7467T>C, c.7471C>T, c.7471delC, c.7472A>G y c.7474\_7475delGA). Curiosamente, la bioinformática mostró que la variante c.7467T>C, no solo alteraba ESEs, sino que también creaba un ESS (Tabla 1). Dentro de la misma zona, se ensayó también c.7492A>G, pues supuestamente generaba un sitio de unión a hnRNPA1. Por otro lado, se ensayaron c.7544C>T, c.7601C>T y c.7617G>A, porque, a pesar de estar fuera de la región de interés, los resultados bioinformáticos mostraban la eliminación de ESEs. Además, c.7617G>A corresponde al último nt del exón. Sin embargo, contrariamente a lo esperado, ninguna de

estas variantes alteró el *splicing* del exón 15, ni siquiera aquellas comprendidas en la región c.7463\_7492.

#### Variantes que crean nuevos sitios de *splicing*

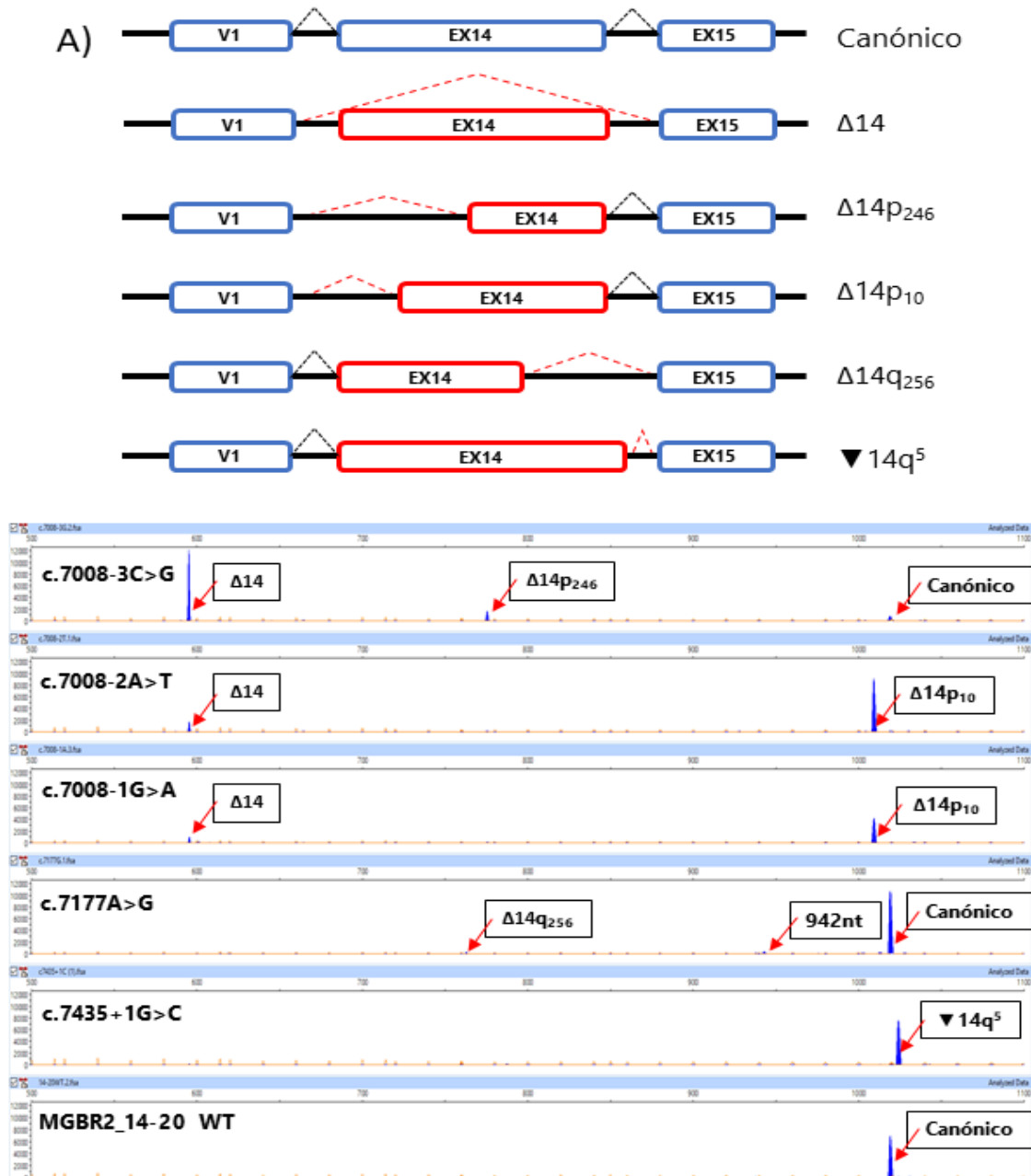
Se seleccionaron y ensayaron 10 variantes que, según los resultados bioinformáticos, creaban nuevos sitios de *splicing* (Tabla 1). Seis en el exón 14 (c.7030A>G, c.7180A>T, c.7182A>G, c.7203A>G, c.7266T>A y c.7418G>A) y cuatro (c.7447A>G, c.7466A>G, c.7501C>T y c.7611\_7615delTAAAC) en el exón 15. Los resultados mostraron que las dos variantes del exón 15 c.7447A>G y c.7501C>T alteraban ligeramente el *splicing*. La primera, c.7447A>G, aunque produjo mayoritariamente el transcrito canónico, encontramos un 10% de la isoforma  $\Delta 15p_{12}$ , generada por el uso del nuevo sitio de *splicing* (Tabla 2). Por otro lado, la variante c.7501C>T alteraba el reconocimiento del aceptor natural del exón 15, pues provocó un 16% de  $\Delta 15p_{83}$  (Tabla 2). Curiosamente, a pesar de que los resultados bioinformáticos indicaban la creación de un nuevo donador fuerte (NNSplice: <0,4→0,96 y MaxEnt: 2,44→10,19) (Tabla 1), éste no fue observado en ninguno de los ensayos. El conjunto de los datos del ensayo funcional y los resultados *in silico*, donde se aprecia la pérdida de la secuencia de unión a 9G8 (SRSF7), nos hace pensar que la variante c.7501C>T podría estar alterando ESEs del exón 15, en lugar de crear un nuevo 5'ss.

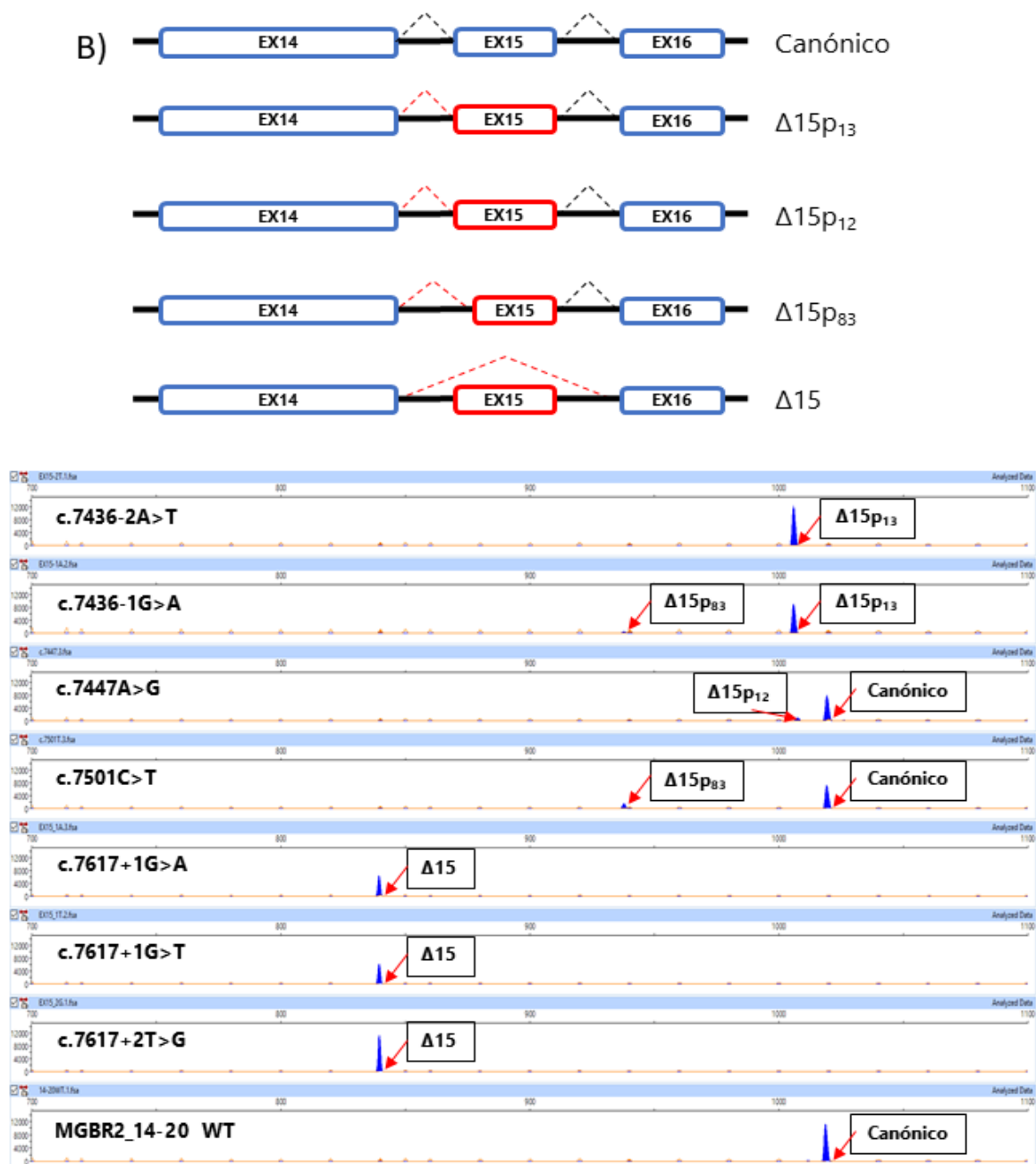
#### **D. ANÁLISIS DE LOS TRANSCRITOS**

A través de los ensayos con el minigen MGBR2\_14-20, caracterizamos nueve isoformas generadas debido al *splicing* aberrante producido por variaciones en la secuencia de DNA de los exones 14 y 15 (Figura 10).

Por un lado, las variantes del exón 14 produjeron un total de seis transcritos diferentes, aunque sólo pudimos caracterizar cinco de ellos:  $\Delta 14$ ,  $\Delta 14p_{10}$ ,  $\Delta 14p_{246}$ ,  $\Delta 14q_{259}$ ,  $\nabla 14q^5$ . Una sexta isoforma de 942nt fue identificada mediante electroforesis capilar. Mientras los transcritos  $\Delta 14$ ,  $\Delta 14p_{10}$  y  $\nabla 14q^5$  se comprobaron mediante secuenciación capilar, los demás fueron deducidos por el tamaño de los fragmentos obtenidos debido a la pequeña cantidad de transcrito producido. Todas las isoformas resultantes de variantes del exón 14, excepto  $\Delta 14p_{246}$ , son delecciones *frameshift* que llevarían a la aparición de un PTC en la proteína.

Por otro lado, las variaciones de la secuencia que afectan al *splicing* del exón 15 de *BRCA2* resultaron en cuatro transcritos diferentes:  $\Delta 15p_{12}$ ,  $\Delta 15p_{13}$ ,  $\Delta 15p_{83}$  y  $\Delta 15$ . Sólo la isoforma  $\Delta 15p_{12}$ , una delección *in-frame*, fue deducida por su tamaño a través de los resultados de electroforesis capilar. Las demás fueron comprobadas por secuenciación.





**Figura 10.** Isoformas encontradas tras el ensayo funcional con el MGBR2\_14-20 de las variantes de *splicing* de los exones 14 y 15 de *BRCA2*. Arriba aparecen representados los eventos de *splicing* encontrados. Las cajas con bordes azules representan exones naturales, los bordes rojos indican exones aberrantes. Las líneas discontinuas muestran las reacciones de *splicing*. Abajo aparecen los electroferogramas mostrando el producto de PCR marcado con FAM (azul). El estándar de tamaño utilizado fue LIZ1200. A) Variantes del exón 14; B) Variantes del exón 15.

## **E. INTERPRETACIÓN CLÍNICA DE LAS VARIANTES DE *SPLICING* EN LOS EXONES 14 Y 15 DE *BRCA2*.**

En total, 12 de las variantes ensayadas produjeron patrones de *splicing* aberrantes. Sin embargo, debido a las diferencias en la conservación de las secuencias, las cantidades relativas de transcrito y el número de nucleótidos delecionados o insertados, cada una de ellas debe ser analizada individualmente.

Los ensayos funcionales indicaron que las variantes intrónicas c.7008-2A>T, c.7008-1G>A, c.7435+1G>C, c.7436-2A>T, c.7436-1G>A, c.7617+1G>A, c.7617+1G>T y c.7617+2T>G resultaban en la ausencia total de transcrito canónico. Además, todas las isoformas producidas por estas variantes eran aberrantes ya que presentaban deleciones o inserciones *frameshift* que generan codones de parada prematuros. Siguiendo los criterios propuestos por ACMG, estas variantes podrían ser consideradas como variantes patogénicas (Tabla 3).

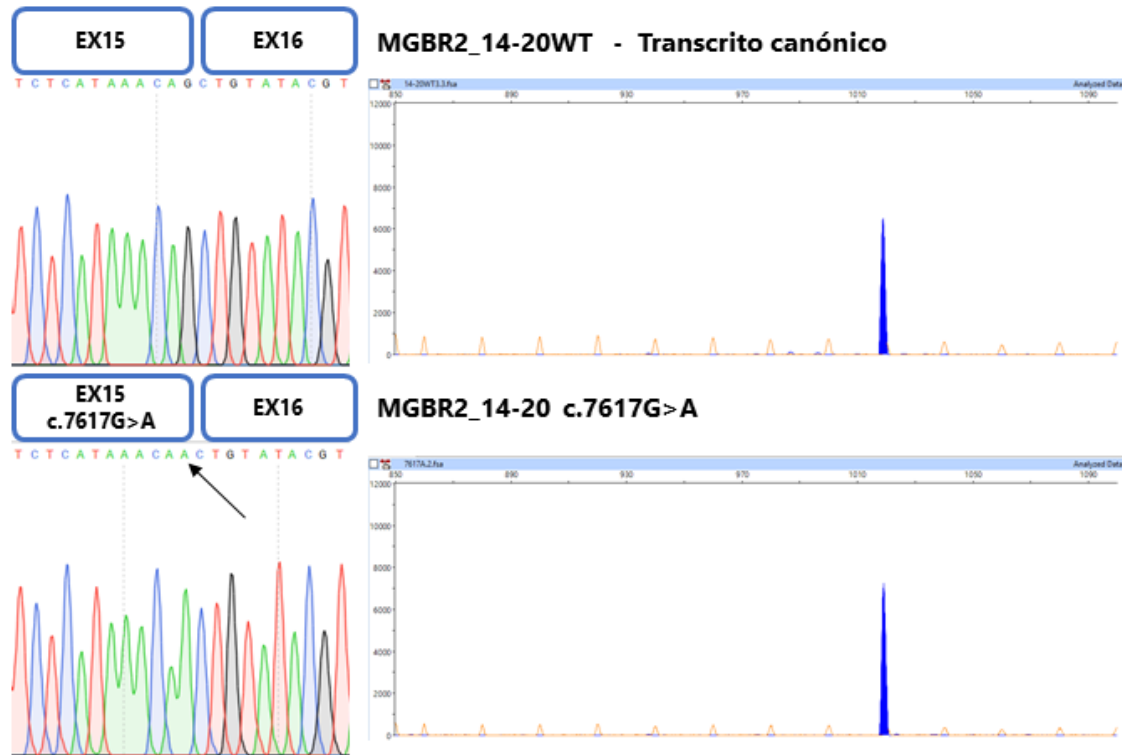
Algunas variantes tienen efectos parciales en el *splicing*. Por un lado, encontramos la variante c.7008-3C>G que produce mayoritariamente la isoforma aberrante  $\Delta 14$  (~80%), pero también un 7% de transcrito wt y un 13% de  $\Delta 14p_{246}$  (Tabla 2). El transcrito  $\Delta 14p_{246}$  supone una deleción *in-frame* (p.Thr2337\_Arg2418del) que elimina parte del sitio de unión a FANCD2 (p.Thr2350\_Val2545). Según los criterios de ACMG, esta variante, clasificada como VUS en ClinVar, debería considerarse como probablemente patogénica. Por otro lado, la variante *nonsense* c.7501C>T (p.Gln2501Ter) generó un 84% de transcrito canónico y un 16% de  $\Delta 15p_{83}$  (Tabla 2). En este caso, puesto que el total de las isoformas portarían PTC, su clasificación clínica no varía con respecto a lo previamente establecido, se considera una variante patogénica (Tabla 3). Sin embargo, debido a la alteración que produce en el *splicing*, la clasificación funcional debería ser mixta: *nonsense* y espliceógena.

Otras variantes como c.7177A>G (p.Met2393Val) y c.7447A>G (p.Ser2483Gly) también dieron lugar mayoritariamente a la isoforma canónica (95% y 90%, respectivamente) (Tabla 2). Sin embargo, en estos casos, el transcrito canónico porta variaciones en la secuencia de mRNA que darán lugar a cambios de aminoácidos en la proteína. Con el objetivo de conocer cómo afectan estos cambios, se realizó un estudio *in silico* utilizando la herramienta

bioinformática PolyPhen (<http://genetics.bwh.harvard.edu/pph2/>). Según los resultados, tanto p.Met2393Val como p.Ser2483Gly eran cambios benignos que supuestamente no afectarían a la funcionalidad de la proteína. Por tanto, en este caso, siguiendo los criterios de ACMG y a falta de resultados funcionales que confirmen el impacto de las variantes, consideramos que ambas deberían permanecer clasificadas como VUS.

Cabe destacar que la variante c.7617G>A (último nucleótido del exón 15), aparece recogida en la base de datos UMD como causante de enfermedad debido a su supuesto efecto en *splicing*: pérdida del exón 15. Sin embargo, los ensayos funcionales en MCF-7 utilizando el MGBR2\_14-20 mostraron un 100% de transcrito canónico (Figura 11). Atendiendo a los datos obtenidos con el minigen, y dado que se trata de una variante sinónima (p.Gln2539=), proponemos una revisión en la clasificación clínica de c.7617G>A.

En conjunto, nuestros datos aportan información para re-clasificar tres variantes VUS (c.7008-3C>G, c.7008-1G>A y c.7436-1G>A) a patogénicas o probablemente patogénicas, según los criterios de ACMG. Además, apoyan la actual clasificación clínica de seis variantes de *splicing* consideradas como patogénicas (c.7008-2A>T, c.7435+1G>C, c.7436-2A>T, c.7617+1G>A, c.7617+1G>T y c.7617+2T>G).



**Figura 11.** Ensayo funcional de c.7671G>A, el último nucleótido del exón 15. A la izquierda, las secuencias de los transcritos generados por MGBR2\_14-20 wt y MGBR2\_14-20 c.7671G>A en células MCF-7. La flecha indica el cambio de nucleótido. A la derecha, los electroferogramas de los transcritos procedentes de los minigenes wt y c.7671G>A. Los picos azules muestran el producto de PCR marcado con FAM, los picos naranjas pertenecen al estándar de tamaño (LIZ1200).

**Tabla 3. Análisis y clasificación de las 12 variantes que alteran el *splicing* de los exones 14 y 15 de *BRCA2* en el minigen MGBR2 14-20.**

| HGVS           | Tipo de variante                  | Isoforma                            | Efecto en la proteína | Clasif. previa | Criterios ACMG                 | Clasf. ACMG         |
|----------------|-----------------------------------|-------------------------------------|-----------------------|----------------|--------------------------------|---------------------|
| <b>EXÓN 14</b> |                                   |                                     |                       |                |                                |                     |
| c.7008-3C>G    | Intrónica                         | $\Delta 14$<br>(r.7008_7435del)     | Thr2337Phefs<br>*17   | VUS            | PS3, PM2,<br>PP3, PP5          | Prob.<br>Patogénica |
| c.7008-2A>T    | Intrónica                         | $\Delta 14p10$<br>(r.7008_7017del)  | Thr2337Asnfs<br>*27   | Patogénica     | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| c.7008-1G>A    | Intrónica                         | $\Delta 14p10$<br>(r.7008_7017del)  | Thr2337Asnfs<br>*27   | VUS            | PVS1, PS3,<br>PM2, PP3         | Patogénica          |
| c.7177A>G      | <i>Missense</i><br>(p.Met2393Val) | Canónico                            |                       | VUS            | PM2, PP3                       | VUS                 |
| c.7435+1G>C    | Intrónica                         | $\nabla 14q5$<br>(r.7435_7435+5ins) | Asp2479Glyfs<br>*4    | Patogénica     | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| <b>EXÓN 15</b> |                                   |                                     |                       |                |                                |                     |
| c.7436-2A>T    | Intrónica                         | $\Delta 15p13$<br>(r.7436_7448del)  | Asp2479Valfs<br>*41   | Patogénica     | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| c.7436-1G>A    | Intrónica                         | $\Delta 15p13$<br>(r.7436_7448del)  | Asp2479Valfs<br>*41   | VUS            | PVS1, PS3,<br>PM2, PP3         | Patogénica          |
| c.7447A>G      | <i>Missense</i><br>(p.Ser2483Gly) | Canónico                            |                       | VUS            | PM2, PP3                       | VUS                 |
| c.7501C>T      | <i>Nonsense</i><br>(p.Gln2501Ter) | Canónico                            |                       | Patogénica     | PSV1, PM2,<br>PP3, PP5         | Patogénica          |
| c.7617+1G>A    | Intrónica                         | $\Delta 15$<br>(r.7436_7617del)     | Asp2479Alafs<br>*8    | Patogénica     | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| c.7617+1G>T    | Intrónica                         | $\Delta 15$<br>(r.7436_7617del)     | Asp2479Alafs<br>*8    | Patogénica     | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| c.7617+2T>G    | Intrónica                         | $\Delta 15$<br>(r.7436_7617del)     | Asp2479Alafs<br>*8    | Patogénica     | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |

# **DISCUSIÓN**



## DISCUSIÓN

---

La clasificación clínica de variantes en genes asociados a cáncer de mama es actualmente uno de los mayores retos a los que se enfrenta la Medicina Genómica. El desarrollo de nuevas tecnologías de secuenciación, acompañado del abaratamiento de los costes, ha llevado a la aparición de miles de variantes de significado clínico desconocido. Comúnmente, el impacto funcional de las variantes se considera patogénico cuando son *nonsense* o *frameshift*, es decir, generan un PTC. También, se consideran patogénicas aquellas *missense* que afectan a aminoácidos esenciales en la proteína. Esta visión, que atiende exclusivamente al proceso de traducción, pasa por alto variantes que podrían ser deletéreas debido a la alteración de otros mecanismos esenciales en el procesamiento del DNA, como por ejemplo la transcripción o el *splicing*. En este contexto, los ensayos funcionales proporcionan información esencial para la clasificación clínica de variantes, que revierte en el consejo genético de pacientes y/o familias portadoras.

Durante el transcurso esta tesis doctoral se han diseñado y construido vectores estables para el estudio de la transcripción y del *splicing* de *BRCA2*, se han estudiado las regiones reguladoras a través de microdeleciones y ensayos de pérdida de función génica mediante RNA pequeños de interferencia (siRNA); se han aplicado técnicas como la electroforesis capilar para analizar los transcritos generados; y se han ensayado más de 200 variantes distribuidas a lo largo del gen *BRCA2*. Además, se han llevado a cabo estudios de interacción RNA-proteína, conocidos como Pull-Down, en el International Centre for Genetic Engineering and Biotechnology en Trieste (Italia), con el fin de conocer los factores que participan en la regulación del *splicing* de *BRCA2*. Los resultados de esta tesis doctoral han permitido la publicación de 5 artículos científicos: 4 presentados en el apartado ARTÍCULOS; y uno en colaboración con el Hospital Clínico San Carlos de Madrid y el Hospital Vall d'Hebron de Barcelona (Montalban et al., 2018); además de otro que se encuentra bajo revisión. Por otro lado, se han establecido colaboraciones estables con otros grupos de investigación.

## RELACIÓN ENTRE EL PROMOTOR DE *BRCA2* Y EL DESARROLLO DEL CMOH

Desde los años 90 es conocido que la inactivación de los genes *BRCA* está asociada a un incremento en el riesgo de sufrir cáncer de mama y/o ovario (Miki et al., 1994; Wooster et al., 1995). Históricamente, las variantes germinales han sido detectadas en las partes codificantes del gen, ya que son las usualmente rastreadas en los laboratorios de diagnóstico genético. Sin embargo, las regiones no codificantes juegan un papel fundamental. De hecho, aproximadamente el 90% de las variantes patogénicas conocidas se encuentran en regiones no codificantes del DNA (Hindorff et al., 2009). A lo largo de esta tesis doctoral se ha estudiado la región promotora del gen *BRCA2*, elemento esencial para el inicio de la transcripción y la expresión génica.

### A. REGIONES REGULADORAS DEL PROMOTOR DE *BRCA2*

El mapeo funcional de secuencias de DNA a través de microdeleciones es una estrategia que permite conocer las regiones que juegan un papel esencial en un determinado proceso. Para estudiar el promotor de *BRCA2*, se clonó la región 5' UTR en el vector reportero pGL4.10 con el gen de la luciferasa (Artículo 2; Figura 1 A). Se realizaron 13 microdeleciones de 50 pb desde la posición -509 hasta +81, tomando como punto de referencia el TSS. Las microdeleciones se diseñaron de manera que cubrieran el segmento donde se han encontrado mayor número de mutaciones transcripcionales, entre -500 y +50 desde el TSS (de Vooght et al., 2009). Los resultados mostraron que el promotor de *BRCA2* está compuesto por todo tipo de secuencias reguladoras. Así, durante el desarrollo de este trabajo, hemos encontrados tanto elementos activadores, como elementos represores o secuencias neutras que no afectaban a la transcripción (Artículo 2; Figura 2). Por ejemplo, la eliminación de las secuencias comprendidas en las microdeleciones 5, 10 y 11 no parecía afectar a la tasa transcripcional del gen, pues la diferencia con el resultado procedente del promotor wt no fue significativa. Por otro lado, las microdeleciones 1, 2, 6, 8, 9 y 13 mostraron una sobreexpresión de luciferasa, lo cual indica que a estas regiones podrían unirse elementos represores de la actividad transcripcional, ya que al ser delecionadas provocaban el efecto observado. El resto del promotor, las zonas comprendidas en las

microdeleciones 3, 4, 7 y 12, apuntan a ser secuencias de unión a factores de transcripción activadores, pues su eliminación reducía significativamente la expresión génica.

Las microdeleciones 1 y 3 fueron las que tuvieron un mayor efecto. La microdelección 3 causó la mayor reducción de la expresión de luciferasa, con una caída de más del 80% con respecto al wt (Artículo 2; Figura 2), lo cual podría representar el resultado esperado de un alelo nulo. Este dato puede explicarse fácilmente si tenemos en cuenta que la región comprendida en la microdelección 3 va desde la posición -59 hasta la -10 desde el TSS, donde se localiza el núcleo del promotor de *BRCA2* (de -66 a +129, TSS) (Davis et al., 1999). A esta zona se unen los GTFs y la RNAPol II, por lo que es lógico que, al ser eliminada, se pierda la expresión del gen. Por otro lado, la microdelección 1 comprendía desde la posición +32 hasta la +81 desde el TSS, formando parte del primer exón de *BRCA2*. Curiosamente, a pesar de pertenecer también al núcleo del promotor, la eliminación de esta secuencia resultó en un aumento la expresión de más de 356% con respecto al wt (Artículo 2; Figura 2), lo que sugiere la presencia de fuertes elementos represores en la zona. En 1999, cuando Davis y colaboradores caracterizaron por primera vez el promotor de *BRCA2*, lo hicieron a través de microdeleciones. Así, determinaron que la región entre -66 y +129 desde el TSS era el fragmento con mayor actividad transcripcional con respecto a su tamaño. Sin embargo, encontraron también que, utilizando como promotor el fragmento comprendido entre las posiciones -17 y +129 desde el TSS, la expresión era prácticamente nula. Según nuestros resultados, la microdelección 3 (de -59 a -10, TSS) reduce la expresión a mínimos, pero tanto la microdelección 1 (de +32 a +81, TSS) como la 2 (de -14 a +36, TSS) aumentaban la expresión de luciferasa. La comparación de sus resultados con los nuestros lleva a sugerir una región concreta dentro del núcleo de promotor fuertemente regulada por factores activadores, la cual estaría comprendida entre la posición -66 y la -14 desde el TSS.

El mapeo funcional del promotor aporta una valiosa información tanto desde el punto de vista biológico como clínico. Por ejemplo, podríamos pensar en un posible rescate de la expresión génica en el caso de alelo nulo a través de la modificación de los elementos represores que se encuentran en las microdeleciones 1 y 2. Sin embargo, para corroborar esta hipótesis, son necesarios experimentos complementarios que confirmen los datos

obtenidos e identifiquen los factores específicos implicados en la regulación de la transcripción de *BRCA2*. Por otro lado, el conocimiento de las regiones reguladoras supone una ventaja a la hora del rastreo de mutaciones patogénicas en el promotor, pues éstas representan áreas importantes que pueden modificar significativamente la expresión del gen. En este trabajo hemos identificado la secuencia comprendida entre la microdelección 8 y la microdelección 1 (región -284\_+81, tomando como referencia el TSS) como una zona rica en elementos reguladores, pues en ella encontramos las mayores modificaciones de la tasa transcripcional. De esta forma, el estudio mediante microdeleciones aporta información esencial, mapeando las zonas donde se sitúan los elementos reguladores y podrían encontrarse variantes que alteren la unión de factores esenciales de la transcripción.

## **B. EFECTO DE LAS VARIANTES DE DNA EN EL PROMOTOR DE *BRCA2***

Durante este estudio, se ensayaron funcionalmente 15 variantes de DNA del promotor de *BRCA2* encontradas en pacientes y reportadas en las bases de datos (Artículo 2; Tabla 1 y Figura 1 B). Después de rastreo genético del promotor de *BRCA2* de 95 pacientes de cáncer de mama *BRCA* negativas, pertenecientes al programa de prevención de cáncer de Castilla y León, se seleccionaron tres variantes germinales: rs3092989\_A, rs563971900\_T y rs206118\_A. Ninguna de las tres modificó significativamente la expresión de luciferasa con respecto al promotor wt (Artículo 2; Figura 3 A). Este resultado sugiere que las variantes no afectan a la transcripción de *BRCA2* y por tanto no estarían implicadas en el desarrollo de cáncer de mama y ovario hereditario. La asociación de estas variantes con la enfermedad ha sido descartada también por otros autores. Por ejemplo, la variante rs3092989\_A no mostró asociación con el cáncer de mama en un estudio realizado en 491 pacientes en China (Liu et al., 2014). Así mismo, varios autores han descartado la variante rs206118\_A como causante de enfermedad (Baynes et al., 2007; Figueiredo et al., 2011; Mucaki et al., 2016). De hecho, esta variante aparece como benigna en dbSNP ([https://www.ncbi.nlm.nih.gov/snp/rs206118#clinical\\_significance](https://www.ncbi.nlm.nih.gov/snp/rs206118#clinical_significance)). Curiosamente, la variante rs563971900\_T ha sido analizada por Burke y colaboradores (c.-296C>T) posteriormente a la publicación del Artículo 2 de esta de tesis. Según sus ensayos de EMSA (*Electrophoretic Mobility Shift Assay*), esta variante se encuentra en una región que contiene

secuencias de específicas reconocidas por proteínas nucleares de MCF-7, y cuya unión se ve gravemente afectada a causa del cambio de nucleótido. Específicamente, encontraron que la variante rs563971900\_T alteraba la unión del factor de transcripción PAX5 al promotor de *BRCA2*. De hecho, previamente los mismos autores habían observado una leve disminución de la expresión de luciferasa a causa de esta variante en células MDA-MB-231 (dos Santos et al., 2018), fenómeno que se repetía en células MCF-7, pero no en células MDA-MB-468 (Burke et al., 2018). El bajo impacto transcripcional de este tipo de variantes hace pensar en una compleja interacción entre los factores de transcripción y los elementos en *cis* del promotor, así como en la posible redundancia de funciones en el proceso. Por otro lado, la diferencia de resultados obtenidos pone de manifiesto la variabilidad entre las distintas líneas celulares que son utilizadas para este tipo de experimentos. Finalmente, debido principalmente a su incidencia en la población control, los autores concluyen que la variante rs563971900\_T no está implicada en el desarrollo de la enfermedad y proponen clasificarla como “no patogénica” o “probablemente no patogénica”, lo cual está apoyado por nuestros datos de expresión de luciferasa (105%).

A partir de la base de datos Ensembl, se seleccionaron para el ensayo 12 variantes raras que mapeaban entre -500 y +75 desde el TSS de *BRCA2*: rs545887954\_T, rs541454524\_G, rs148060832\_A, rs36221751\_A, rs565737252\_A, rs533226610\_A, rs551887850\_G, rs11571571\_C, rs570548398\_T, rs556172218\_T, rs55880202\_T y rs567110692\_A (Artículo 2; Tabla 1). Entre ellas, siete aumentaban significativamente la expresión génica, tres la disminuían y dos no variaban con respecto al wt (Artículo 2; Figura 3 A). Al contrario que en las variantes encontradas en pacientes, los SNPs rs545887954\_T y rs556172218\_T, que no variaban la expresión, aparecen sin clasificar. Además, no hemos encontrado en la bibliografía ningún estudio que los mencione. Por otro lado, tres de los SNPs estudiados disminuían entre un 17 y un 27% la expresión de luciferasa con respecto al wt: rs551887850\_G, rs570548398\_T y rs55880202\_T (Artículo 2; Figura 3 A). Puesto que las tres variantes causan una reducción moderada de la expresión génica, no podrían ser considerados como alelos nulos. La variante rs55880202\_T (c.-175C>T) ha sido posteriormente analizada en células MDA-MB-468 y células MCF-7 y no se ha encontrado

un efecto significativo en la expresión de luciferasa (Burke et al., 2018). La clasificación de variantes que varían ligeramente la tasa transcripcional supone un reto en la medicina genómica, pues, aunque es cierto que la disminución de la expresión de *BRCA2* podría llevar al desarrollo de cáncer de mama y/o ovario, no se conoce el umbral mínimo de expresión de *BRCA2* a partir del cual se determina la existencia de haplo-insuficiencia. Estudios recientes en *BRCA1* sugieren que un 20-30% de expresión es suficiente para mantener el papel de supresor tumoral de esta proteína, y por tanto no confieren riesgo de desarrollo de cáncer de mama (de la Hoya et al., 2016). Por otro lado, estudios en otros genes de reparación de daño en el DNA (*MLH1* y *MSH2*), señalan que una reducción del 50% en la expresión del gen es suficiente para relacionarlo con la aparición de la enfermedad (Liu et al., 2016). La clasificación clínica de variantes del promotor en una tarea complicada debido al desconocimiento de las regiones promotoras y de los niveles de expresión necesarios para un correcto funcionamiento de la proteína, además éstos pueden variar en función del tejido, del momento del desarrollo, de las condiciones ambientales y de otros muchos factores. Por ello, se hace necesario el estudio en profundidad de estas regiones en los genes asociados a enfermedad.

Los otros siete SNPs estudiados aumentaban significativamente la expresión de luciferasa con respecto al wt: rs541454524\_G, rs148060832\_A, rs36221751\_A, rs565737252\_A, rs533226610\_A, rs11571571\_C y rs567110692\_A (Artículo 2; Figura 3 A). El aumento de la expresión, que variaba entre 113-163%, se asocia a la pérdida o modificación de elementos represores en el promotor o bien a la creación de elementos activadores. Dada la función supresora de tumores de *BRCA2*, su sobreexpresión probablemente tenga un efecto protector, disminuyendo el riesgo de padecer la enfermedad (Harper et al., 2015). Un ejemplo lo encontramos en la variante rs1799943\_A (c.-26G>A) de *BRCA2*, que aumenta hasta dos veces la expresión del gen en células HeLa y células MCF-7, y ejerce un papel protector frente al riesgo de padecer cáncer de mama (Gochhait et al., 2007). Asimismo, el alelo rs4942440\_A también se asocia a una disminución del riesgo de cáncer mama debido a su efecto en la sobreexpresión de *BRCA2* (Maia et al., 2012). Con estos antecedentes, podríamos suponer que las variantes encontradas tienen un efecto protector frente a la

enfermedad. Sin embargo, debido a que se trata de variantes raras, no existen estudios de asociación que apoyen esta teoría, por tanto, no pueden ser clasificadas como variantes protectoras. Además, por otro lado, la sobreexpresión de genes reparadores de daño en el DNA ha sido asociada con el desarrollo de metástasis en diferentes tumores, entre los que se encuentra el cáncer de mama (Kauffmann et al., 2008). También ha sido relacionada con resistencia terapéutica y un mal pronóstico clínico. Este fenómeno, aparentemente contradictorio, puede ser explicado desde el punto de vista de la progresión de la enfermedad. Así, para la aparición y crecimiento del tumor es necesaria una inestabilidad genómica que confiere a las células la capacidad de dividirse descontroladamente. Sin embargo, al comienzo de la metástasis, las células tumorales deben introducirse en el torrente sanguíneo mediante extravasación, viajar a través de él y colonizar el nuevo tejido. Estas funciones, todas ellas específicas y reguladas, requieren de resistencia y estabilidad genómica, que se encontraría potenciada cuando existe sobreexpresión de *BRCA2*. De hecho, hace 20 años se asoció la sobreexpresión de *BRCA2* con una mayor agresividad en tumores de mama esporádicos (Bièche et al., 1999). Además, se ha visto que la expresión de *RAD51*, un factor clave en la reparación de daño a DNA, está incrementada en muchos tumores, probablemente contribuyendo a la resistencia al tratamiento (Klein, 2008). Dado el doble efecto de la sobreexpresión de los genes reparadores, la clasificación de SNPs basada únicamente en ensayos funcionales sería inapropiada, pues el incremento con respecto al wt podría ser a la vez beneficioso y perjudicial. En estos casos, los estudios de asociación cobran vital importancia a la hora de decidir el papel que juegan las variantes en el desarrollo y evolución del tumor.

La actividad transcripcional está regulada a distintos niveles. Por un lado, una mayor actividad está relacionada con una condensación menor de la cromatina, lo que la hace más accesible a factores de transcripción, pero también más sensible a digestión por DNasa I. Estas regiones pueden encontrarse en la enciclopedia de elementos del DNA conocida como ENCODE (Consortium, 2012), son los llamados *clusters* de hipersensibilidad a DNasa I. De los SNPs que alteraban significativamente la expresión de luciferasa, nueve mapeaban en dos de estas regiones (Artículo 2; Figura 3 B). En el *cluster* 61, más alejado del TSS, desde -

710 hasta -296, se encuentra parte del promotor proximal y en él mapea la variante rs541454524\_G, que aumentaba la expresión génica. En el *cluster* 121 encontramos ocho SNPs reguladores, cinco aumentando la expresión génica (rs36221751\_A, rs565737252\_A, rs533226610\_A, rs11571571\_C, rs567110692\_A) y tres disminuyéndola (rs551887850\_G, rs570548398\_T y rs55880202\_T). El *cluster* 121 está situado desde la posición -185 hasta la +105, tomando como referencia el TSS. En él localiza el núcleo del promotor, donde se unen la mayoría de los GTFs y la RNAPol II. Esta región es crítica en la regulación transcripcional ya que cualquier cambio de nucleótido podría modificar secuencias de unión de factores esenciales.

En resumen, la mayoría de las variantes ensayadas (10/15) alteraban significativamente la tasa de transcripción, aunque estos cambios eran moderados, sobre todo en aquellas variantes que disminuían la expresión (17-21%). Entre las variantes, no encontramos ningún caso de alelo nulo que pudiese asociarse con la aparición de la enfermedad. Un gran número de los SNPs reguladores (8/10) mapeaban en el *cluster* 121, indicando la importancia de esta región en la transcripción de *BRCA2*. Los datos obtenidos muestran el promotor de *BRCA2* como un promotor sensible a cambios, señalando la necesidad de incorporar el rastreo genético del mismo en la rutina clínica. Por otro lado, son necesarios más estudios funcionales que aporten información sobre elementos reguladores, variantes y niveles mínimos de expresión génica, para ayudar a mejorar el diagnóstico molecular del CMOH.

## PROCESAMIENTO DEL MRNA DE *BRCA2* Y SU RELACIÓN CON EL CMOH

El procesamiento del mRNA o *splicing* de *BRCA2* es un proceso altamente complejo y, a pesar de los más de 20 años transcurridos desde el descubrimiento del gen y de su implicación en el desarrollo de varios tipos de tumores, poco se conoce sobre las secuencias implicadas en el reconocimiento de los exones. De hecho, fue en 2016 cuando se publicó por primera vez el patrón de isoformas fisiológicas de *splicing* alternativo de *BRCA2* (Fackenthal et al., 2016). Por otro lado, tras décadas de rastreo genético en pacientes con CMOH y otras patologías hereditarias, los genes *BRCA* se han convertido en los genes de predisposición a cáncer más secuenciados (Cline et al., 2018). Además, el incremento en el uso de la secuenciación masiva en centros de investigación, hospitales y empresas ha aumentado exponencialmente el número de variantes encontradas. La dificultad de conocer el efecto de una variante de DNA en la funcionalidad de la proteína implica que el incremento en el número de variantes encontradas vaya asociado a un incremento en el número de VUS. En este contexto, los estudios funcionales de *splicing* cobran vital importancia, pues cualquier cambio de nucleótido puede, teóricamente, afectar al reconocimiento de los exones. De hecho, se estima que hasta el 50% de las variantes exónicas podrían estar alterando el *splicing*, y que aproximadamente el 60% de las variantes asociadas al desarrollo de enfermedades son variantes espliceogénicas (Lopez-Bigas et al., 2005). Un ejemplo se puede observar en el trabajo de Sanz y colaboradores, realizado en nuestro laboratorio, donde las variantes c.439C>T (p.Gln147Ter) y c.9234C>T (p.Val3078=) de *BRCA2*, previamente clasificadas como variantes *nonsense* y sinónima, respectivamente, resultaron en la pérdida de los exones correspondientes debido a la alteración de SREs (Sanz et al., 2010), por lo que deberían ser consideradas como variantes espliceogénicas. Los resultados de esta tesis muestran secuencias de *BRCA2* esenciales para el reconocimiento de los exones, así como factores de *splicing* implicados en dicha regulación. Además, se ha realizado un estudio exhaustivo sobre el efecto espliceogénico de variantes candidatas encontradas en las principales bases de datos.

## **A. MINIGENES COMO HERRAMIENTAS PARA EL ESTUDIO DE VARIANTES DE *SPLICING*.**

Resultados previos de nuestro grupo habían demostrado la fiabilidad de los minigenes basados en pSAD (Patente P201231427-CSIC) para el estudio de variantes de *splicing* en *BRCA2*, pues el resultado de ocho de las variantes ensayadas en el minigen MGBR2\_19-27 coincidían con los obtenidos en muestras de pacientes (Acedo et al., 2015). En este trabajo se han validado los dos minigenes utilizados (MGBR2\_14-20 y MGBR2\_2-9) a través de un total de 27 variantes que reproducían un resultado igual o muy similar a lo previamente reportado en la bibliografía. Con el minigen MGBR2\_14-20, ensayamos 13 variantes de las que había datos publicados: c.7008-2A>T (Colombo et al., 2009, 2013), c.7617+1G>A (Gutiérrez-Enríquez et al., 2009), c.7617+2T>G (Vreeswijk et al., 2009), c.7618-1G>A (Whiley et al., 2011), c.7805G>C y c.8023A>G (Bonnet et al., 2008), c.7806-9T>G (Kwong et al., 2008), c.7975A>G (Houdayer et al., 2012), c.7976G>A (Hofmann et al., 2003; Wu et al., 2005), c.7976G>C y c.7988A>T (Farrugia et al., 2008), c.7992T>A (Théry et al., 2011), y c.8168A>G (Walker et al., 2010). De la misma manera, con el minigen MGBR2\_2-9 encontramos 14 variantes que coincidían con lo reportado: c.67+3A>G (Parsons et al., 2015), c.316+3del (Muller et al., 2011), c.316+5G>C (Bonnet et al., 2008), c.426-12\_426-8del (Zhang et al., 2009), c.467A>G (Sanz et al., 2010), c.516+1G>T (Whiley et al., 2011), c.517G>T, c.572A>G y c.617C>G (Gaildrat et al., 2012), c.627C>A y c.627C>T (Di Giacomo et al., 2013), c.631G>A (Pensabene et al., 2009), c.631+3A>G y c.681+4A>G (Houdayer et al., 2012). Por otro lado, detectamos una discordancia en la variante c.68-7T>A. Mientras los análisis con el minigen mostraban aproximadamente un 3% de  $\Delta 3$  (Artículo 4; Tabla Suplementaria S6), en un estudio reciente en células linfoblastoides encontraron este transcrito incrementado hasta en un 20%, aunque finalmente concluyeron que no se trataba de una variante patogénica (Colombo et al., 2018). Estas variaciones en los resultados pueden deberse a diversos factores como las diferencias técnicas entre laboratorios (aislamiento de RNA, retrotranscriptasas, DNA polimerasas, ciclos de amplificación, técnicas de análisis, *primers* y longitud del amplificado, etc.) o el tipo celular utilizado, ya que el *splicing* es dependiente de tejido. Se pueden encontrar variaciones en los transcritos debido también a efectos de contexto genómico, ya que la presencia de SNPs, el momento del desarrollo o los factores

ambientales podrían modular el proceso (Lee and Rio, 2015). En un trabajo realizado en colaboración con el Hospital Clínico San Carlos de Madrid y el Hospital Vall d'Hebron de Barcelona, analizamos en distintas familias y con el minigen MGBR2\_14-20, la variante c.7976+5G>T de *BRCA2*, observando ciertas diferencias en las isoformas encontradas, aunque todos los resultados coincidían en  $\Delta 17$  como transcrito principal (Montalban et al., 2018). Otras evidencias que demuestran la validez de los minigenes como herramientas de análisis de variantes de *splicing* las podemos encontrar en los resultados obtenidos con el MGBR2\_2-9 wt, que mostró patrones fisiológicos del *splicing* alternativo de *BRCA2* previamente caracterizados ( $\Delta 3$ ,  $\Delta 5$ ,  $\Delta 5,6$  y  $\Delta 6q7$ ). Las isoformas  $\Delta 3$ ,  $\Delta 5$  y  $\Delta 6q,7$  fueron observadas en muestras de tumor de mama y en células MCF-7, mientras que  $\Delta 5,6$  había sido detectada en sangre y líneas linfoblastoides (Fackenthal et al., 2016). Así mismo, el transcrito  $\Delta 5,6$  había sido previamente observado en células COS7 como resultado de un minigen basado en pSPL3 que contenía los exones 5 y 6 de *BRCA2* (Hansen et al., 2010). Otro ejemplo de que los minigenes son una herramienta válida para el estudio de variantes de *splicing*, lo encontramos en un reciente estudio de 30 variantes en los genes *MLH1*, *MSH2*, *MSH6* y *PMS2*, asociados al síndrome de Lynch. En él se ha demostrado que los resultados obtenidos de pacientes son prácticamente idénticos a los observados en minigenes (van der Klift et al., 2015). En resumen, el conjunto de los resultados obtenidos demuestra la validez de los minigenes para el estudio de variantes de *splicing*.

Los minigenes MGBR2\_2-9 y MGBR2\_14-20 son herramientas simples, precisas y robustas con las que poder ensayar variantes de *splicing* comprendidas en dichas regiones de *BRCA2*. Los minigenes son especialmente útiles cuando no se dispone de RNA de los pacientes portadores de la variante a estudiar. Además, presentan ciertas ventajas como el análisis específico del alelo portador de la variante, sin influencias del alelo wt; la identificación y cuantificación precisa de todos los transcritos generados, a través de la inhibición del NMD; la utilización de distintos tipos celulares y distintas condiciones; y la posibilidad de ensayar variantes artificiales para estudiar la regulación de los exones.

## **B. SECUENCIAS REGULADORAS Y FACTORES DE *SPLICING* IMPLICADOS EN EL RECONOCIMIENTO DE LOS EXONES DE *BRCA2*.**

Para un correcto procesamiento del mRNA, a parte de las secuencias básicas de *splicing*, las cuales se encuentran altamente conservadas, son precisas otras secuencias que actúan como elementos reguladores (SREs). Estos elementos, *enhancer* o silenciadores, son necesarios para el reconocimiento de ciertos exones. El mapeo de dichas secuencias aporta una información de gran utilidad a la hora de analizar y clasificar variantes en un determinado gen, pues la alteración de estos elementos en *cis* conllevaría un *splicing* aberrante. En esta tesis se han mapeado funcionalmente 13 de los 26 exones codificantes de *BRCA2*. Se han realizado 38 microdeleciones exónicas abarcando hasta 55 nt desde los límites exón/intrón, dado que está aceptado que la mayor densidad de *enhancers* se encuentra entre 10 y 20 nt desde los sitios de *splicing* (De Conti et al., 2012; Fairbrother et al., 2004). En total, nueve de los 13 exones analizados resultaron estar controlados por ESEs: los exones del 3 al 8 (Artículo 4; Figura 2 A); el exón 15 (Resultados; Figura 9); y los exones 17 y 18 (Artículo 1; Figura 2 A y B). Todos estos exones contienen secuencias esenciales a las que se unen factores de *splicing* necesarios para su reconocimiento. Éstas constituirían *hotspots* para mutaciones espliceogénicas. De hecho, en el minigen MGBR2\_2-9, 12 de las variantes espliceogénicas estaban localizadas en estas regiones: c. c.97G>A, c.100G>A (exón 3), c.352C>T, c.368\_372del5, c.378A>G, c.379G>T, c.383A>G (exón 4), c.440A>T, c.441A>G, c.451G>A, c.467A>G y c.470\_474del (exón 5) (Artículo 4; Tabla 1); y en el minigen MGBR2\_14-20, siete de las variantes que alteraban el *splicing* del exón 18 mapeaban en la región reguladora: c.7985C>G, c.7988A>T, c.7992T>A, c.8007A>G, c.8009C>A, c.8009C>T y c.8009C>G (Artículo 1; Tabla 1).

Los minigenes pueden utilizarse para estudiar secuencias reguladoras en *cis*, así como los factores de *splicing* que se unen a ellas (Cooper, 2005). Su plasticidad permite incorporar modificaciones para testar hipótesis y elucidar los elementos esenciales en una secuencia determinada. En esta tesis doctoral se han resuelto detalladamente las secuencias que controlan el reconocimiento del exón 18 de *BRCA2* (Artículo 1; Figura 2) (Anexos; Figura Suplementaria 1). Ahora sabemos que la región c.7979\_8013 es esencial para el *splicing* de

dicho exón y que cualquier cambio de nucleótido podría alterar el *splicing* de *BRCA2* e incrementar el riesgo de CMOH. Por otro lado, nuestros datos sugieren un efecto sinérgico de los factores de *splicing* implicados en este proceso, ya que en los ensayos de pérdida de función mediante siRNA, observamos la alteración del *splicing* al inhibir SFSR2 (SC35), pero los resultados no fueron tan impactantes como los obtenidos con las microdeleciones (Anexos; Figura Suplementaria 1). Asimismo, nuestros estudios de regulación del exón 3 de *BRCA2*, sugieren que SRSF3 (SRp20) está participando activamente en el reconocimiento del exón (Artículo 4; Figura 2 B). Aunque, de la misma manera que en el exón 18, al inhibir este factor, el reconocimiento del exón se veía solo ligeramente afectado (17,5%  $\Delta 3$ ), mientras que la delección de la región reguladora (c.95\_124del) provocaba aproximadamente un 46% de  $\Delta 3$  (Artículo 2; Tabla Suplementaria S4). Probablemente debido a la degeneración de las secuencias de unión, estos factores sean reemplazables y/o actúen de manera sinérgica. Siguiendo esta hipótesis, intentamos silenciar varios factores a la vez, sin embargo, la viabilidad de las células se veía demasiado comprometida. Curiosamente, dentro la región reguladora del exón 3, mapeaban las variantes c.97G>A y c.100G>A, cuyo resultado fue mayoritariamente  $\Delta 3$  (71% y 57%, respectivamente) (Artículo 4; Tabla Suplementaria S6). Aunque las posiciones c.97 y c.100 no pertenecen a la supuesta secuencia de unión de SRp20, se encuentran muy próximas a ella, formando según Mfold (<http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form>) una estructura secundaria en forma de lazo que se ve alterada por los cambios de nucleótido c.97G>A y c.100G>A (Artículo 4; Figura Suplementaria S3). Estos datos reflejan la complejidad de la regulación del proceso de *splicing*, en la que no sólo se deben tener en cuenta las secuencias básicas, sino que entran en juego otros elementos como los factores de *splicing* o la estructura secundaria del RNA.

Si comparamos las regiones de *BRCA2* analizadas durante esta tesis con los dos minigenes, del exón 2 al 9 (MGBR2\_2-9) y del exón 14 al 18 (MGBR2\_14-20), observamos que la primera tiene una regulación mucho más compleja. Desde el exón 3 hasta el exón 8 aparecen secuencias esenciales para el reconocimiento de exones (Artículo 4; Figura 2 A). Además, en esta región, las secuencias reguladoras en *cis* de un determinado exón parecen

tener efecto en el *splicing* del exón vecino. Por ejemplo, la delección de la región reguladora del exón 4 (c.344\_397) mostró transcritos minoritarios como  $\Delta 4,5$  y  $\Delta 4,5,6$ ; la delección de c.494\_514 en el exón 6, produjo en torno a un 24% de  $\Delta 5,6$ ; y la delección de la región reguladora del exón 8 (c.634\_678) provocó aproximadamente un 20% de la isoforma  $\Delta 6q,7,8$  (Artículo 4; Tabla Suplementaria S4). Estos resultados muestran como los elementos reguladores contenidos en un exón pueden afectar al reconocimiento de otros exones, es decir, observamos una regulación inter-exónica, la cual pone de manifiesto una vez más la importancia del contexto genómico en el reconocimiento de los exones (Buratti et al., 2006). A su vez, nuestros resultados sugieren que el exón 5 de *BRCA2* está altamente regulado, pues las regiones reguladoras de los exones vecinos también afectan a su reconocimiento. Además, las dos microdelecciones de 25 nt en exón 5 resultaron principalmente en la isoforma  $\Delta 5$  (Artículo 4; Tabla Suplementaria S4). De hecho, en el trabajo con el MGBR2\_2-9, hemos hallado tres variantes que provocan aberraciones en el *splicing* por la alteración de ESEs en este exón: c.440A>T, c.441A>G. y c.451G>A (Artículo 4; Tabla 1); demostrando que el exón 5 constituiría un *hotspot* para mutaciones espliceogénicas en *BRCA2*.

Las secuencias básicas de *splicing*, aunque están altamente conservadas (especialmente los sitios donador y aceptor), pueden variar. Un ejemplo lo encontramos en el donador del exón 17 de *BRCA2*, un donador atípico "GC". Este tipo de donadores se encuentra aproximadamente en el 1% de los intrones humanos y suelen estar asociados a *splicing* alternativo (Kralovicova et al., 2011). Como se ha mencionado previamente, con el minigen MGBR2\_14-20, estudiamos la variante c.7976+5G>T, localizada en la posición +5 del donador del exón 17 e identificada en tres familias de CMOH no relacionadas en el Hospital de Universitario de Vall d'Hebron (HUVH), en el Hospital Clínico San Carlos (HCSC) y en el Instituto de Biología y Genética Molecular (IBGM) (Montalban et al., 2018). Aunque se encontraron algunas diferencias, los resultados obtenidos por los tres laboratorios coincidían en que la variante c.7976+5G>T provocaba principalmente la pérdida del exón 17. La frecuencia del nucleótido guanina en la posición +5 del intrón es del 80% (Zhang, 1998), además, en concreto en el exón 17 la conservación de la secuencia del donador cobra mayor importancia para compensar el donador "GC" (Kralovicova et al., 2011). Existen otros

intrones de *BRCA2* donde la alteración de la posición +5 produce alteraciones en el *splicing*. Durante esta tesis se ha ensayado c.316+5G>C en el donador del exón 3, obteniendo  $\Delta 3$  como transcrito principal (~90%) (Artículo 4; Tabla Suplementaria S6). Otra variante de este tipo, analizada por nuestro laboratorio, fue c.8754+5G>A en el donador del exón 21 de *BRCA2*, que provocaba el uso de un 5' ss críptico a 46 nt del sitio natural de *splicing* (Acedo et al., 2015). En otros genes, existen también casos de variantes espliceogénicas en esta posición, como c.2657+5G>A en el gen *CFTR* (Highsmith et al., 1997) o c.116+5G>C en el gen *MLH1* (Arnold et al., 2009). Sin embargo, la variante c.7435+5T>C en el intrón 14 de *BRCA2*, seleccionada para su estudio debido a estas evidencias, no causó alteración alguna en el *splicing* del correspondiente exón, posiblemente debido a la baja conservación del nucleótido timina en la posición +5 (5%). En conjunto, todos estos resultados sugieren un efecto espliceogénico cuando la guanina de la posición +5 en el intrón es reemplazada por otro nucleótido y ponen de manifiesto la importancia de la conservación de las secuencias de *splicing* más allá de las posiciones típicamente observadas (-1, -2, +1 y +2).

Otro elemento que debe tenerse en cuenta en el proceso de regulación del *splicing* es el uso de sitios crípticos. Éstos son secuencias que coinciden con las secuencias conservadas del aceptor o del donador, pero en condiciones normales pasan desapercibidas para el espliceosoma. Los sitios crípticos se utilizan en determinados contextos, por ejemplo, cuando una variante altera el sitio de *splicing* natural o cuando ésta elimina una secuencia de unión a ESS que está inhibiendo su reconocimiento (Cartegni et al., 2002). A lo largo de *BRCA2* nos encontramos con varios sitios crípticos no identificados por los programas bioinformáticos, pero que eran utilizados como resultado de algunas de las variantes, generando transcritos aberrantes. Por ejemplo, la variante c.67G>A, última posición del exón 2, activaba el uso de un donador "GC" críptico 4 nt *upstream* del natural, provocando  $\Delta 2q_4$  (Artículo 4; Figura 4 B). Por el contrario, la alteración de los sitios +1, +2 y +3 del intrón 2, resultaron en la pérdida del exón 2 (Artículo 4; Tabla 1).

Otros ejemplos los observamos en las variantes que afectan al aceptor del exón 17. Se ensayaron cinco variantes de este tipo: c.7806-9T>G, c.7806-2A>G, c.7806-1G>A, c.7806-1G>T y c.7806-1\_7806-2dup; cada una resultó en un patrón diferente de transcritos (Artículo

1; Figura 3). Mientras las variantes c.7806-1G>A y c.7806-1\_7806-2dup generaban aceptores *de novo* reconocidos por el espliceosoma, el resto de las variantes provocaban el uso de aceptores crípticos:  $\Delta 17p_{69}$ ,  $\Delta 17p_{20}$  y  $\nabla 17p^8$ . En conjunto, los resultados demuestran la dificultad de predecir el efecto de una variante sobre el *splicing* de un exón. Contrariamente a lo obtenido con los ensayos mediante microdeleciones, a través de los cuales observamos que el extremo 5' de *BRCA2* presentaba una regulación más compleja, en los exones 14-18 se observó que era más prevalente el uso de sitios crípticos. En todos los exones, desde el 14 hasta el 18, encontramos variantes que activaban el reconocimiento de estos sitios. El conocimiento de los sitios crípticos es importante a la hora de predecir el efecto de una determinada variante. Por ejemplo, un cambio en los sitios 3' o 5' de *splicing*, normalmente resultará en la pérdida de dicho exón, pero si se activa el uso de un sitio críptico, el resultado generado será un transcrito con una inserción o delección, que podría ser *in-frame*, y por tanto de interpretación conflictiva en la funcionalidad de la proteína, o bien podría ser *frameshift*, provocando un PTC. Cavallari y colaboradores observaron que la variante en la posición +1 del intrón 6 del factor de coagulación VII (cuya ausencia se considera letal) daba lugar principalmente a la pérdida del exón (*frameshift*), pero también activaba el uso de un donador críptico que provocaba la delección *in-frame* de los últimos 30 nt del exón. La proteína resultante, carente de 10 aa, era funcional y capaz de llevar a cabo la coagulación (Cavallari et al., 2012). Otra posibilidad, la encontramos cuando una variante modifica la secuencia de unión a un silenciador que actuaba reprimiendo el uso del sitio críptico. Por ejemplo, un estudio reciente demostró la existencia de dos donadores crípticos (Cr1 y Cr2) localizados a 23 y a 51 nt, respectivamente, de la frontera exón/intrón del exón 7 del gen *SMN*. Según los resultados, el sitio Cr1 no se utilizaba debido a la unión de ESS-N1, un factor de silenciamiento de *splicing*. Curiosamente, observaron que la activación del uso de Cr1, restauraba la inclusión del exón 7 (Singh et al., 2017).

El mapeo funcional de secuencias reguladoras aporta una información precisa sobre las regiones necesarias para el reconocimiento de los exones. Estas secuencias son reconocidas por factores de *splicing* que pueden variar en función del tejido, del tipo celular, del momento del desarrollo, de las condiciones fisiológicas o de los estímulos externos, entre

otros, dando lugar a diferentes patrones de *splicing* alterativo (Lee and Rio, 2015). En los últimos años se han desarrollado distintas herramientas bioinformáticas que tratan de predecir los SREs (Grodecká et al., 2017). Sin embargo, los resultados obtenidos no son precisos y no están recomendados para uso en investigación clínica. A lo largo de esta tesis nos hemos encontrado con resultados que corroboran la falta de precisión de este tipo de programas informáticos. Un ejemplo lo encontramos en el estudio de los exones 14 y 15 de *BRCA2*, donde se seleccionaron 26 variantes que, según la bioinformática, alteraban SREs (Resultados; Tabla 1). Sin embargo, tras el ensayo funcional, ninguna de ellas afectó al *splicing*. Únicamente, la variante del exón 15 c.7501C>T, seleccionada por la supuesta creación de un nuevo 5'ss, resultó tener un efecto a través de la alteración de los ESEs del exón 15 (Resultados; Tabla 2). Otro ejemplo lo encontramos en las variantes c.7738C>T y c.7772A>G del exón 16 que, aunque fueron seleccionadas por crear sitios de unión a hnRNPA (Artículo 3; Tabla 2), se comportaron como variantes neutras desde el punto de vista del *splicing*. Entre las 200 variantes ensayadas en esta tesis, aproximadamente la mitad (101) fueron seleccionadas por alterar SREs, de las cuales sólo 30 generaron transcritos aberrantes. Cabe destacar que más de la mitad de estas variantes reguladoras (16/30) se encontraban localizadas en regiones de microdeleciones positivas, lo cual recalca el valor de la estrategia de mapeo funcional para la selección de variantes de DNA en SREs activos. Mediante esta técnica, estudios previos de nuestro grupo han demostrado la presencia de ESEs activos en los exones 19, 20, 21 y 23 (Acedo et al., 2015), elevando a 13 (3 a 8, 15, 17 a 21 y 23) el número de exones regulados por SREs en *BRCA2*.

Normalmente, las alteraciones provocadas por variantes que modifican SREs suelen tener efectos parciales con difícil interpretación clínica. La identificación de secuencias esenciales para el *splicing* en genes asociados a enfermedad facilita la tarea de selección y clasificación de estas variantes. Además, los datos obtenidos a través de estudios funcionales son datos fiables que se pueden incorporar a los algoritmos utilizados por las herramientas bioinformáticas para mejorar su precisión.

### C. VARIANTES EN *BRCA2*: EFECTO EN EL *SPLICING* Y SIGNIFICADO CLÍNICO

Normalmente, se consideran variantes de *splicing* aquellas que alteran las secuencias más conservadas (posiciones  $\pm 1$  o  $\pm 2$  de los ss). Sin embargo, debido a la presencia de elementos reguladores en *cis*, cualquier cambio en la secuencia de DNA podría, potencialmente, alterar el *splicing*. De hecho, como se ha mencionado previamente, se estima que el 60% de las variantes asociadas a enfermedad podrían ser espliceogénicas (López-Bigas et al., 2005). Durante este trabajo se han ensayado funcionalmente 200 variantes repartidas a lo largo de 13 exones de *BRCA2* (y sus secuencias intrónicas flanqueantes), desde el exón 2 hasta el exón 9 con el MGBR2\_2-9 (Artículo 4), y desde el exón 14 hasta el exón 18 con el MGBR2\_14-20 (Artículos 1, 3 y Resultados). Hemos encontrado que más de la mitad (104) alteraban el *splicing*. Entre las variantes espliceogénicas de *BRCA2*, 49 eran exónicas y 55 intrónicas. La mayoría de las variantes exónicas que alteraban el *splicing* eran variantes *missense*, aunque también encontramos *nonsense*, sinónimas y *frameshift*. Por ejemplo, la variante del exón 3 c.97G>A está clasificada como *missense* porque provocaría un cambio de glutámico a lisina en la proteína (p.Glu33Lys). Sin embargo, según nuestros resultados, sólo el 23% del transcrito procedente de dicho alelo portaría el cambio de aminoácido, mientras que más del 70% correspondería a  $\Delta 3$  (p.D23\_L105del) (Artículo 4; Tabla Suplementaria S6). Otro ejemplo es c.7802A>G (p.Tyr2601Cys) en el exón 16, que provoca aproximadamente la mitad del transcrito canónico portador de la variante y la otra mitad con una delección *frameshift* de 4 nt en el extremo 3' del exón ( $\Delta 6q4$ ) (Artículo 3; Figura 2 C). En el caso de las variantes sinónimas, que a menudo pasan desapercibidas o son filtradas por los algoritmos bioinformáticos, es importante tener en cuenta que podrían alterar el *splicing* (Sauna and Kimchi-Sarfaty, 2011; Supek et al., 2014). Un ejemplo lo encontramos en la variante c.441A>G en el exón 5 de *BRCA2* (p.Gln147=) que provoca un 96% de la isoforma  $\Delta 5$  (Artículo 4; Figura 3), una delección *frameshift* (p.P143Gfs\*23). También en el exón 18 de *BRCA2* la variante sinónima c.8331G>A (p.Lys2777=) generaba más de 50% de  $\Delta 18$  (p.Y2660Ffs\*43) (Artículo 1; Figura 4). Por otro lado, las variantes *nonsense* normalmente son clasificadas como patogénicas, pues generan PTC que alteran la proteína. También en el exón 18 de *BRCA2* encontramos la variante

c.8009C>A, clasificada como *nonsense* porque generaría un PTC en la posición 2670 de BRCA2 (p.Ser2670X). Sin embargo, c.8009C>A produce más de un 90% de  $\Delta 18$  (Artículo 1; Figura 4), debido supuestamente a la modificación de SREs en el exón 18, lo que generaría una delección *frameshift* (p.Y2660Ffs\*43). Estas variantes, cuyo resultado es mayoritariamente la alteración del proceso de *splicing*, deberían ser anotadas como variantes espliceogénicas, pues el efecto causado en la proteína no se corresponde con su clasificación actual.

Durante el desarrollo de esta tesis doctoral se ha llevado a cabo un estudio exhaustivo de variantes de *splicing* en *BRCA2* mediante ensayos funcionales con minigenes. Utilizando el minigen MGBR2\_2-9, se han ensayado funcionalmente 83 variantes, 53 de las cuales resultaron en una alteración del proceso de *splicing* (Artículo 4; Tabla Suplementaria S6). Por otro lado, con el MGBR2\_14-20 se ensayaron 117, de las cuales 51 alteraron el *splicing* [(Artículo 1; Tabla 1); (Artículo 3; Tabla 3); (Resultados; Tabla 2)]. El resultado más frecuente tras el ensayo de estas variantes fue *skipping* o pérdida del exón. La pérdida de todos los exones estudiados, excepto el exón 3 y el exón 17, se traduce en una delección *frameshift* y la aparición de un PTC en la proteína. Por tanto, las variantes que generan estos transcritos, inactivarían BRCA2. Por otro lado, los exones 3 y 17, aunque mantienen la pauta de lectura, pertenecen regiones importantes del gen. El exón 3 codifica la secuencia de unión a PALB2, esencial para la funcionalidad de la proteína, pues la eliminación de este dominio provoca graves consecuencias en la estructura y funcionalidad de BRCA2 (Xia et al., 2006). Entre las variantes actualmente clasificadas como patogénicas, encontramos varias que provocan principalmente  $\Delta 3$  (c.316+2T>C, c.316+3delA, c.316+5G>C y c.316+6T>C). Por su parte, el exón 17 codifica una región altamente conservada del dominio de unión a DNA y necesaria para llevar a cabo la recombinación homóloga. De hecho, se ha demostrado que la variante espliceogénica c.7976G>A (R2659K), que provoca  $\Delta 17$ , altera la funcionalidad de BRCA2, estando así asociada al desarrollo de la enfermedad (Farrugia et al., 2008; Wu et al., 2005). Luego, a pesar de ser delecciones *in-frame*, las variantes que produzcan mayoritariamente  $\Delta 3$  o  $\Delta 17$ , deberían considerarse como variantes patogénicas o probablemente patogénicas.

La interpretación clínica de variantes de *splicing* es una tarea compleja que debe tener en cuenta diferentes factores como la cantidad relativa de transcrito aberrante, el

mantenimiento de la pauta de lectura o la presencia de isoformas minoritarias capaces de mantener la funcionalidad de la proteína. De acuerdo con el trabajo de Vallée y colaboradores, cuando una variante genera más del 60% de transcritos aberrantes, debe considerarse como una alteración grave de *splicing* (Vallée et al., 2016). Siguiendo este criterio y como resultado de este trabajo, podemos afirmar que 73 de las 200 variantes ensayadas mediante minigenes alteraban gravemente al *splicing* (36 afectaban a los exones 2-9 y 37 a los exones 14-18). La recopilación de las variantes que alteraban gravemente el *splicing* aparece en la Tabla Anexa 5. Cabe destacar, que en este recuento hemos incluido la variante del exón 18 c.8331G>A, que se encuentra en el límite de la clasificación, puesto que generó un 40.7% de transcrito canónico (Artículo 1, Tabla Suplementaria S3).

Este trabajo comenzó evaluando las variantes que afectaban a los exones 17 y 18 de *BRCA2* con el minigen MGBR2\_14-20 (Artículo 1). Se clasificaron según ACMG 52 variantes comprendidas en esta región, de las cuales 12 fueron clasificadas como patogénicas y 8 como probablemente patogénicas desde el punto de vista del *splicing*. También, se analizó la variante c.7976+5G>T, que resultó ser una variante probablemente patogénica (clase 4 según los criterios ENIGMA) (Montalban et al., 2018). En la región comprendida entre los exones 14 y 16 se evaluaron un total de 20 variantes [(Artículo 3; Tabla 4); (Resultados; Tabla 3)]. Desde el punto de vista del *splicing*, 13 fueron clasificadas como patogénicas y cuatro como probablemente patogénicas. En total, utilizando el MGBR2\_14-20 hemos clasificado 37 variantes como patogénicas o probablemente patogénicas, de las cuales 16 estaban previamente clasificadas como VUS (Tabla Anexa 5). La variante del exón 16 c.7802A>G no aparece en el listado porque generó un 54% de transcrito canónico. Ésta fue probablemente la variante más conflictiva: según ACMG, c.7802A>G debería ser clasificada como probablemente patogénica; por el contrario, los criterios de ENIGMA indican que debería considerarse como clase 3 (Artículo 3; Tabla 4). Por otro lado, con el MGBR2\_2-9 evaluamos, siguiendo los criterios de ACMG, un total de 36 variantes que afectaban gravemente al *splicing*. Según nuestros resultados, 28 de ellas podrían clasificarse como variantes patogénicas y cinco como probablemente patogénicas. Las otras tres variantes, no cumplían los requisitos necesarios, por tanto, permanecieron como VUS. Cabe destacar que 13 de las

33 variantes clasificadas como patogénicas o probablemente patogénicas en la región 2-9 aparecían previamente anotadas como VUS (Artículo 4; Tabla 1) (Tabla Anexa 5). Además, la clasificación de 17 de las variantes como patogénicas coincidía con lo previamente reportado. Durante este trabajo, evaluamos también las 36 variantes siguiendo los criterios propuestos por ENIGMA. En este caso, debido a que las normas impuestas por esta guía son más restrictivas, sólo una de las variantes fue clasificada como clase 5 (c. c.470\_474del), la mayoría (29 variantes) fueron clasificadas como clase 4 y seis como clase 3. La principal diferencia entre las clases 4 y 5 propuestas por ENIGMA radica en la probabilidad de la patogenicidad de la variante según el modelo multifactorial, de manera que las variantes de clase 5 tienen una probabilidad >99% de ser patogénicas, mientras que en las de clase 4 la probabilidad se sitúa entre el 99 y el 95%. Sin embargo, a pesar de esta diferencia, las medidas clínicas recomendadas cuando se encuentra una variante en los genes *BRCA* son equivalentes para variantes de clase 4 y de clase 5, en ambos casos se activan las medidas de prevención. En total, 29 variantes (30 si tenemos en cuenta c.7802A>G) fueron reclasificadas desde VUS a variantes patogénicas o probablemente patogénicas (Tabla Anexa 5). La reclasificación de variantes de significado clínico desconocido a variantes causantes de la enfermedad (patogénicas o probablemente patogénicas / clase 4 o 5) tiene repercusión directa en la clínica aumentando el número de familias que podrían beneficiarse de las medidas preventivas y terapéuticas que se ofrecen a los portadores de mutaciones germinales en los genes *BRCA*. De esta forma, esta tesis doctoral aporta datos fiables sobre el procesamiento del RNA de *BRCA2*, generando un catálogo funcional con aplicabilidad directa en la clínica.

## **PRESENTE Y FUTURO EN EL DIAGNÓSTICO DE VUS, ENSAYOS FUNCIONALES Y TRATAMIENTOS.**

En los últimos años, el avance tecnológico en secuenciación, así como el abaratamiento de los costes, han hecho posible la implementación de secuenciación masiva, tanto en hospitales para el diagnóstico clínico, como en centros de investigación. El uso rutinario de este tipo de técnicas incrementa exponencialmente el número de variantes encontradas en los genes asociados a enfermedad. En las principales bases de datos (ClinVar, UMD y BIC) se van incluyendo las nuevas variantes encontradas. Sin embargo, debido, por una parte, a la gran cantidad de datos recogidos cada día, y por otra, a la participación de compañías privadas en estudios masivos, no tenemos acceso a todas las variantes detectadas. Por este motivo, los números que aparecen en estas bases de datos sean probablemente una subestimación de los reales.

La aparición de nuevas variantes trae como consecuencia el incremento en el número de VUS, pues el aumento en la capacidad de detección de variantes no ha llevado consigo un incremento en los estudios funcionales necesarios. En concreto, los datos de ClinVar para los genes *BRCA* (a fecha de marzo de 2019) recogen unas 7.000 variantes diferentes en *BRCA1*, de las cuales aproximadamente un 30% son VUS; y unas 10.000 en *BRCA2*, de las cuales en torno a un 40% son VUS; aunque, como se ha mencionado antes, probablemente se trate de una subestimación. La interpretación clínica de las variantes es uno de los mayores retos en el diagnóstico molecular del CMOH, particularmente si consideramos que las variantes patogénicas en los *BRCA* alcanzan a explicar solo en torno al 20% de los casos y que, al añadir los genes de moderado y bajo riesgo, esta cifra sólo se incrementa ligeramente (Davies et al., 2017; Lynch et al., 2013). Además, el conocimiento del impacto funcional de las variantes es crítico para el cálculo del riesgo asociado al desarrollo de la enfermedad, las estrategias de prevención y las opciones de tratamiento. Por ejemplo, el olaparib fue el primer fármaco aprobado por la FDA basado en la inhibición de PARP para el tratamiento de distintos tipos de cáncer de ovario en pacientes que portasen mutaciones en los genes *BRCA*, que más tarde, fue aprobado también para el tratamiento de tumores de mama metastásicos en pacientes con mutaciones germinales en *BRCA*. Además, la

identificación de variantes que alteren la funcionalidad de *BRCA2* no solo es beneficiosa para pacientes de cáncer de mama u ovario, sino también para pacientes con otro tipo de tumores con mutaciones en *BRCA2*, puesto que podrían beneficiarse de las mismas terapias. De hecho, recientemente se ha publicado un estudio que confirma el papel deletéreo de las mutaciones germinales en *BRCA2* en el cáncer de próstata metastásico resistente a la castración, sugiriendo la utilización de esta clase de terapias para el tratamiento de estos tumores (Castro et al., 2019).

Las variantes de efecto desconocido o VUS son habitualmente ignoradas debido a la falta de pruebas de su asociación con la enfermedad. Las VUS suelen ser variantes *missense*, sinónimas o deleciones/inserciones *in-frame*. No obstante, teóricamente cualquier cambio en la secuencia de DNA puede alterar mecanismos reguladores en el metabolismo de los ácidos nucleicos. Por ejemplo, se estima que hasta un tercio de las variantes patogénicas asociadas a CMOH están afectando el *splicing* (Castéra et al., 2014). El mayor problema reside en la dificultad para reconocer qué variantes van a causar cambios cualitativos y cuantitativos en los patrones de *splicing*. Existen para ello, herramientas bioinformáticas basadas en secuencias consenso de elementos en *cis*, como las utilizadas durante el transcurso de esta tesis. Sin embargo, según nuestra experiencia, son válidas para el rastreo de las regiones más conservadas como el sitio donador o aceptor (posiciones  $\pm 1$  y  $\pm 2$ ), pero no son precisas cuando se trata elementos menos conservados como los SREs. Estas herramientas, son usadas como método de selección previo a los ensayos funcionales, que son los que aportan verdaderos datos sobre cómo afecta una variante al procesamiento del RNA del gen en cuestión. En esta tesis doctoral se han llevado a cabo estudios funcionales mediante RT-PCR y análisis de los transcritos por electroforesis capilar y secuenciación SANGER. Sin embargo, existen otros métodos para determinar si una variante altera el *splicing*. En 2017, se publicó el primer artículo en el que se analizaban los transcritos de 11 genes asociados a CMOH (entre ellos *BRCA1* y *BRCA2*) mediante RNA-seq, presentando esta técnica como una forma rápida de analizar patrones aberrantes de *splicing* que podrían estar relacionados con el CMOH (Davy et al., 2017). Curiosamente, en este estudio una de las variantes utilizadas como control de *splicing* aberrante fue c.7975A>G de *BRCA2*. Esta

variante, ensayada en células MCF-7 con el minigen MGBR2\_14-20 durante el desarrollo de esta tesis, resultó en una pérdida parcial del exón 17, corroborando los resultados reportados en la bibliografía (Caux-Moncoutier et al., 2011), así como los obtenidos mediante los ensayos con el minigen MGBR2\_14-20 (Artículo 1; Tabla 1). Otra técnica utilizada para testar VUS, son los ensayos funcionales basados en células madre embrionarias de ratón (mESC). Esta técnica consiste en introducir el gen *BRCA2* mutado en mESCs que porten *Brca2* condicional en hemicigosis. De esta manera, si al delecionar *Brca2*, el gen *BRCA2* con la variante es capaz de rescatar la viabilidad celular, sabremos si dicha variante está afectando a la funcionalidad de la proteína (Kuznetsov et al., 2008). Utilizando esta técnica, recientemente, Mesman y colaboradores testaron un conjunto de variantes *missense* de *BRCA2*, de las cuales siete VUS fueron incapaces de rescatar el fenotipo letal, y otras cuatro VUS lo rescataron ligeramente (Mesman et al., 2018). Entre las VUS analizadas se encontraba c.7753G>A, una variante del exón 16 de *BRCA2* analizada en esta tesis doctoral (Artículo 3; Tabla 3). El resultado de ambos estudios indica que, a pesar de los resultados bioinformáticos, c.7753G>A no afecta el *splicing*, pero la variación *missense* si afecta a la funcionalidad de la proteína. Curiosamente, en el trabajo de Mesman, todas las VUS incapaces de rescatar el fenotipo, excepto c.93G>T (domino de unión a PALB2), se encontraban alterando la secuencia de aminoácidos del dominio de unión a DNA de *BRCA2*, lo que recalca la importancia de la integridad de esta región en la proteína.

Existen muchas enfermedades causadas por alteraciones en el *splicing*. Éstas pueden clasificarse en dos grandes grupos: aquellas causadas por alteraciones en *cis* (*cis-acting mutations*), cuando la alteración de *splicing* se produce por cambios en la secuencia del gen alterado; y aquellas causadas por alteraciones en *trans* (*trans-acting factors*), cuando son los elementos de la maquinaria de *splicing* los que se encuentran alterados. Por tanto, las terapias basadas en la corrección de los patrones de *splicing* pueden llevarse a cabo desde dos puntos de vista: modificando los componentes del espliceosoma, es decir, alterando el proceso a nivel global; o modificando específicamente el *splicing* del gen causante de la enfermedad. La mayoría de las terapias centradas en corregir el *splicing* consisten en el uso de oligonucleótidos complementarios al pre-mRNA de interés. Los oligonucleótidos usados

en terapia suelen tener una longitud de 15 a 30 nucleótidos (Levin, 2019). Estos, una vez entran en la célula, se unen a la molécula de mRNA (o pre-mRNA) complementaria. El efecto que provocan puede variar en función de cómo hayan sido diseñados: destrucción de mRNA por corte enzimático, modificación del *splicing* o cambio en la función de un RNA regulador. Entre los que modifican los patrones de *splicing*, un ejemplo es el nusinersen, empleado para el tratamiento de la atrofia muscular espinal, una enfermedad autosómica recesiva causada por mutaciones en el gen *SMN1* y caracterizada por la pérdida de la fuerza muscular debido a la afectación de las motoneuronas de la médula espinal. El gen *SMN1* tiene un parálogo llamado *SMN2*, en el que, debido a un cambio de nucleótido, la maquinaria de *splicing* no reconoce el exón 7, dando lugar a una alteración que termina en la degradación del mRNA. Nusinersen es un oligonucleótido diseñado para hibridar con la secuencia que contiene la mutación en *SMN2* y bloquear la señal que conduce a la pérdida del exón. Así, nusinersen rescata la inclusión del exón 7 y permite la producción de una proteína funcional, capaz de restaurar la funcionalidad neuronal en pacientes con mutaciones patogénicas en *SMN1* (Levin, 2019; Ottesen, 2017). Otro ejemplo de tratamiento mediante oligonucleótidos lo encontramos en la distrofia muscular de Duchenne, causada por mutaciones en el gen *DMD*. El oligonucleótido llamado eteplirsén, aprobado actualmente por la FDA, provoca la pérdida del exón 51, una delección *in frame* capaz de restaurar la pauta de lectura de la proteína. Como resultado, se genera una proteína más pequeña, pero funcional. Aproximadamente el 14% de los pacientes portan mutaciones que pueden ser rescatadas a través de la delección del exón 51 y por tanto tratadas con eteplirsén (Lim et al., 2017). Existen modificaciones de estos oligonucleótidos que combinan distintas estrategias, como los oligonucleótidos bifuncionales que son capaces de reclutar factores de *splicing* que modulen el proceso (Skordis et al., 2003). Aunque los oligonucleótidos son herramientas efectivas, el diseño de cada uno debe ser específico para cada caso, y teniendo en cuenta que cada oligonucleótido es considerado como un nuevo agente terapéutico, estas aproximaciones resultan impensables para el tratamiento de pacientes con variantes raras como es el caso de *BRCA2*. Actualmente, aunque hay otros en ensayo, nusinersen y

eteplirsen son los únicos dos oligonucleótidos aprobados que actúan a nivel de *splicing* (Stein and Castanotto, 2017).

Existen otras estrategias terapéuticas para la corrección del *splicing*. Una de ellas es la modificación del factor U1 del espliceosoma. La técnica consiste en restaurar la complementariedad entre el pre-RNA mutado y U1 mediante la modificación de la secuencia de unión a RNA de U1. Ya en 2008, el grupo de Franco Pagani demostró que la modificación de U1 era una estrategia válida para la corrección de mutaciones de donador del factor de coagulación VII (Pinotti et al., 2008). Más tarde, el mismo grupo desarrolló ExSpeU1 (*Exonic specific U1 snRNA*) para corregir errores en los genes causantes de hemofilia B, fibrosis quística y atrofia muscular espinal (Fernandez Alanis et al., 2012). Recientemente, se ha publicado por primera vez la restauración, mediante este método (ExSpeU1), de una variante en la posición +2 del intrón 3 del factor de coagulación IX, causante de hemofilia B (Scalet et al., 2019). No obstante, y a pesar de su elevado potencial como estrategia terapéutica, actualmente no existe ningún tratamiento aprobado basado en la modificación de este factor de *splicing*. Otra estrategia es la conocida como SMaRT (*Spliceosome-mediated RNA Trans-splicing*). SMaRT se basa en la capacidad del espliceosoma para llevar a cabo el *splicing* entre dos hebras de pre-mRNA diferentes (Mansfield et al., 2004). Este método consiste en diseñar un oligonucleótido llamado PTM (*pre-transplicing molecule*) para reemplazar la secuencia mutada. El PTM debe contener una secuencia complementaria al intrón donde se quiere reparar el *splicing*, seguido del exón wt con secuencias de *splicing* conservadas. Así el espliceosoma seleccionará la nueva hebra obteniendo un transcrito canónico. Esta aproximación se ha ensayado en distintas enfermedades causadas por mutaciones de *splicing* como fibrosis quística, hemofilia B, atrofia muscular espinal o distrofia muscular de Duchenne, pero en el mejor de los casos solo entre el 20-40% del mRNA obtenido carecía de la alteración (Berger et al., 2016). Aun así, SMaRT representa una buena estrategia para la corrección del *splicing* en enfermedades con variantes raras, pues un solo PTM podría reparar el efecto de múltiples variantes.

También se han llevado a cabo algunas de estas aproximaciones a nivel del *splicing* para el tratamiento de determinados tipos de cáncer. Wan y colaboradores utilizaron un

oligonucleótido para inducir la pérdida del exón 15 del gen *HER2*, sobre-expresado en muchos tipos de tumores (entre ellos tumor de mama ERBB2). Así, consiguieron disminuir la expresión de HER2 en la línea celular SK-BR-3, lo que llevó a una bajada de la proliferación y un aumento de la apoptosis (Wan et al., 2009). Otro ejemplo lo encontramos en melanoma, donde un 30% de los pacientes con mutaciones en el gen *BRAF* desarrollan resistencia al tratamiento con vemurafenib, un anticuerpo monoclonal inhibidor de RAS quinasa. El mecanismo que confiere la resistencia consiste a menudo en la pérdida del dominio de unión a RAS (exones 3-5 de *BRAF*) debido a una alteración de *splicing*. Salton y colaboradores, consiguieron modular el *splicing in vitro* e *in vivo*, a través de la inhibición de SF3B1, un factor de *splicing* necesario para la pérdida del dominio de unión a RAS (Salton et al., 2015).

En todos los casos mencionados, las herramientas utilizadas muestran resultados prometedores para el tratamiento de enfermedades causadas por alteraciones en el *splicing*. Sin embargo, aunque se han superado algunas barreras como la síntesis química de oligonucleótidos, hoy en día rápida, económica y accesible, aún quedan obstáculos que superar antes de poder incorporar dichas aproximaciones a la clínica. Entre los principales desafíos encontramos: los efectos *off-target*, la toxicidad y los métodos de administración del fármaco. Afortunadamente, debido al conocimiento de la secuencia del genoma humano y el avance en los programas bioinformáticos, podemos diseñar secuencias altamente específicas, lo cual ha reducido considerablemente los efectos *off-target*. No obstante, es importante tener en cuenta la posibilidad de que dichos efectos pueden ocurrir, pues se trata de secuencias de 15-30 nt que podrían hibridar inespecíficamente. Por otro lado, la administración de oligonucleótidos ha sido asociada con la activación de la respuesta inmune innata, lo que podría provocar graves efectos adversos durante este tipo de tratamiento (Levin, 2019). El método de administración sigue siendo el mayor de los retos a superar, tanto los oligonucleótidos como los vectores que contienen snRNA U1 modificado o SMaRT deben ser administrados y transportados hasta el núcleo, donde podrán ejercer su acción. Esto se convierte en una difícil tarea debido a la naturaleza hidrofílica de los ácidos nucleicos, que les confiere una capacidad limitada para atravesar la membrana celular. Actualmente, los métodos basados en virus, como retrovirus, adenovirus o virus adeno-

asociados, para introducir ácidos nucleicos en las células, son los más exitosos debido a su alta capacidad en la transfección. Sin embargo, estos presentan ciertas desventajas: inmunogenicidad, integración aleatoria en el genoma (retrovirus y lentivirus) y potencial de transmisión. Por otro lado, existen métodos no basados en virus como las nanopartículas lipídicas, que representan una verdadera promesa en terapia génica. La principal ventaja de las nanopartículas es la libertad en el diseño, por ejemplo, pueden modificarse hasta obtener valores inmunogénicos prácticamente nulos, pueden ser diseñadas para la administración por distintas vías, pueden dirigirse específicamente a un tejido concreto mediante la incorporación de señalizadores y puede obtenerse una liberación lenta del fármaco si fuese deseable. Sin embargo, y a pesar de los progresos obtenidos durante los últimos años, la mayoría de las nanopartículas están aún en desarrollo (Suñé-Pou et al., 2017).

En resumen, las nuevas técnicas de secuenciación masiva generan gran cantidad de datos, apareciendo nuevas variantes cuyo efecto sobre el metabolismo de los ácidos nucleicos es a menudo desconocido. Estas variantes de significado clínico desconocido deberían ser evaluadas, pues podrían estar afectando a procesos fundamentales como la transcripción o el *splicing*. Existen distintos métodos para evaluar el riesgo asociado a una variante de DNA, entre los que destacan los estudios de asociación en poblaciones, las variables familiares o epidemiológicas (segregación, co-ocurrencia o historia familiar de cáncer), las aproximaciones bioinformáticas, los ensayos funcionales y el análisis multifactorial, que engloba todos los parámetros anteriores. Aunque es recomendable evaluar una variante utilizando diferentes métodos, realmente, son los ensayos funcionales los que aportan datos para comprender cómo afecta una determinada variante a la actividad génica. Cuando una variante es candidata de afectar al *splicing*, se recomienda comprobar los transcritos generados en una muestra del paciente portador de la misma. No obstante, cuando no se dispone de dicha muestra, los minigenes representan una herramienta útil para la comprobación del efecto. Además, los minigenes permiten el estudio detallado de regiones concretas de un gen, aportando información precisa, tanto de las variantes como de elementos reguladores de *splicing*. De hecho, entre los principales beneficios de los minigenes encontramos la flexibilidad a la hora de estudiar cualquier cambio en la secuencia

de DNA, así como la capacidad evaluar el efecto de variantes artificiales; y la posibilidad de cuantificar las isoformas resultantes sin la interferencia del alelo wt. Toda la información generada a través del estudio funcional de variantes repercute en la comprensión de los mecanismos celulares básicos, que luego son utilizados para predecir la aparición de la enfermedad o para crear estrategias de tratamiento que permitan rescatar el efecto de mutaciones patogénicas. En concreto, el conocimiento de la regulación de los procesos de transcripción y *splicing* en distintas condiciones revierte, tanto en la mejora del diagnóstico clínico, como en la precisión de los algoritmos utilizados en las herramientas bioinformáticas de predicción, que a su vez son utilizados en los primeros pasos en la clasificación clínica de variantes candidatas.



# CONCLUSIONES



## CONCLUSIONES

---

1. El mapeo funcional del promotor de *BRCA2* mostró la presencia de seis regiones activadoras y tres represoras:
  - a. La eliminación de la región  $-59_{-}10$  desde el TSS (microdeleción 3) provocó la mayor caída de la tasa transcripcional ( $>80\%$ ), sugiriendo ser una zona de unión a factores esenciales como la RNAPol II y los GTFs.
  - b. La eliminación de las secuencias  $+32_{+}81$  (microdeleción 1) y  $-14_{+}36$  (microdeleción 2) provocó un aumento del 356% y 180% respectivamente, de la expresión de luciferasa, sugiriendo la presencia de fuertes elementos represores en dicha región.
2. El promotor de *BRCA2* es altamente sensible a cambios de nucleótido:
  - a. La mayoría de los SNPs ensayados (10/15) alteraban significativamente la transcripción. Tres de ellos la reducían ligeramente (17-28%).
  - b. El *cluster* 121, situado desde la posición -185 hasta la +105, concentra ocho de las 10 variantes reguladoras que modifican sensiblemente la actividad transcripcional.
3. El reconocimiento de al menos nueve de los 27 exones de *BRCA2* está regulado por SREs:
  - a. Los exones 3 a 8, 15, 17 y 18 de *BRCA2* están regulados por *enhancers* de *splicing*.
  - b. El factor de *splicing* SFSR3 (SRp20) está implicado en el reconocimiento del exón 3, mientras que el SRSF2 (SC35) está implicado en el reconocimiento del exón 18.
  - c. El reconocimiento del exón 5 está fuertemente regulado por *enhancers* de *splicing* y efectos de contexto genómico de exones vecinos (4 y 6). La identificación de tres variantes espliceogénicas (c.440A>T, c.441A>G, y c.451G>A) sugiere la existencia de un *hotspot* para variantes de *splicing*.
  - d. Los exones 14 a 18 contienen sitios críticos de *splicing* que pueden ser activados por variantes de DNA.

- e. De las variantes espliceogénicas que alteraban SREs, más de la mitad (16/30) se encontraban localizadas en regiones de microdeleciones positivas.
4. La mayoría de las variantes ensayadas (104/200) alteraban el *splicing* de *BRCA2*, de las cuales el 70% (73/104) provocaron graves alteraciones:
- a. Cualquier tipo de variante puede alterar el *splicing*, ya que 55 variantes eran intrónicas y 49 exónicas de todos los tipos (*missense*, *sinónimas*, *nonsense* y *frameshift*).
  - b. Diecisiete variantes clasificadas en función de su efecto en la proteína deberían ser consideradas como variantes espliceogénicas: 14 *missense*, una *nonsense*, una sinónima y una *frameshift*.
  - c. El evento aberrante de *splicing* más común fue la pérdida del exón, seguido por el uso de sitios crípticos de *splicing*.
  - d. Veintisiete de las variantes estudiadas produjeron resultados similares a los previamente publicados, demostrando la reproducibilidad y fiabilidad de los minigenes MGBR2\_2-9 y MGBR2\_14-20 como herramientas apropiadas para este tipo de estudios.
  - e. Un total de 74 variantes de *BRCA2* (30 de ellas VUS) han sido clasificadas como patogénicas o probablemente patogénicas según las normas de ACMG, confirmando la utilidad de la tecnología de minigenes para la caracterización inicial de variantes de *splicing*.

# **BIBLIOGRAFÍA**



## BIBLIOGRAFÍA

---

- Acedo, A., Hernández-Moro, C., Curiel-García, Á., Díez-Gómez, B., and Velasco, E. a (2015). Functional Classification of BRCA2 DNA Variants by Splicing Assays in a Large Minigene with 9 Exons. *Hum. Mutat.* 36, 210–21. doi:10.1002/humu.22725.
- Acedo, A., Sanz, D. J., Duran, M., Infante, M., Perez-Cabornero, L., Miner, C., et al. (2012). Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res.* 14, R87. doi:10.1186/bcr3202.
- Allender, Y., and Arias, A. (2008). Cáncer de mama y ovario hereditario. *Rev Per Ginecol Obs.* 54, 194–198.
- Alonso Sánchez M. et al. (2006). CÁNCER HEREDITARIO. *SEOM*. Available at: <http://contenidos.institutoroche.es/pdf/libro1.pdf>.
- Änkö, M.-L. (2014). Regulation of gene expression programmes by serine–arginine rich splicing factors. *Semin. Cell Dev. Biol.* 32, 11–21. doi:10.1016/j.semcdb.2014.03.011.
- Anna, A., and Monika, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.*, 1–16. doi:10.1007/s13353-018-0444-7.
- Arnold, S., Buchanan, D. D., Barker, M., Jaskowski, L., Walsh, M. D., Birney, G., et al. (2009). Classifying MLH1 and MSH2 Variants Using Bioinformatic Prediction , Splicing Assays , Segregation , and Tumor Characteristics. doi:10.1002/humu.20936.
- Baynes, C., Healey, C. S., Pooley, K. A., Scollen, S., Luben, R. N., Thompson, D. J., et al. (2007). Common variants in the ATM, BRCA1, BRCA2, CHEK2 and TP53 cancer susceptibility genes are unlikely to increase breast cancer risk. *Breast Cancer Res.* 9, R27. doi:10.1186/bcr1669.
- Berger, A., Maire, S., Gaillard, M.-C., Sahel, J.-A., Hantraye, P., and Bemelmans, A.-P. (2016). mRNA *trans*-splicing in gene therapy for genetic diseases. *Wiley Interdiscip. Rev. RNA* 7, 487–498. doi:10.1002/wrna.1347.
- Berget, S. M., and Sharp, P. A. A spliced sequence at the 5'-terminus of adenovirus late mRNA. *Brookhaven Symp. Biol.*, 332–44. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/754866> [Accessed October 18, 2017].
- Beroud, C., Letovsky, S. I., Braastad, C. D., Caputo, S. M., Beaudoux, O., Bignon, Y. J., et al. (2016). BRCA Share: A Collection of Clinical BRCA Gene Variants. *Hum. Mutat.* 37, 1318–1328. doi:10.1002/humu.23113.
- Bièche, I., Noguès, C., and Lidereau, R. (1999). Overexpression of BRCA2 gene in sporadic breast tumours. *Oncogene* 18, 5232–8. doi:10.1038/sj.onc.1202903.
- Bonnet, C., Krieger, S., Vezain, M., Rousselin, a, Tournier, I., Martins, a, et al. (2008). Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *J. Med. Genet.* 45, 438–46. doi:10.1136/jmg.2007.056895.
- Bosch, A., Eroles, P., Zaragoza, R., Viña, J. R., and Lluch, A. (2010). Triple-negative breast cancer: Molecular features, pathogenesis, treatment and current lines of research. *Cancer Treat. Rev.* 36, 206–215. doi:10.1016/j.ctrv.2009.12.002.
- Boulton, S. J. (2006). Cellular functions of the BRCA tumour-suppressor proteins. *Biochem. Soc. Trans.*

34, 633–45. doi:10.1042/BST0340633.

- Broca, P. (1866). *Traité des tumeurs*. Paris: P. Asselin Available at: <http://www.worldcat.org/title/traite-des-tumeurs/oclc/490652636> [Accessed April 12, 2018].
- Bruwer, A., Bargen, J. A., and Kierland, R. R. (1954). Surface pigmentation and generalized intestinal polyposis; (Peutz-Jeghers syndrome). *Proc. Staff Meet. Mayo Clin.* 29, 168–71. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/13145642>.
- Buratti, E., Baralle, M., and Baralle, F. E. (2006). Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res.* 34, 3494–510. doi:10.1093/nar/gkl498.
- Burke, L. J., Sevcik, J., Gambino, G., Tudini, E., Mucaki, E. J., Shirley, B. C., et al. (2018). BRCA1 and BRCA2 5' noncoding region variants identified in breast cancer patients alter promoter activity and protein binding. *Hum. Mutat.* 39, 2025–2039. doi:10.1002/humu.23652.
- Burrows, N. P., Nicholls, a C., Richards, a J., Luccarini, C., Harrison, J. B., Yates, J. R., et al. (1998). A point mutation in an intronic branch site results in aberrant splicing of COL5A1 and in Ehlers-Danlos syndrome type II in two British families. *Am. J. Hum. Genet.* 63, 390–8. doi:10.1086/301948.
- Butler, J. E. F., and Kadonaga, J. T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16, 2583–2592. doi:10.1101/gad.1026202.The.
- Buys, S. S., Sandbach, J. F., Gammon, A., Patel, G., Kidd, J., Brown, K. L., et al. (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. doi:10.1002/cncr.30498.
- Cartegni, L., Chew, S. L., and Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3, 285–98. doi:10.1038/nrg775.
- Castéra, L., Krieger, S., Rousselin, A., Legros, A., Baumann, J.-J., Bruet, O., 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, 1305–13. doi:10.1038/ejhg.2014.16.
- Castro, E., Romero-Laorden, N., Del Pozo, A., Lozano, R., Medina, A., Puente, J., et al. (2019). PROREPAIR-B: A Prospective Cohort Study of the Impact of Germline DNA Repair Mutations on the Outcomes of Patients With Metastatic Castration-Resistant Prostate Cancer. *J. Clin. Oncol.* JCO1800358. doi:10.1200/JCO.18.00358.
- Caux-Moncoutier, V., Castéra, L., Tirapo, C., Michaux, D., Rémon, M.-A., Laugé, A., et al. (2011). EMMA, a cost- and time-effective diagnostic method for simultaneous detection of point mutations and large-scale genomic rearrangements: application to BRCA1 and BRCA2 in 1,525 patients. *Hum. Mutat.* 32, 325–334. doi:10.1002/humu.21414.
- Cavallari, N., Balestra, D., Branchini, A., Maestri, I., Chuamsunrit, A., Sasanakul, W., et al. (2012). Activation of a cryptic splice site in a potentially lethal coagulation defect accounts for a functional protein variant. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1822, 1109–1113. doi:10.1016/J.BBADIS.2012.03.001.
- Chabot, B., and Shkreta, L. (2016). Defective control of pre-messenger RNA splicing in human disease. *J. Cell Biol.* 212, 13–27. doi:10.1083/jcb.201510032.
- Chen, S., Iversen, E. S., Friebel, T., Finkelstein, D., Weber, B. L., Eisen, A., et al. (2006). Characterization of BRCA1 and BRCA2 mutations in a large United States sample. *J. Clin. Oncol.* 24, 863–71. doi:10.1200/JCO.2005.03.6772.

- Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/902310> [Accessed October 18, 2017].
- Cline, M. S., Liao, R. G., Parsons, M. T., Paten, B., Alquaddoomi, F., Antoniou, A., et al. (2018). BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. *PLOS Genet.* 14, e1007752. doi:10.1371/journal.pgen.1007752.
- Colombo, M., De Vecchi, G., Caleca, L., Foglia, C., Ripamonti, C. B., Ficarazzi, F., et al. (2013). Comparative In Vitro and In Silico Analyses of Variants in Splicing Regions of BRCA1 and BRCA2 Genes and Characterization of Novel Pathogenic Mutations. *PLoS One* 8, e57173. doi:10.1371/journal.pone.0057173.
- Colombo, M., Lòpez-Perolio, I., Meeks, H. D., Caleca, L., Parsons, M. T., Li, H., et al. (2018). The BRCA2 c.68-7T > A variant is not pathogenic: A model for clinical calibration of spliceogenicity. *Hum. Mutat.* 39, 729–741. doi:10.1002/humu.23411.
- Colombo, M., Ripamonti, C. B., Pensotti, V., Foglia, C., Peissel, B., Pierotti, M. A., et al. (2009). An unusual BRCA2 allele carrying two splice site mutations. *Ann. Oncol.* 20, 1143–1144. doi:10.1093/annonc/mdp241.
- Consortium, T. E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi:10.1038/nature11247.
- Cooper, D. N. (2002). Human gene mutation in pathology and evolution. *J. Inherit. Metab. Dis.* 25, 157–182. doi:10.1023/A:1015621710660.
- Cooper, T. a (2005). Use of minigene systems to dissect alternative splicing elements. *Methods* 37, 331–40. doi:10.1016/j.ymeth.2005.07.015.
- Crossley, M., and Brownlee, G. G. (1990). Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B. *Nature* 345, 444–446. doi:10.1038/345444a0.
- Curtin, N. (2014). PARP inhibitors for anticancer therapy. *Biochem. Soc. Trans.* 42, 82–8. doi:10.1042/BST20130187.
- Danino, Y. M., Even, D., Ideses, D., and Juven-Gershon, T. (2015). The core promoter: At the heart of gene expression. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1849, 1116–1131. doi:10.1016/j.bbagr.2015.04.003.
- Davies, H., Glodzik, D., Morganella, S., Yates, L. R., Staaf, J., Zou, X., et al. (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* 23, 517–525. doi:10.1038/nm.4292.
- Davis, P. L., Miron, a, Andersen, L. M., Iglehart, J. D., and Marks, J. R. (1999a). Isolation and initial characterization of the BRCA2 promoter. *Oncogene* 18, 6000–6012. doi:10.1038/sj.onc.1202990.
- Davis, P. L., Miron, A., Andersen, L. M., Iglehart, J. D., and Marks, J. R. (1999b). Isolation and initial characterization of the BRCA2 promoter. 6000–6012.
- Davy, G., Rousselin, A., Goardon, N., Castéra, L., Harter, V., Legros, A., et al. (2017). Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. *Eur. J. Hum. Genet.* 25, 1147–1154. doi:10.1038/ejhg.2017.116.
- De Conti, L., Baralle, M., and Buratti, E. (2012). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* 4, 49–60. doi:10.1002/wrna.1140.

- de la Hoya, M., Soukarieh, O., López-Perolio, I., Vega, A., Walker, L. C., van Ierland, Y., et al. (2016). Combined genetic and splicing analysis of BRCA1 c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum. Mol. Genet.* 25, 2256–2268. doi:10.1093/hmg/ddw094.
- de Vooght, K. M. K., van Wijk, R., and van Solinge, W. W. (2009). Management of gene promoter mutations in molecular diagnostics. *Clin. Chem.* 55, 698–708. doi:10.1373/clinchem.2008.120931.
- Deans, A. J., and West, S. C. (2011). DNA interstrand crosslink repair and cancer. *Nat. Rev. Cancer* 11, 467–80. doi:10.1038/nrc3088.
- Di Giacomo, D., Gaildrat, P., Abuli, A., Abdat, J., Frébourg, T., Tosi, M., et al. (2013). Functional analysis of a large set of BRCA2 exon 7 variants highlights the predictive value of hexamer scores in detecting alterations of exonic splicing regulatory elements. *Hum. Mutat.* 34, 1547–57. doi:10.1002/humu.22428.
- dos Santos, E. S., Caputo, S. M., Castera, L., Gendrot, M., Briaux, A., Breault, M., et al. (2018). Assessment of the functional impact of germline BRCA1/2 variants located in non-coding regions in families with breast and/or ovarian cancer predisposition. *Breast Cancer Res. Treat.* 168, 311–325. doi:10.1007/s10549-017-4602-0.
- Eccles, D. M., Mitchell, G., Monteiro, A. N. A., Schmutzler, R., Couch, F. J., Spurdle, A. B., et al. (2015). *BRCA1* and *BRCA2* genetic testing—pitfalls and recommendations for managing variants of uncertain clinical significance. *Ann. Oncol.* 26, 2057–2065. doi:10.1093/annonc/mdv278.
- Esashi, F., Christ, N., Gannon, J., Liu, Y., Hunt, T., Jasin, M., et al. (2005). CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* 434, 598–604. doi:10.1038/nature03404.
- Fackenthal, J. D., Yoshimatsu, T., Zhang, B., de Garibay, G. R., Colombo, M., De Vecchi, G., et al. (2016). Naturally occurring <em>BRCA2</em> alternative mRNA splicing events in clinically relevant samples. *J. Med. Genet.* 53, 548 LP-558. Available at: <http://jmg.bmj.com/content/53/8/548.abstract>.
- Fairbrother, W. G., Holste, D., Burge, C. B., and Sharp, P. a. (2004). Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol.* 2. doi:10.1371/journal.pbio.0020268.
- Farber-Katz, S., Hsuan, V., Wu, S., Landrith, T., Vuong, H., Xu, D., et al. (2018). Quantitative Analysis of BRCA1 and BRCA2 Germline Splicing Variants Using a Novel RNA-Massively Parallel Sequencing Assay. *Front. Oncol.* 8, 286. doi:10.3389/fonc.2018.00286.
- Farrugia, D. J., Agarwal, M. K., Pankratz, V. S., Deffenbaugh, A. M., Pruss, D., Frye, C., et al. (2008). Functional assays for classification of BRCA2 variants of uncertain significance. *Cancer Res.* 68, 3523–31. doi:10.1158/0008-5472.CAN-07-1587.
- Fernandez Alanis, E., Pinotti, M., Dal Mas, A., Balestra, D., Cavallari, N., Rogalska, M. E., et al. (2012). An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum. Mol. Genet.* 21, 2389–2398. doi:10.1093/hmg/dds045.
- Figueiredo, J. C., Brooks, J. D., Conti, D. V., Poynter, J. N., Teraoka, S. N., Malone, K. E., et al. (2011). Risk of contralateral breast cancer associated with common variants in BRCA1 and BRCA2: potential modifying effect of BRCA1/BRCA2 mutation carrier status. *Breast Cancer Res. Treat.* 127, 819–829. doi:10.1007/s10549-010-1285-1.
- Fisher, B., Costantino, J. P., Wickerham, D. L., Cecchini, R. S., Cronin, W. M., Robidoux, A., et al. (2005). Tamoxifen for the Prevention of Breast Cancer: Current Status of the National Surgical Adjuvant

- Breast and Bowel Project P-1 Study. *JNCI J. Natl. Cancer Inst.* 97, 1652–1662. doi:10.1093/jnci/dji372.
- Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., et al. (1998). Genetic Heterogeneity and Penetrance Analysis of the BRCA1 and BRCA2 Genes in Breast Cancer Families. *Am. J. Hum. Genet.* 62, 676–689. doi:10.1086/301749.
- Fraile-Bethencourt, E., Díez-Gómez, B., Velásquez-Zapata, V., Acedo, A., Sanz, D. J., and Velasco, E. a. (2017). Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. *PLOS Genet.* 13, e1006691. doi:10.1371/journal.pgen.1006691.
- Fraile-Bethencourt, E., Valenzuela-Palomo, A., Díez-Gómez, B., Acedo, A., and Velasco, E. A. (2018). Identification of Eight Spliceogenic Variants in BRCA2 Exon 16 by Minigene Assays. *Front. Genet.* 9, 188. doi:10.3389/fgene.2018.00188.
- Gaildrat, P., Krieger, S., Di Giacomo, D., Abdat, J., Révillion, F., Caputo, S., et al. (2012). Multiple sequence variants of BRCA2 exon 7 alter splicing regulation. *J. Med. Genet.* 49, 609–17. doi:10.1136/jmedgenet-2012-100965.
- Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. *Hum. Genet.* 135, 851–67. doi:10.1007/s00439-016-1683-5.
- Gilbert, W. (1978). Why genes in pieces? *Nature* 271, 501–501. doi:10.1038/271501a0.
- Gochhait, S., Bukhari, S. I. A., Bairwa, N., Vadhera, S., Darvishi, K., Raish, M., et al. (2007). Implication of BRCA2 -26G>A 5' untranslated region polymorphism in susceptibility to sporadic breast cancer and its modulation by p53codon 72 Arg>Pro polymorphism. *Breast Cancer Res.* 9, R71. doi:10.1186/bcr1780.
- Goldgar, D. E., Easton, D. F., Deffenbaugh, A. M., Monteiro, A. N. A., Tavtigian, S. V, Couch, F. J., et al. (2004). Integrated Evaluation of DNA Sequence Variants of Unknown Clinical Significance : Application to BRCA1 and BRCA2. 535–544.
- Green, R., Green, A., Simms, M., Pater, A., Robb, J., and Green, J. (2003). Germline hMLH1 promoter mutation in a Newfoundland HNPCC kindred. *Clin. Genet.* 64, 220–227. doi:10.1034/j.1399-0004.2003.t01-1-00110.x.
- Grodecká, L., Buratti, E., and Freiburger, T. (2017). Mutations of pre-mRNA splicing regulatory elements: Are predictions moving forward to clinical diagnostics? *Int. J. Mol. Sci.* 18. doi:10.3390/ijms18081668.
- Guidugli, L., Carreira, A., Caputo, S. M., Ehlen, A., Galli, A., Monteiro, A. N. a, et al. (2014). Functional assays for analysis of variants of uncertain significance in BRCA2. *Hum. Mutat.* 35, 151–64. doi:10.1002/humu.22478.
- Gutiérrez-Enríquez, S., Coderch, V., Masas, M., Balmaña, J., and Díez, O. (2009). The variants BRCA1 IVS6-1G>A and BRCA2 IVS15+1G>A lead to aberrant splicing of the transcripts. *Breast Cancer Res. Treat.* 117, 461–465. doi:10.1007/s10549-008-0154-7.
- Habara, Y., Takeshima, Y., Awano, H., Okizuka, Y., Zhang, Z., Saiki, K., et al. (2008). In vitro splicing analysis showed that availability of a cryptic splice site is not a determinant for alternative splicing patterns caused by +1G-->A mutations in introns of the dystrophin gene. *J. Med. Genet.* 46, 542–7. doi:10.1136/jmg.2008.061259.
- Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., et al. (1990). Linkage of

- early-onset familial breast cancer to chromosome 17q21. *Science* 250, 1684–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2270482> [Accessed April 12, 2018].
- Han, S. P., Tang, Y. H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. *Biochem. J.* 430, 379–92. doi:10.1042/BJ20100396.
- Hansen, T. V. O., Steffensen, A. Y., Jønson, L., Andersen, M. K., Ejlersen, B., and Nielsen, F. C. (2010). The silent mutation nucleotide 744 G → A, Lys172Lys, in exon 6 of BRCA2 results in exon skipping. *Breast Cancer Res. Treat.* 119, 547–550. doi:10.1007/s10549-009-0359-4.
- Hardison, R. C., Chui, D. H. K., Giardine, B., Riemer, C., Patrinos, G. P., Anagnou, N., et al. (2002). HbVar: A relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. *Hum. Mutat.* 19, 225–233. doi:10.1002/humu.10044.
- Harper, A. R., Nayee, S., and Topol, E. J. (2015). Protective alleles and modifier variants in human health and disease. *Nat. Rev. Genet.* 16, 689–701. doi:10.1038/nrg4017.
- Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., et al. (2012). Dynamic Protein-Protein Interaction Wiring of the Human Spliceosome. *Mol. Cell* 45, 567–580. doi:10.1016/J.MOLCEL.2011.12.034.
- Hemminki, A., Markie, D., Tomlinson, I., Avizienyte, E., Roth, S., Loukola, A., et al. (1998). A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature* 391, 184–187. doi:10.1038/34432.
- Highsmith, W. E., Burch, L. H., Zhou, Z., Olsen, J. C., Strong, T. V., Smith, T., et al. (1997). Identification of a splice site mutation (2789+5 G→A) associated with small amounts of normal CFTRmRNA and mild cystic fibrosis. *Hum. Mutat.* 9, 332–338. doi:10.1002/(SICI)1098-1004(1997)9:4<332::AID-HUMU5>3.0.CO;2-7.
- Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., et al. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9362–9367. doi:10.1073/pnas.0903103106.
- Hofmann, W., Horn, D., Hüttner, C., Classen, E., and Scherneck, S. (2003). The BRCA2 variant 8204G→A is a splicing mutation and results in an in frame deletion of the gene. *J. Med. Genet.* 40, e23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12624152> [Accessed March 4, 2019].
- Horn, S., Figl, A., Rachakonda, P. S., Fischer, C., Sucker, A., Gast, A., et al. (2013). TERT Promoter Mutations in Familial and Sporadic Melanoma. *Science* (80-. ). 339, 959–961. doi:10.1126/science.1230062.
- Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., et al. (2012). Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum. Mutat.* 33, 1228–38. doi:10.1002/humu.22101.
- Howlett, N. G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., De Die-Smulders, C., et al. (2002). Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* 297, 606–9. doi:10.1126/science.1073834.
- Huang, R. C., Maheshwari, N., and Bonner, J. (1960). Enzymatic synthesis of RNA. *Biochem. Biophys. Res. Commun.* 3, 689–94. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/13716400> [Accessed October 18, 2017].
- Hurwitz, J., Furth, J. J., Anders, M., Ortiz, P. J., and August, J. T. (1961). The enzymatic incorporation of ribonucleotides into RNA and the role of DNA. *Cold Spring Harb. Symp. Quant. Biol.* 26, 91–100.

Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14450184> [Accessed October 18, 2017].

- Jenne, D. E., Reimann, H., Nezu, J., Friedel, W., Loff, S., Jeschke, R., et al. (1998). Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat. Genet.* 18, 38–43. doi:10.1038/ng0198-38.
- Jung, H., Lee, D., Lee, J., Park, D., Kim, Y. J., Park, W.-Y., et al. (2015). Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* 47, 1242–1248. doi:10.1038/ng.3414.
- Kauffmann, A., Rosselli, F., Lazar, V., Winnepeninckx, V., Mansuet-Lupo, A., Dessen, P., et al. (2008). High expression of DNA repair pathways is associated with metastasis in melanoma patients. *Oncogene* 27, 565–573. doi:10.1038/sj.onc.1210700.
- Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. *Nat. Rev. Mol. Cell Biol.* 13, 700–12. doi:10.1038/nrm3454.
- Khamis, A., Palmen, J., Lench, N., Taylor, A., Badmus, E., Leigh, S., et al. (2015). Functional analysis of four LDLR 5'UTR and promoter variants in patients with familial hypercholesterolaemia. *Eur. J. Hum. Genet.* 23, 790–795. doi:10.1038/ejhg.2014.199.
- Kinzler, K. W., and Vogelstein, B. (1997). Gatekeepers and caretakers. *Nature* 386, 761–763. doi:10.1038/386761a0.
- Klein, H. L. (2008). The consequences of Rad51 overexpression for normal and tumor cells. *DNA Repair (Amst)*. 7, 686–693. doi:10.1016/J.DNAREP.2007.12.008.
- Knudson, A. G., and Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* 68, 820–3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/5279523> [Accessed May 7, 2018].
- Kralovicova, J., Hwang, G., Asplund, a. C., Churbanov, A., Smith, C. I. E., and Vorechovsky, I. (2011). Compensatory signals associated with the activation of human GC 5' splice sites. *Nucleic Acids Res.* 39, 7077–7091. doi:10.1093/nar/gkr306.
- Krush, A. J. (1979). Contributions of Pierre Paul Broca to Cancer Genetics HISTORY OF SCIENCE CONTRIBUTIONS OF PIERRE PAUL BROCA TO CANCER GENETICS. Available at: <http://digitalcommons.unl.edu/tnas> [Accessed April 13, 2018].
- Kurian, A. W., Sigal, B. M., and Plevritis, S. K. (2010). Survival Analysis of Cancer Risk Reduction Strategies for *BRCA1/2* Mutation Carriers. *J. Clin. Oncol.* 28, 222–231. doi:10.1200/JCO.2009.22.7991.
- Kuznetsov, S. G., Liu, P., and Sharan, S. K. (2008). Mouse embryonic stem cell-based functional assay to evaluate mutations in *BRCA2*. *Nat. Med.* 14, 875–81. doi:10.1038/nm.1719.
- Kwong, A., Wong, L. P., Chan, K. Y. K., Ma, E. S. K., Khoo, U. S., and Ford, J. M. (2008). Characterization of the pathogenic mechanism of a novel *BRCA2* variant in a Chinese family. *Fam. Cancer* 7, 125–133. doi:10.1007/s10689-007-9155-7.
- Larsen, M. J., Thomassen, M., Gerdes, A.-M., and Kruse, T. A. (2014). Hereditary breast cancer: clinical, pathological and molecular characteristics. *Breast Cancer (Auckl)*. 8, 145–55. doi:10.4137/BCBCR.S18715.
- Lee, Y., and Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* 84, 291–323. doi:10.1146/annurev-biochem-060614-034316.
- Levin, A. A. (2019). Treating Disease at the RNA Level with Oligonucleotides. *N. Engl. J. Med.* 380, 57–

70. doi:10.1056/nejmra1705346.

- Levine, A., and Durbin, R. (2001). A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res.* 29, 4006–13. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11574683> [Accessed June 5, 2018].
- Li, F. P., Fraumeni, J. F. J., Mulvihill, J. J., Blattner, W. A., Dreyfus, M. G., Tucker, M. A., et al. (1988). A cancer family syndrome in twenty-four kindreds. *Cancer Res.* 48, 5358–5362.
- Li, J., Zou, C., Bai, Y., Wazer, D. E., Band, V., and Gao, Q. (2006). DSS1 is required for the stability of BRCA2. *Oncogene* 25, 1186–1194. doi:10.1038/sj.onc.1209153.
- Liaw, D., Marsh, D. J., Li, J., Dahia, P. L. M., Wang, S. I., Zheng, Z., et al. (1997). Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat. Genet.* 16, 64. Available at: <http://dx.doi.org/10.1038/ng0597-64>.
- Lim, K. R. Q., Maruyama, R., and Yokota, T. (2017). Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug Des. Devel. Ther.* 11, 533–545. doi:10.2147/DDDT.S97635.
- Liu, L., Fang, Y., Fan, J. L., Hu, J. M., Xu, X. T., Jin, X. H., et al. (2014). BRCA2 promoter polymorphism is associated with breast cancer prognosis in Chinese women. *Chin. Med. J. (Engl)*. 127, 2012–2015. doi:10.3760/cma.j.issn.0366-6999.20130978.
- Liu, Q., Thompson, B. A., Ward, R. L., Hesson, L. B., and Sloane, M. A. (2016). Understanding the Pathogenicity of Noncoding Mismatch Repair Gene Promoter Variants in Lynch Syndrome. *Hum. Mutat.* 37, 417–426. doi:10.1002/humu.22971.
- Lloyd, K. M. (1963). Cowden's Disease. *Ann. Intern. Med.* 58, 136. doi:10.7326/0003-4819-58-1-136.
- Lo, T., Pellegrini, L., Venkitaraman, A. R., and Blundell, T. L. (2003). Sequence fingerprints in BRCA2 and RAD51: implications for DNA repair and cancer. *DNA Repair (Amst)*. 2, 1015–1028. doi:10.1016/S1568-7864(03)00097-1.
- Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigo, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 579, 1900–1903.
- López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 579, 1900–3. doi:10.1016/j.febslet.2005.02.047.
- Lynch, H. T., and Krush, A. J. (1971). Carcinoma of the breast and ovary in three families. *Surg. Gynecol. Obstet.* 133, 644–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/4328854> [Accessed April 12, 2018].
- Lynch, H. T., Snyder, C., and Casey, M. J. (2013). Hereditary ovarian and breast cancer: What have we learned. *Ann. Oncol.* 24, viii83–viii95. doi:10.1093/annonc/mdt313.
- Maia, A.-T., Antoniou, A. C., O'Reilly, M., Samarajiwa, S., Dunning, M., Kartsonaki, C., et al. (2012). Effects of BRCA2 cis-regulation in normal breast and cancer risk amongst BRCA2 mutation carriers. *Breast Cancer Res.* 14, R63. doi:10.1186/bcr3169.
- Malkin, D., Li, F., Strong, L., Fraumeni, J., Nelson, C., Kim, D., et al. (1990). Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science (80-. )*. 250, 1233–1238. doi:10.1126/science.1978757.
- Manning, K. S., and Cooper, T. A. (2017). The roles of RNA processing in translating genotype to phenotype. *Nat. Rev. Mol. Cell Biol.* 18, 102–114. doi:10.1038/nrm.2016.139.

- Mansfield, S. G., Chao, H., and Walsh, C. E. (2004). RNA repair using spliceosome-mediated RNA trans-splicing. *Trends Mol. Med.* 10, 263–268. doi:10.1016/J.MOLMED.2004.04.007.
- Mau, C., and Untch, M. (2017). Prophylactic Surgery: For Whom, When and How? *Breast Care* 12, 379–384. doi:10.1159/000485830.
- McGuire, A. M., Pearson, M. D., Neafsey, D. E., and Galagan, J. E. (2008). Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol.* 9, R50. doi:10.1186/gb-2008-9-3-r50.
- Menzel, T., Nähse-Kumpf, V., Kousholt, A. N., Klein, D. K., Lund-Andersen, C., Lees, M., et al. (2011). A genetic screen identifies BRCA2 and PALB2 as key regulators of G2 checkpoint maintenance. *EMBO Rep.* 12, 705–712. doi:10.1038/embor.2011.99.
- Mesman, R. L. S., Calléja, F. M. G. R., Hendriks, G., Morolli, B., Misovic, B., Devilee, P., et al. (2018). The functional impact of variants of uncertain significance in BRCA2. *Genet. Med.* 0, 1–10. doi:10.1038/s41436-018-0052-2.
- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266, 66–71.
- Milner, J., Ponder, B., Hughes-Davies, L., Seltsmann, M., and Kouzarides, T. (1997). Transcriptional activation functions in BRCA2. *Nature* 386, 772–3. doi:10.1038/386772a0.
- Montalban, G., Fraile-Bethencourt, E., López-Perolio, I., Pérez-Segura, P., Infante, M., Durán, M., et al. (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.*, 1–6. doi:10.1002/humu.23583.
- Mucaki, E. J., Caminsky, N. G., Perri, A. M., Lu, R., Laederach, A., Halvorsen, M., et al. (2016). A unified analytic framework for prioritization of non-coding variants of uncertain significance in heritable breast and ovarian cancer. *BMC Med. Genomics* 9, 19. doi:10.1186/s12920-016-0178-5.
- Muller, D., Rouleau, E., Schultz, I., Caputo, S., Lefol, C., Bièche, I., et al. (2011). An entire exon 3 germline rearrangement in the BRCA2 gene: pathogenic relevance of exon 3 deletion in breast cancer predisposition. *BMC Med. Genet.* 12, 121. doi:10.1186/1471-2350-12-121.
- Narod, S. A., Feunteun, J., Lynch, H. T., Watson, P., Conway, T., Lynch, J., et al. (1991). Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet (London, England)* 338, 82–3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1676470> [Accessed April 12, 2018].
- Narod, S. a, and Foulkes, W. D. (2004). BRCA1 and BRCA2: 1994 and beyond. *Nat. Rev. Cancer* 4, 665–76. doi:10.1038/nrc1431.
- Nelen, M. R., van Staveren, W. C. G., Peeters, E. A. J., Ben Hassel, M., Gorlin, R. J., Hamm, H., et al. (1997). Germline Mutations in the PTEN/MMAC1 Gene in Patients With Cowden Disease. *Hum. Mol. Genet.* 6, 1383–1387. Available at: <http://dx.doi.org/10.1093/hmg/6.8.1383>.
- Nielsen, F. C., van Overeem Hansen, T., and Sørensen, C. S. (2016). Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat. Rev. Cancer* 16, 599–612. doi:10.1038/nrc.2016.72.
- Osheim, Y. N., Miller, O. L., and Beyer, A. L. (1985). RNP particles at splice junction sequences on Drosophila chorion transcripts. *Cell* 43, 143–51. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3935315> [Accessed May 30, 2018].
- Ottesen, E. W. (2017). ISS-N1 makes the first FDA-approved drug for spinal muscular atrophy. *Transl. Neurosci.* 8, 1–6. doi:10.1515/tnsci-2017-0001.

- Padgett, R. a (2012). New connections between splicing and human disease. *Trends Genet.* 28, 147–54. doi:10.1016/j.tig.2012.01.001.
- Parsons, M. T., Whiley, P. J., Beesley, J., Drost, M., De Wind, N., Thompson, B. A., et al. (2015). Consequences of germline variation disrupting the constitutional translational initiation codon start sites of MLH1 and BRCA2: use of potential alternative start sites and implications for predicting variant pathogenicity HHS Public Access. *Mol Carcinog* 54, 513–522. doi:10.1002/mc.22116.
- Pellagatti, A., Armstrong, R. N., Steeples, V., Sharma, E., Repapi, E., Singh, S., et al. (2018). Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. *Blood*, blood-2018-04-843771. doi:10.1182/blood-2018-04-843771.
- Pellegrini, L., Yu, D. S., Lo, T., Anand, S., Lee, M., Blundell, T. L., et al. (2002). Insights into DNA recombination from the structure of a RAD51–BRCA2 complex. *Nature* 420, 287–293. doi:10.1038/nature01230.
- Pensabene, M., Spagnoletti, I., Capuano, I., Condello, C., Pepe, S., Contegiacomo, a, et al. (2009). Two mutations of BRCA2 gene at exon and splicing site in a woman who underwent oncogenetic counseling. *Ann. Oncol.* 20, 874–8. doi:10.1093/annonc/mdn724.
- Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747–752. doi:10.1038/35021093.
- Petrucelli, N., Daly, M. B., and Feldman, G. L. (2016). BRCA1 and BRCA2 Hereditary Breast and Ovarian Cancer. *GeneReviews® [Internet]* 00, 1–11. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20301425> [Accessed April 13, 2018].
- Pickrell, J. K., Pai, A. A., Gilad, Y., and Pritchard, J. K. (2010). Noisy Splicing Drives mRNA Isoform Diversity in Human Cells. *PLoS Genet.* 6, e1001236. doi:10.1371/journal.pgen.1001236.
- Pilié, P. G., Tang, C., Mills, G. B., and Yap, T. A. (2019). State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat. Rev. Clin. Oncol.* 16, 81–104. doi:10.1038/s41571-018-0114-z.
- Pinotti, M., Rizzotto, L., Balestra, D., Lewandowska, M. A., Cavallari, N., Marchetti, G., et al. (2008). U1-snRNA-mediated rescue of mRNA processing in severe factor VII deficiency. *Blood* 111, 2681–4. doi:10.1182/blood-2007-10-117440.
- Plon, S. E., Eccles, D. M., Easton, D., Foulkes, W. D., Genuardi, M., Greenblatt, M. S., et al. (2008). Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum. Mutat.* 29, 1282–1291. doi:10.1002/humu.20880.
- Quinet, A., Lemaçon, D., and Vindigni, A. (2017). Replication Fork Reversal: Players and Guardians. *Mol. Cell* 68, 830–833. doi:10.1016/j.molcel.2017.11.022.
- Radice, P., De Summa, S., Caleca, L., and Tommasi, S. (2011). Unclassified variants in BRCA genes: guidelines for interpretation. *Ann. Oncol.* 22 Suppl 1, i18–23. doi:10.1093/annonc/mdq661.
- Rebbeck, T. R., Friebel, T. M., Friedman, E., Hamann, U., Huo, D., Kwong, A., et al. (2018). Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations. *Hum. Mutat.* 39, 593–620. doi:10.1002/humu.23406.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (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, 405–423. doi:10.1038/gim.2015.30.

- Rizzo, J. M., and Buck, M. J. (2012). Key principles and clinical applications of “next-generation” DNA sequencing. *Cancer Prev. Res. (Phila)*. 5, 887–900. doi:10.1158/1940-6207.CAPR-11-0432.
- Roy, R., Chun, J., and Powell, S. N. (2012). BRCA1 and BRCA2 : different roles in a common pathway of genome protection. Available at: <http://dx.doi.org/10.1038/nrc3181>.
- Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 16, 129–143. doi:10.1038/nrm3952.
- Salton, M., Kasprzak, W. K., Voss, T., Shapiro, B. A., Poulikakos, P. I., and Misteli, T. (2015). Inhibition of vemurafenib-resistant melanoma by interference with pre-mRNA splicing. *Nat. Commun.* 6, 7103. doi:10.1038/ncomms8103.
- Sanz, D. J., Acedo, A., Infante, M., Durán, M., Pérez-Cabornero, L., Esteban-Cardenosa, E., et al. (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. doi:10.1158/1078-0432.CCR-09-2564.
- Sauna, Z. E., and Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* 12, 683–91. doi:10.1038/nrg3051.
- Scalet, D., Maestri, I., Branchini, A., Bernardi, F., Pinotti, M., and Balestra, D. (2019). Disease-causing variants of the conserved +2T of 5' splice sites can be rescued by engineered U1snRNAs. *Hum. Mutat.* 40, 48–52. doi:10.1002/humu.23680.
- Seizl, M., Hartmann, H., Hoeg, F., Kurth, F., Martin, D. E., Söding, J., et al. (2011). A Conserved GA Element in TATA-Less RNA Polymerase II Promoters. *PLoS One* 6, e27595. doi:10.1371/journal.pone.0027595.
- Shahid, T., Soroka, J., Kong, E. H., Malivert, L., McIlwraith, M. J., Pape, T., et al. (2014). Structure and mechanism of action of the BRCA2 breast cancer tumor suppressor. *Nat. Struct. Mol. Biol.* doi:10.1038/nsmb.2899.
- Sharma, N., Sosnay, P. R., Ramalho, A. S., Douville, C., Franca, A., Gottschalk, L. B., et al. (2014). Experimental Assessment of Splicing Variants Using Expression Minigenes and Comparison with In Silico Predictions. *Hum. Mutat.* 35, 1249–1259. doi:10.1002/humu.22624.
- Singh, N. N., Del Rio-Malewski, J. B., Luo, D., Ottesen, E. W., Howell, M. D., and Singh, R. N. (2017). Activation of a cryptic 5' splice site reverses the impact of pathogenic splice site mutations in the spinal muscular atrophy gene. *Nucleic Acids Res.* 45, 12214–12240. doi:10.1093/nar/gkx824.
- Skordis, L. A., Dunckley, M. G., Yue, B., Eperon, I. C., and Muntoni, F. (2003). Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc. Natl. Acad. Sci.* 100, 4114–4119. doi:10.1073/pnas.0633863100.
- Slavin, T. P., Maxwell, K. N., Lilyquist, J., Vijai, J., Neuhausen, S. L., Hart, S. N., et al. (2017). The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. *npj Breast Cancer* 3, 22. doi:10.1038/s41523-017-0024-8.
- Spurdle, A. B., Couch, F. J., Hogervorst, F. B. L., Radice, P., and Sinilnikova, O. M. (2008). Prediction and assessment of splicing alterations: implications for clinical testing. *Hum. Mutat.* 29, 1304–13. doi:10.1002/humu.20901.
- Spurdle, A. B., Healey, S., Devereau, A., Hogervorst, F. B. L., Monteiro, A. N. A., Nathanson, K. L., et al.

- (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. doi:10.1002/humu.21628.
- Stein, C. A., and Castanotto, D. (2017). FDA-Approved Oligonucleotide Therapies in 2017. *Mol. Ther.* 25, 1069–1075. doi:10.1016/j.ymthe.2017.03.023.
- Stevens, A. (1960). Incorporation of the adenine ribonucleotide into RNA by cell fractions from *E. coli* B. *Biochem. Biophys. Res. Commun.* 3, 92–96. doi:10.1016/0006-291X(60)90110-8.
- Suñé-Pou, M., Prieto-Sánchez, S., Boyero-Corral, S., Moreno-Castro, C., Yousfi, Y. El, Suñé-Negre, J. M., et al. (2017). Targeting splicing in the treatment of human disease. *Genes (Basel)*. 8. doi:10.3390/genes8030087.
- Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T., and Lehner, B. (2014). Synonymous mutations frequently act as driver mutations in human cancers. *Cell* 156, 1324–35. doi:10.1016/j.cell.2014.01.051.
- Tai, Y. C., Domchek, S., Parmigiani, G., and Chen, S. (2007). Breast Cancer Risk Among Male BRCA1 and BRCA2 Mutation Carriers. *JNCI J. Natl. Cancer Inst.* 99, 1811–1814. doi:10.1093/jnci/djm203.
- Tavtigian, S. V., Simard, J., Rommens, J., Couch, F., Shattuck-Eidens, D., Neuhausen, S., et al. (1996). The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nat. Genet.* 12, 333–337. doi:10.1038/ng0396-333.
- Théry, J. C., Krieger, S., Gaildrat, P., Révillion, F., Buisine, M.-P., Killian, A., et al. (2011). Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. *Eur. J. Hum. Genet.* 19, 1052–8. doi:10.1038/ejhg.2011.100.
- Vallée, M. P., Di Sera, T. L., Nix, D. A., Paquette, A. M., Parsons, M. T., Bell, R., et al. (2016). Adding In Silico Assessment of Potential Splice Aberration to the Integrated Evaluation of *BRCA* Gene Unclassified Variants. *Hum. Mutat.* 37, 627–639. doi:10.1002/humu.22973.
- van der Klift, H. M., Jansen, A. M. L., van der Steenstraten, N., Bik, E. C., Tops, C. M. J., Devilee, P., et al. (2015). Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol. Genet. Genomic Med.* 3, 327–345. doi:10.1002/mgg3.145.
- Vaz-Draco, R., Custódio, N., and Carmo-Fonseca, M. (2017). Deep intronic mutations and human disease. *Hum. Genet.* 136, 1093–1111. doi:10.1007/s00439-017-1809-4.
- Villate, O., Ibarluzea, N., Fraile-Bethencourt, E., Valenzuela, A., Velasco, E. A., Grozeva, D., et al. (2018). Functional Analyses of a Novel Splice Variant in the CHD7 Gene, Found by Next Generation Sequencing, Confirm Its Pathogenicity in a Spanish Patient and Diagnose Him with CHARGE Syndrome. *Front. Genet.* 9, 26–31. doi:10.3389/fgene.2018.00007.
- Vreeswijk, M. P. G., Kraan, J. N., van der Klift, H. M., Vink, G. R., Cornelisse, C. J., Wijnen, J. T., et al. (2009). Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Hum. Mutat.* 30, 107–14. doi:10.1002/humu.20811.
- Wagner, S. D., and Berglund, J. A. (2014). "Alternative Pre-mRNA Splicing," in (Humana Press, Totowa, NJ), 45–54. doi:10.1007/978-1-62703-980-2\_4.
- Walker, L. C., Whiley, P. J., Couch, F. J., Farrugia, D. J., Healey, S., Eccles, D. M., et al. (2010). Detection of splicing aberrations caused by BRCA1 and BRCA2 sequence variants encoding missense substitutions: implications for prediction of pathogenicity. *Hum. Mutat.* 31, E1484–505.

doi:10.1002/humu.21267.

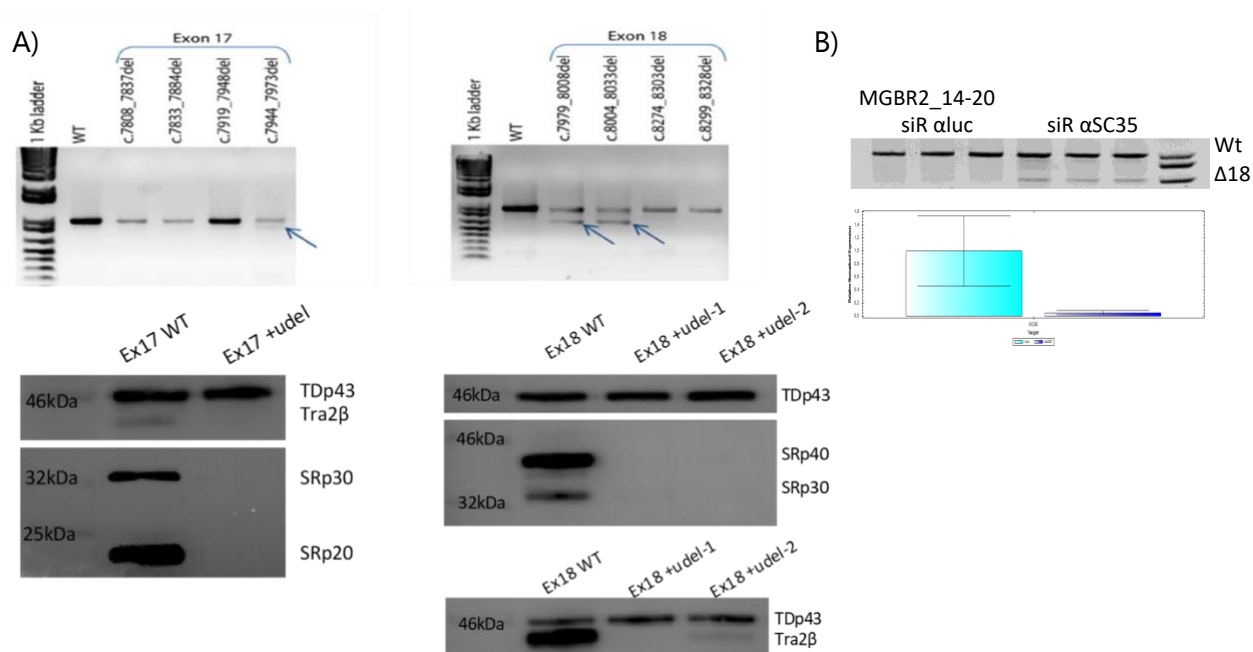
- Wan, J., Sazani, P., and Kole, R. (2009). Modification of HER2 pre-mRNA alternative splicing and its effects on breast cancer cells. *Int. J. Cancer* 124, 772–777. doi:10.1002/ijc.24052.
- Wang, G.-S., and Cooper, T. a (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* 8, 749–61. doi:10.1038/nrg2164.
- Wang, J., Zhang, J., Li, K., Zhao, W., and Cui, Q. (2012). SpliceDisease database: linking RNA splicing and disease. *Nucleic Acids Res.* 40, D1055–9. doi:10.1093/nar/gkr1171.
- Wang, X., Andreassen, P. R., and D'Andrea, A. D. (2004). Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol. Cell. Biol.* 24, 5850–62. doi:10.1128/MCB.24.13.5850-5862.2004.
- Weiss, S. B., and Gladstone, L. (1959). A MAMMALIAN SYSTEM FOR THE INCORPORATION OF CYTIDINE TRIPHOSPHATE INTO RIBONUCLEIC ACID <sup>1</sup>. *J. Am. Chem. Soc.* 81, 4118–4119. doi:10.1021/ja01524a087.
- Whiley, P. J., Guidugli, L., Walker, L. C., Healey, S., Thompson, B. a, Lakhani, S. R., et al. (2011). Splicing and multifactorial analysis of intronic BRCA1 and BRCA2 sequence variants identifies clinically significant splicing aberrations up to 12 nucleotides from the intron/exon boundary. *Hum. Mutat.* 32, 678–87. doi:10.1002/humu.21495.
- Will, C. L., and Lührmann, R. (2011). Spliceosome structure and function. TL - 3. *Cold Spring Harb. Perspect. Biol.* 3 VN-re, 1–23. doi:10.1101/cshperspect.a003707.
- Wimmer, K., Roca, X., Beiglböck, H., Callens, T., Etzler, J., Rao, A. R., et al. (2007). Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5' splice-site disruption. *Hum. Mutat.* 28, 599–612. doi:10.1002/humu.20493.
- Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., et al. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378, 789–792. doi:10.1038/378789a0.
- Wu, K., Hinson, S. R., Ohashi, A., Farrugia, D., Wendt, P., Tavtigian, S. V, et al. (2005). Functional Evaluation and Cancer Risk Assessment of BRCA2 Unclassified Variants. 417–426.
- Xia, B., Sheng, Q., Nakanishi, K., Ohashi, A., Wu, J., Christ, N., et al. (2006a). Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol. Cell* 22, 719–29. doi:10.1016/j.molcel.2006.05.022.
- Xia, B., Sheng, Q., Nakanishi, K., Ohashi, A., Wu, J., Christ, N., et al. (2006b). Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2. *Mol. Cell* 22, 719–729. doi:10.1016/j.molcel.2006.05.022.
- Yang, C., Bolotin, E., Jiang, T., Sladek, F. M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 389, 52–65. doi:10.1016/j.gene.2006.09.029.
- Yang, H., Jeffrey, P. D., Miller, J., Kinnucan, E., Sun, Y., Thoma, N. H., et al. (2002). BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* 297, 1837–48. doi:10.1126/science.297.5588.1837.
- Zhang, L., Bacaes, R., Boyar, S., Hudis, C., Nafa, K., and Offit, K. (2009). cDNA analysis demonstrates that the BRCA2 intronic variant IVS4-12del5 is a deleterious mutation. *Mutat. Res.* 663, 84–9. doi:10.1016/j.mrfmmm.2008.11.010.

Zhang, M. (1998). Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.* 7, 919–932. doi:10.1093/hmg/7.5.919.

**ANEXOS**

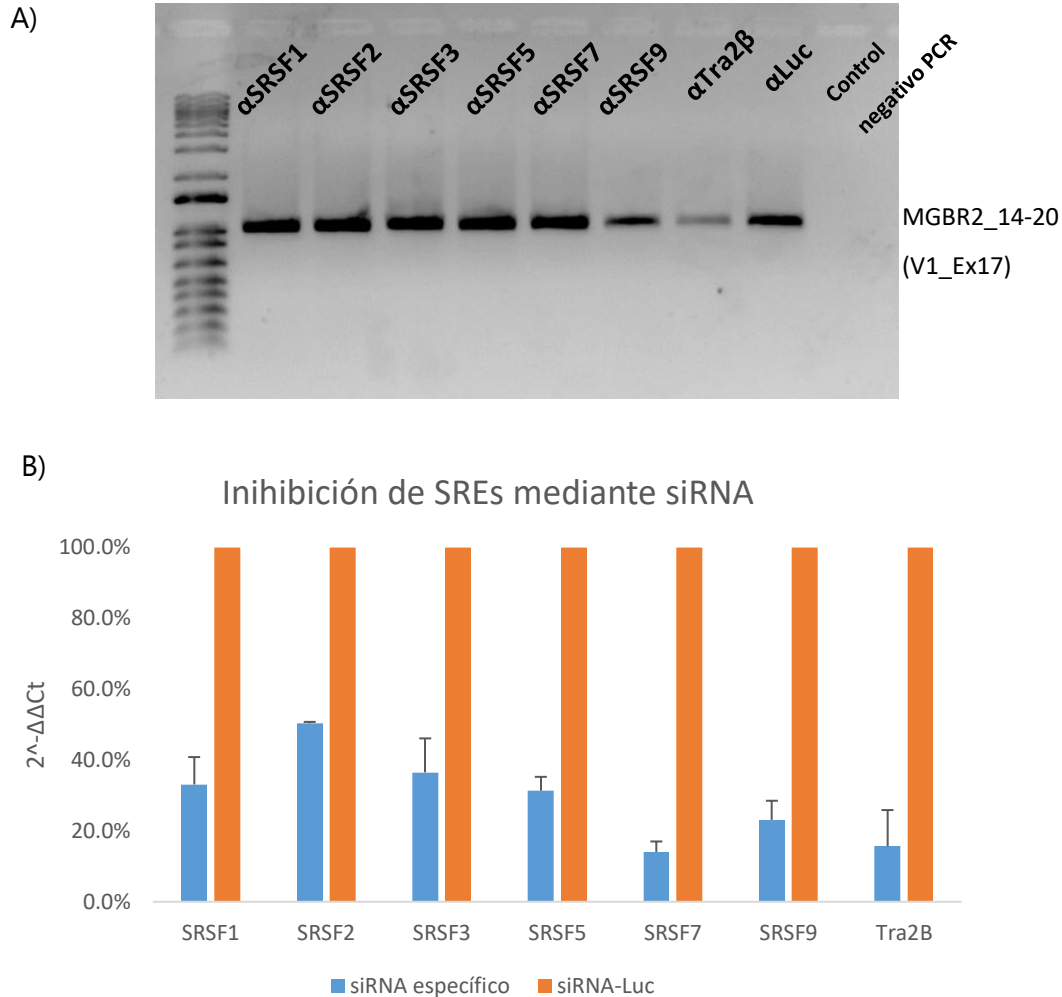


## ANEXOS



**FIGURA SUPLEMENTARIA 1.** A) Resultado del ensayo de Pull-Down sobre las microdeleciones positivas de los exones 17 y 18. En la parte superior aparecen los geles de agarosa donde se muestra el *skipping* de los exones en cuestión como consecuencia de la microdelección. En la parte inferior aparecen los resultados de Western-blot con los anticuerpo anti-SR-*family* y anti-Tra2β; se utilizó anti-TDp43 como control del experimento. B) Silenciamiento de SC35 (SRSF2) mediante siRNA y comprobación por qPCR. En la parte superior aparece en un gel de agarosa el transcrito resultante del minigen MGBR2\_14-20 en condiciones normales (siR αluc) y bajo el silenciamiento de SC35 (SRSF2) (siR αSC35). En la parte inferior aparece el resultado de la qPCR que muestra la reducción de la expresión de SC35 (SRSF2).





**FIGURA SUPLEMENTARIA 2.** Efecto de la inhibición de las principales SRp y Tra2β. A) Resultado de la amplificación del transcrito producido por el MGBR2\_14-20 en células MCF-7 transfectadas con los correspondientes siRNA. El RNA fue purificado y convertido en cDNA como previamente se ha explicado. La amplificación se realizó con los *primers* pMAD\_607FW y RTBR2\_Ex17RV2. El símbolo α seguido del nombre del *enhancer* de *splicing* indica el siRNA utilizado para la transfección. Como control negativo se utilizó un siRNA dirigido al gen de la luciferasa (Luc). Como patrón de tamaño se utilizó 1 Kb Plus DNA Ladder (Thermo Fisher Scientific, MA, EEUU). B) Cuantificación por qPCR de la inhibición de los *enhancers* de *splicing* mediante siRNA. La amplificación se realizó con 10ng de cDNA utilizando SG qPCR Master Mix (Eurx, Gdansk, Polonia) y siguiendo las indicaciones del fabricante. Como control interno se utilizó GAPDH.



**TABLA ANEXA 1. Secuencias de los oligonucleótidos de mutagénesis dirigida para las microdeleciones en los exones 14 y 15 de *BRCA2*.**

| <b>Microdeleciones Exón 14 <i>BRCA2</i> (5'→3')</b> |   |
|---|---|
| c.7010_7039del                                      | TATATATTTTCTCCCCATTGCAGCACAAATTTTACCGCACCTGGTCAAGA            |
|   | TCTTGACCAGGTGCGGTAAAATTTGTGCTGCAATGGGGAGAAAATATATA            |
| c.7035_7064del                                      | AACTAAGGAACGTCAAGAGATACAATTTCTGTCTAAATCTCATTTGTATG            |
|   | CATACAAATGAGATTAGACAGAAATTGTATCTCTTGACGTTCCCTTAGTT            |
| c.7378_7407del                                      | AATGACAATGAGATTCATCAGTTTAACAAATTCACAAAGTGTGAAGAAGAACCTTTAGGT  |
|   | ACCTAAAGGTTCTTCTTCACACTTTGTGAATTTGTTAAACTGATGAATCTCATTGTCATT  |
| c.7402_7432del                                      | AACAAAAACAACCTCCAATCAAGCAGCAGCTTAGGTATTGTATGACAATTTGTGTGATGAA |
|   | TTCATCACACAAATTGTCATACAATACCTAAGCTGCTGCTTGATTGGAGTTGTTTTGTT   |
| <b>Microdeleciones Exón 15 <i>BRCA2</i> (5'→3')</b> |   |
| c.7438_7467del                                      | GCTAAGTATTTATTCTTTGATAGATATACAGGATATGCGAATTAAGAAGA            |
|   | TCTTCTTAATTCGCATATCCTGTATATCTATCAAAGAATAAATACTTAGC            |
| c.7463_7492del                                      | TAGATTTAATTACAAGTCTTCAGAATGCCAAACAAAGGCAACGCGTCTTCCACAGCCAG   |
|   | CTGGCTGTGGAAAGACGCGTTGCCTTTGTTTGGCATTCTGAAGACTTGAATTAATCTA    |
| c.7561_7590del                                      | TGCAAAAACATCCACTCTGCCTCGAGTTCCTCTGCGTGTTCTCATAAAC             |
|   | GTTTATGAGAACACGCAGAGGGAACTCGAGGCAGAGTGGATGTTTTTGCA            |
| c.7586_7615del                                      | ATCTCTCTGAAAGCAGCAGTAGGAGAGGTATGTGTTGTCTACAATACTG             |
|   | CAGTATTGTAGACAAACACATACCTCTCCTACTGCTGCTTTCAGAGAGAT            |



**TABLA ANEXA 2. Secuencias de siRNA para el silenciamiento de proteínas SR y Tra2 $\beta$ ; y secuencias de los oligonucleótidos empleados para comprobar dicho silenciamiento mediante qPCR.**

| <b>Secuencias de siRNA</b>                      |                                |
|---|--------------------------------|
| SRSF1 (SF2/ASF)                                 | 5' ACGAUUGCCGCAUCUACGU 3'      |
|   | 5' ACGUAGAUGCGGCAAUCGU 3'      |
| SRSF2 (SC35)                                    | 5' AAUCCAGGUCGCGAUCGAA 3'      |
|   | 5' UUCGAUCGCGACCUGGAUU 3'      |
| SRSF3 (SRp20)                                   | 5' GAGUGGAACUGUCGAAUGG 3'      |
|   | 5' CCAUUCGACAGUCCACUC 3'       |
| SRSF5 (SRp40)                                   | 5' CCACCUGUAAGAACAGAAA 3'      |
|   | 5' UUUCUGUUCUACAGGUGG 3'       |
| SRSF7 (9G8)                                     | 5' GGAGAGUUAGAAAGGGCUU 3'      |
|   | 5' AAGCCCUUUCUAAUCUCUC 3'      |
| SRSF9 (SRp30c)                                  | 5' GGAUAUGCCUGCGUAAA 3'        |
|   | 5' UUUACGCAGGGCAUAUUC 3'       |
| Tra2 $\beta$                                    | 5' GGAGGAUACAGAUACGUU 3'       |
|   | 5' AACGUGAUCUGUAUCCUC 3'       |
| Luciferasa                                      | 5' UAAGGCUAUGAAGAGAUAC 3'      |
|   | 5' GUAUCUCUUAUAGCCUUA 3'       |
| <b>Secuencias de oligonucleótidos para qPCR</b> |                                |
| SRSF1 (SF2)                                     | 5' CATCGACCTCAAGAATCGCC 3'     |
|   | 5' CCCATCGTAATCATAGCCGT 3'     |
| SRSF2 (SC35)                                    | 5' TTCGCCTTCGTTGCTTTC 3'       |
|   | 5' CCGTAGCGCGCCATTTGCA 3'      |
| SRSF3 (SRp20)                                   | 5' TGCTTTTGTTGAATTTGAAGA 3'    |
|   | 5' CATTGACAGTTCCACTCTT 3'      |
| SRSF5 (SRp40)                                   | 5' AATGATAGACGAAATGCTCC 3'     |
|   | 5' CCGCAAACGTTACTTCCCC 3'      |
| SRSF7 (9G8)                                     | 5' GGGCTTTCAGTTATTATGGT 3'     |
|   | 5' GGCAGGTGGTCTATCAAAA 3'      |
| SRSF9 (SRp30c)                                  | 5' CTGGGGATGTCTGTTATGCT 3'     |
|   | 5' TGAGAGCGGAATTTGGTGT 3'      |
| Tra2 $\beta$                                    | 5' TTGATGGGCGTAGGATCAGAGTTG 3' |
|   | 5' TCCTCTGCATAGTAATCCCGACG 3'  |
| GAPDH   | 5' TGCCAAATATGATGACATCAAGAA 3' |
|   | 5' GGAGTGGGTGTCGCTGTTG 3'      |



**TABLA ANEXA 3. Colección de 294 variantes comprendidas entre el intrón 13 y el intrón 15 de BRCA2 procedente de las principales bases de datos (ClinVar, UMD y BIC). Las variantes están anotadas siguiendo los criterios de HGVS.**

| Variantes en el exón 14 y regiones intrónicas flanqueantes |           |                  |           |           |
|--|-----------|------------------|-----------|-----------|
| c.7008-20_c.7008-17del4                                    | c.7066T>C | c.7175A>G        | c.7256T>G | c.7349A>T |
| c.7008-20A>G   | c.7067T>A | c.7177A>G        | c.7258G>T | c.7350T>C |
| c.7008-17T>G   | c.7069C>T | c.7178T>C        | c.7260A>G | c.7351G>A |
| c.7008-15delTT   | c.7071G>T | c.7180A>T        | c.7261C>G | c.7354A>G |
| c.7008-14delT  | c.7072T>C | c.7182A>G        | c.7261C>T | c.7355A>G |
| c.7008-13C>T   | c.7073C>G | c.7185C>T        | c.7266T>A | c.7356T>C |
| c.7008-10C>T   | c.7078T>C | c.7185C>A        | c.7270A>G | c.7359G>A |
| c.7008-5T>C  | c.7081C>T | c.7187T>A        | c.7273A>G | c.7360A>T |
| c.7008-3C>G  | c.7082A>G | c.7187T>C        | c.7278T>A | c.7360A>G |
| c.7008-2A>T  | c.7082A>T | c.7186_7187delTT | c.7279A>G | c.7366C>T |
| c.7008-1G>A  | c.7086G>A | c.7188G>T        | c.7282T>C | c.7366C>G |
| c.7009A>G  | c.7088A>G | c.7188G>A        | c.7283T>A | c.7370T>G |
| c.7010C>T  | c.7090G>T | c.7191T>A        | c.7287G>A | c.7373A>G |
| c.7011A>G  | c.7090G>A | c.7193C>G        | c.7294A>G | c.7375A>T |
| c.7012A>G  | c.7093C>T | c.7195A>G        | c.7296A>G | c.7376A>G |
| c.7015A>G  | c.7095T>A | c.7196C>T        | c.7299A>G | c.7394C>T |
| c.7017G>C  | c.7096C>G | c.7203A>G        | c.7301A>C | c.7397C>T |
| c.7017G>T  | c.7098G>A | c.7205C>T        | c.7303C>T | c.7398A>G |
| c.7021C>T  | c.7102T>C | c.7207A>G        | c.7304A>G | c.7402G>A |
| c.7022G>T  | c.7102T>G | c.7208C>T        | c.7307A>G | c.7404A>C |
| c.7022G>A  | c.7104G>A | c.7210A>T        | c.7309A>G | c.7409T>C |
| c.7024C>T  | c.7107A>C | c.7215C>A        | c.7311T>A | c.7410C>G |
| c.7030A>G  | c.7115C>A | c.7218T>G        | c.7313A>G | c.7411A>G |

|            |           |           |           |              |
|------------|-----------|-----------|-----------|--------------|
| c.7034A>G  | c.7115C>G | c.7219G>C | c.7314T>A | c.7413A>G    |
| c.7037A>G  | c.7118G>C | c.7222C>T | c.7316G>C | c.7414A>G    |
| c.7039C>G  | c.7119C>G | c.7224A>G | c.7317A>G | c.7415A>C    |
| c.7040C>A  | c.7119C>T | c.7232A>C | c.7319A>C | c.7416G>T    |
| c.7042A>C  | c.7124T>G | c.7234A>G | c.7319A>G | c.7417T>C    |
| c.7045T>C  | c.7126G>C | c.7235C>T | c.7322G>C | c.7418G>A    |
| c.7047delT | c.7133C>G | c.7237A>G | c.7326T>C | c.7428A>G    |
| c.7049C>T  | c.7137A>G | c.7239A>G | c.7330G>C | c.7429C>A    |
| c.7050C>T  | c.7140T>A | c.7241C>T | c.7330G>T | c.7429C>T    |
| c.7051G>A  | c.7150C>A | c.7241C>G | c.7331A>G | c.7431T>C    |
| c.7052C>G  | c.7152A>G | c.7242A>G | c.7331A>T | c.7435+1G>C  |
| c.7053A>T  | c.7157C>A | c.7242A>T | c.7339A>G | c.7435+3A>G  |
| c.7055C>T  | c.7160C>T | c.7244A>G | c.7341T>C | c.7435+5T>C  |
| c.7057G>C  | c.7162A>G | c.7248T>C | c.7342A>C | c.7435+6G>A  |
| c.7060C>T  | c.7168A>G | c.7251C>T | c.7343A>G | c.7435+7T>G  |
| c.7063G>T  | c.7170T>G | c.7252A>G | c.7347T>G | c.7435+10G>A |
| c.7066T>G  | c.7172A>T | c.7253G>A | c.7347T>C |              |

| Variantes en el exón 15 y regiones intrónicas flanqueantes |                   |                 |                    |                     |
|--|-------------------|-----------------|--------------------|---------------------|
| c.7436-43T>G   | c.7469T>C         | c.7505G>C       | c.7556_7557insC    | c.7598C>G           |
| c.7436-22c>T   | c.7471C>T         | c.7506C>G       | c.7558C>T          | c.7600G>A           |
| c.7436-17T>G   | c.7471delC        | c.7506C>T       | c.7559G>A          | c.7601C>T           |
| c.7436-14T>G   | c.7472A>G         | c.7507G>A       | c.7559G>T          | c.7602G>C           |
| c.7436-10T>C   | c.7474_7475delGA  | c.7509C>G       | c.7561A>G          | c.7610A>G           |
| c.7436-4A>G  | c.7478T>G         | c.7512T>G       | c.7561delA         | c.7611_7615delTAAAC |
| c.7436-4A>T  | c.7480C>T         | c.7521A>T       | c.7562T>C          | c.7612A>T           |
| c.7436-2A>G  | c.7481G>A         | c.7522G>A       | c.7565_7568delCTCT | c.7617G>A           |
| c.7436-2A>T  | c.7484T>C         | c.7523G>T       | c.7565C>T          | c.7617+1G>A         |
| c.7436-2delAGAT  | c.7485dup         | c.7525_7526insA | c.7567_7568delCT   | c.7617+1G>T         |
| c.7436-1G>A  | c.7487A>C         | c.7527T>C       | c.7575A>G          | c.7617+2T>G         |
| c.7438T>G  | c.7491_7493delGAA | c.7528C>T       | c.7579delG         | c.7617+4T>C         |
| c.7443_7443delT  | c.7491G>A         | c.7529T>C       | c.7580_7582delTAG  | c.7617+14delTA      |
| c.7447A>G  | c.7492A>G         | c.7534C>T       | c.7583G>A          | c.7617+17A>G        |
| c.7448G>A  | c.7495C>T         | c.7537G>A       | c.7586G>A          | c.7617+19T>G        |
| c.7463G>A  | c.7499G>C         | c.7542_7549dup  | c.7593_7593delT    |                     |
| c.7464A>C  | c.7501C>T         | c.7543_7543delA | c.7595_7596insTT   |                     |
| c.7466A>G  | c.7503A>T         | c.7543_7544insA | c.7596_7609del     |                     |
| c.7467_7468insT  | c.7504C>T         | c.7544C>T       | c.7596C>T          |                     |
| c.7467T>C  | c.7505G>A         | c.7552C>G       | c.7597T>G          |                     |



**TABLA ANEXA 4. Secuencias de los oligonucleótidos utilizados para la mutagénesis dirigida de las variantes candidatas de alterar el *splicing* de los exones 14 y 15 de BRCA2.**

| <b>Variantes (HGVS)</b> | <b>Oligonucleótidos <i>forward/reverse</i> (5'→3')</b> | <b>Variantes (HGVS)</b> | <b>Oligonucleótidos <i>forward/reverse</i> (5'→3')</b> |
|-------------------------|--|-------------------------|--|
| c.7008-5T>C             | CTTATATATTTTCTCCCATCGCAGCACAACTAAGGAACGT               | c.7435+6G>A             | GAAGAAGAACCTTTAGGTATTATATGACAATTTGTGTGATGA             |
|                         | ACGTTCCCTAGTTGTGCTGCGATGGGGAGAAAATATATAAG              |                         | TCATCACACAAATTGTCATATAATACCTAAAGGTTCTTCTTC             |
| c.7008-3C>G             | TATATATTTTCTCCCATTTGGAGCACAACTAAGGAACGTCA              | c.7435+7T>G             | GAAGAACCTTTAGGTATTGGATGACAATTTGTGTGATGA                |
|                         | TGACGTTCCCTAGTTGTGCTCCAATGGGGAGAAAATATATA              |                         | TCATCACACAAATTGTCATCCAATACCTAAAGGTTCTTC                |
| c.7008-2A>T             | ATATATTTTCTCCCATTTGCTGCACAACTAAGGAACGTCAA              | c.7435+10G>A            | GAACCTTTAGGTATTGTATAACAATTTGTGTGATGA                   |
|                         | TTGACGTTCCCTAGTTGTGCGCAATGGGGAGAAAATATAT               |                         | TCATCACACAAATTGTTATACAATACCTAAAGGTTCTTC                |
| c.7008-1G>A             | TATATTTTCTCCCATTTGCAACACAACTAAGGAACGTCAAG              | c.7436-22C>T            | AATTTCAATTTTATTTTGTGAAGTATTATTCTTTGATA                 |
|                         | CTTGACGTTCCCTAGTTGTGTTGCAATGGGGAGAAAATATA              |                         | TATCAAAGAATAAACTTAACAAAAATAAAATTGAAATT                 |
| c.7009A>G               | TATTTTCTCCCATTTGCAGCGCAACTAAGGAACGTCAAGAG              | c.7436-14T>G            | ATTTTATTTTGTGAAGTATGTATTCTTTGATAGATTTAA                |
|                         | CTCTTGACGTTCCCTAGTTGCGCTGCAATGGGGAGAAAATA              |                         | TTAAATCTATCAAAGAATACATACTTAGCAAAAATAAAAT               |
| c.7010C>T               | ATTTTCTCCCATTTGCAGCATAACTAAGGAACGTCAAGAGA              | c.7436-4A>G             | GCTAAGTATTTATTCTTTGGTAGATTTAATTACAAGTCTT               |
|                         | TCTCTTGACGTTCCCTAGTTATGCTGCAATGGGGAGAAAAT              |                         | AAGACTTGTAATTAATCTACCAAAGAATAAATACTTAGC                |
| c.7030A>G               | CAACTAAGGAACGTCAAGAGGTACAGAATCCAAATTTTAC               | c.7436-4A>T             | GCTAAGTATTTATTCTTTGTTAGATTTAATTACAAGTCTT               |
|                         | GTAATAATTTGGATTCTGTACCTCTTGACGTTCCCTAGTTG              |                         | AAGACTTGTAATTAATCTAACAAGAATAAATACTTAGC                 |
| c.7170T>G               | TCTGCTACAAGAAAGGAAAAAATGAGACACTTGAT                    | c.7436-2A>T             | CTAAGTATTTATTCTTTGATTGATTTAATTACAAGTCTTC               |
|                         | ATCAAGTGCTCATTTTTTCTTTCTTGAGCAGA                       |                         | GAAGACTTGTAATTAATCAATCAAAGAATAAATACTTAG                |
| c.7182A>G               | ACAAGAAATGAAAAAATGAGGCACCTGATTACTACAGGCAG              | c.7436-1G>A             | TTATTCTTTGATAAATTTAATTACAAGTCTTCAGAATGC                |
|                         | CTGCCTGTAGTAATCAAGTGCTCATTTTTTCTTTCTTG                 |                         | GCATTCTGAAGACTTGTAATTAATTTATCAAAGAATAA                 |
| c.7266T>A               | TTTCACAGAGTTGAACAGTGAGTTAGGAATATTAACCTGG               | c.7447A>G               | CTTTGATAGATTTAATTACAGGTCTTCAGAATGCCAGAGA               |
|                         | CCAAGTTAATATTCCTAACTCACTGTTCAACTCTGTGAAA               |                         | TCTCTGGCATTCTGAAGACCTGTAATTAATCTATCAAAG                |

|           |  |                         |  |
|-----------|--|-------------------------|--|
| c.7418G>A | GCAGCTGTAACCTTTCACAAAGTATGAAGAAGAACCTTTAGGTATT | c.7466A>G               | AGTCTTCAGAATGCCAGAGGTATACAGGATATGCGAATTA |
|           | AATACCTAAAGGTTCTTCTTCATACTTTGTGAAAGTTACAGCTGC  |                         | TAATTCGCATATCCTGTATACCTCTGGCATTCTGAAGACT |
| c.7024C>T | CAGCACAACTAAGGAACGTTAAGAGATACAGAATCCAAAT       | c.7467T>C               | TCTTCAGAATGCCAGAGACATACAGGATATGCGAATTAA  |
|           | ATTTGGATTCTGTATCTCTTAACGTTCCCTAGTTGTGCTG       |                         | TTAATTCGCATATCCTGTATGTCTCTGGCATTCTGAAGA  |
| c.7037A>G | GAACGTCAAGAGATACAGAGTCCAAATTTACCGCACC          | c.7471C>T               | AGAATGCCAGAGATATATAGGATATGCGAATTAAGAAG   |
|           | GGTGCGGTAAAATTTGGACTCTGTATCTCTTGACGTTT         |                         | CTTCTTAATTCGCATATCCTATATATCTCTGGCATTCT   |
| c.7157C>A | ACATCCATTTTATCAAGTTTATGCTACAAGAAAT             | c.7471delC              | AGAATGCCAGAGATATAAGGATATGCGAATTAAGAAG    |
|           | ATTTCTGTAGCATAACTTGATAAAATGGATGT               |                         | CTTCTTAATTCGCATATCCTTATATCTCTGGCATTCT    |
| c.7177A>G | TCAAGTTTCTGCTACAAGAAATGAAAAAGTGAGACACTTGATTAC  | c.7472A>G               | TCAGAATGCCAGAGATATACGGGATATGCGAATTAAGAAG |
|           | GTAATCAAGTGCTCACTTTTTCATTTCTGTAGCAGAACTTGA     |                         | CTTCTTAATTCGCATATCCCGTATATCTCTGGCATTCTGA |
| c.7180A>T | CTACAAGAAATGAAAAATGTGACACTTGATTACTACAGG        | c.7474_7475delGA        | AGAATGCCAGAGATATACAGTATGCGAATTAAGAAGAAAC |
|           | CCTGTAGTAATCAAGTGTCACATTTTTTCATTTCTTGTAG       |                         | GTTTCTTCTTAATTCGCATACTGTATATCTCTGGCATTCT |
| c.7203A>G | CACTTGATTACTACAGGCAGGCCAACCAAGTCTTTG           | c.7492A>G               | AGGATATGCGAATTAAGAAGGAACAAAGGCAACGCGTCTT |
|           | CAAAGACTTTGGTTGGCCTGCCTGTAGTAATCAAGTG          |                         | AAGACGCGTTGCCTTTGTTCTTCTTAATTCGCATATCCT  |
| c.7261C>G | ACATTTTCACAGAGTTGAAGAGTGTTAGGAATATTAAC         | c.7501C>T               | GAATTAAGAAGAAACAAAGGTAACGCGTCTTTCCACA    |
|           | GTTAATATTCTAACACACTCTTCAACTCTGTGAAAATGT        |                         | TGTGGAAAGACGCGTTACCTTTGTTCTTCTTAATTC     |
| c.7294A>G | ATATTAACCTGGAGGAAAACGGACAAAAGCAAAACATTGAT      | c.7544C>T               | AGTCTGTATCTTGCAAAAATATCCACTCTGCCTCGAATC  |
|           | ATCAATGTTTTGCTTTTGTCCGTTTTCTCCAAGTTAATAT       |                         | GATTGAGGCAGAGTGGATATTTTGAAGATACAGACT     |
| c.7296A>G | ATTAACCTGGAGGAAAACAGGCAAAAGCAAAACATTGATGG      | c.7598C>G               | CAGTAGGAGGCCAAGTTCCTGTGCGTGTTCTCATAAAC   |
|           | CCATCAATGTTTTGCTTTTGCCTGTTTTCTCCAAGTTAAT       |                         | GTTTATGAGAACACGCACAGGGAACCTGGCCTCCTACTG  |
| c.7330G>T | TTGATGGACATGGCTCTGATTATAGTAAAAATAAGATTAAT      | c.7601C>T               | GGAGGCCAAGTTCCTCTGTGTGTTCTCATAAACAGGTA   |
|           | ATTAATCTTATTTTACTATAATCAGAGCCATGTCCATCAA       |                         | TACCTGTTTATGAGAACACACAGAGGGAACCTGGCCTCC  |
| c.7339A>G | GGCTCTGATGATAGTAAAGATAAGATTAATGACAAATGAG       | c.7611_7615<br>delTAAAC | GTTCCCTCTGCGTGTTCTCAAGGTATGTGTTTGTCTAC   |
|           | CTCATTGTCATTAATCTTATCTTTACTATCATCAGAGCC        |                         | GTAGACAAACACATACCTTGAGAACACGCAGAGGGAAC   |
| c.7397C>T | AAACAACTCCAATCAAGCAGTAGCTGTAACCTTTCACAAA       | c.7617G>A               | TCTGCGTGTTCTCATAAACAAGTATGTGTTTGTCTACAA  |
|           | TTTGTGAAAGTTACAGCTACTGCTTGATTGGAGTTGTTT        |                         | TTGTAGACAAACACATACTTGTTTATGAGAACACGCAGA  |

|             |   |
|-------------|---|
| c.7428A>G   | ACAAAGTGTGAAGAAGAGCCTTTAGGTATTGTATGA          |
|             | TCATACAATACCTAAAGGCTCTTCTTCACACTTTGT          |
| c.7435+1G>C | CAAAGTGTGAAGAAGAACCTTTAGCTATTGTATGACA         |
|             | TGTCATACAATAGCTAAAGGTTCTTCTTCACACTTTG         |
| c.7435+3A>G | TGTGAAGAAGAACCTTTAGGTGTTGTATGACAATTTGTGTGATGA |
|             | TCATCACACAAATTGTCATACAACACCTAAAGGTTCTTCTTCACA |
| c.7435+5T>C | CCTTTAGGTATCGTATGACAATTTGTGTGATGAATTTTGGC     |
|             | GGCAAAAATTCATCACACAAATTGTCATACGATACCTAAAGG    |

|             |  |
|-------------|--|
| c.7617+1G>A | CTGCGTGTTCTCATAAACAGATATGTGTTTGTCTACAATA |
|             | TATTGTAGACAAACACATATCTGTTTATGAGAACACGCAG |
| c.7617+1G>T | CTGCGTGTTCTCATAAACAGTTATGTGTTTGTCTACAATA |
|             | TATTGTAGACAAACACATAACTGTTTATGAGAACACGCAG |
| c.7617+2T>G | TTCTCATAAACAGGGATGTGTTTGTCTACAATACTGATGG |
|             | CCATCAGTATTGTAGACAAACACATCCCTGTTTATGAGAA |



**TABLA ANEXA 5. Análisis y clasificación de las 73 variantes que alteraban gravemente el *splicing*.**

| Variantes de <i>BRCA2</i>                    | Tipo de Variante                | % Transcrito canónico | Principales transcritos producidos | Efecto en la proteína | Clasificación previa | Criterios de ACMG        | Nueva clasificación según ACMG |
|--|---------------------------------|-----------------------|------------------------------------|-----------------------|----------------------|--------------------------|--------------------------------|
| <b>Variantes analizadas con el MGBR2_2-9</b> |                                 |                       |                                    |                       |                      |                          |                                |
| c.67G>A                                      | <i>Missense</i><br>(p.Asp23Asn) | 7.7%                  | Δ2q <sub>4</sub> (r.64_67del)      | p.A22Ifs*2            | VUS/Prob. Patogénica | PVS1, PS3, PM2, PP3, PP5 | Patogénica                     |
| c.67+1G>T                                    | Intrónica                       | -                     | Δ2 (r.-39_67del)                   | p.M?                  | Patogénica           | PVS1, PS3, PM2, PP3, PP5 | Patogénica                     |
| c.67+2T>C                                    | Intrónica                       | -                     | Δ2 (r.-39_67del)                   | p.M?                  | Patogénica           | PVS1, PS3, PM2, PP3, PP5 | Patogénica                     |
| c.67+3A>G                                    | Intrónica                       | 2,4%                  | Δ2 (r.-39_67del)                   | p.M?                  | VUS                  | PVS1, PS3, PM2, PP3, PP5 | Patogénica                     |
| c.97G>A                                      | <i>Missense</i><br>(p.Glu33Lys) | 23,3%                 | Δ3 (r.68_316del)                   | p.D23_L105del         | VUS                  | PM2, PM4, PP3, PP5       | Prob. Patogénica               |
| c.100G>A                                     | <i>Missense</i><br>(p.Glu34Lys) | 33,7%                 | Δ3 (r.68_316del)                   | p.D23_L105del         | VUS                  | PM2, PM4, PP3, PP5       | Prob. Patogénica               |
| c.316+2T>C                                   | Intrónica                       | -                     | Δ3 (r.68_316del)                   | p.D23_L105del         | Patogénica           | PVS1, PS3, PM2, PP3, PP5 | Patogénica                     |
| c.316+3delA                                  | Intrónica                       | -                     | Δ3 (r.68_316del)                   | p.D23_L105del         | VUS                  | PVS1, PS3, PM2, PP3, PP5 | Patogénica                     |

|              |                               |       |   |                                      |                              |                          |                  |
|--------------|-------------------------------|-------|---|--------------------------------------|------------------------------|--------------------------|------------------|
| c.316+5G>C   | Intrónica                     | -     | Δ3 (r.68_316del)  | p.D23_L105del                        | Patogénica                   | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.316+6T>C   | Intrónica                     | 21,9% | Δ3 (r.68_316del)  | p.D23_L105del                        | VUS                          | PS3, PM2, PP3, PP5       | Prob. Patogénica |
| c.317-2A>G   | Intrónica                     | -     | Δ4;Δ(4,5);Δ(4,5,6) (r.[317_425del;317_475del;317_516del]) | p.[G106Vfs*10;G106_Ser158;R107Sfs*9] | Prob. Patogénica             | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.426-12del5 | Intrónica                     | 24,7% | Δ5 (r.426_475del)   | p.P143Gfs*23                         | VUS/Prob. Patogénica         | PVS1, PS3, PM2, PP5      | Patogénica       |
| c.426-2A>T   | Intrónica                     | -     | Δ5 (r.426_475del)   | p.P143Gfs*23                         | Patogénica                   | PVS1, PS3, PP3, PP5      | Patogénica       |
| c.441A>G     | Sinónima (p.Gln147Gln)        | -     | Δ5 (r.426_475del)   | p.P143Gfs*23                         | Prob. Benigna                | PS3, PP3                 | VUS              |
| c.451G>A     | Missense (p.Val151Ile)        | -     | Δ5;Δ(5,6) (r.[426_475del;426_516del])                     | p.[P143Gfs*23;S142Rfs*13]            | VUS                          | PS3, PM2, PP3            | Prob. Patogénica |
| c.467A>G     | Missense (p.Asp156Gly)        | -     | Δ5q9 (r.467_475del)                                       | p.D156_S158del                       | VUS/Prob. Benigna            | PM4, PP3, PP5            | VUS              |
| c.470_474del | Frameshift (p.Lys157Serfs*24) | 16,7% | Δ5 (r.426_475del)   | p.P143Gfs*23                         | Patogénica                   | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.475+1G>T   | Intrónica                     | -     | Δ5 (r.426_475del)   | p.P143Gfs*23                         | Patogénica/ Prob. Patogénica | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.475+3A>T   | Intrónica                     | -     | Δ5 (r.426_475del)   | p.P143Gfs*23                         | Patogénica                   | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.476-2A>G   | Intrónica                     | -     | Δ6;Δ(5,6) (r.[476_516del;r.426_516del])                   | p.[V159Gfs*10;S142Rfs*13]            | Patogénica                   | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |

|              |                                  |       |  |                           |                                 |                          |                  |
|--------------|----------------------------------|-------|--|---------------------------|---------------------------------|--------------------------|------------------|
| c.476-3C>A   | Intrónica                        | 29,8% | $\Delta(5,6); \Delta 6$<br>(r.[476_516del;r.426_516del]) | p.[V159Gfs*10;S142Rfs*13] | Patogénica                      | PS3, PP3, PP5            | Prob. Patogénica |
| c.516+1G>T   | Intrónica                        | -     | $\Delta 6; \Delta(5,6)$<br>(r.[476_516del;426_516del])   | p.[V159Gfs*10;S142Rfs*13] | Patogénica                      | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.516+2T>C   | Intrónica                        | -     | $\Delta 6; \Delta(5,6)$<br>(r.[476_516del;426_516del])   | p.[V159Gfs*10;S142Rfs*13] | Patogénica                      | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.516+4delAA | Intrónica                        | 6,4%  | $\Delta(5,6), \Delta 6$<br>(r.[426_516del;476_516del])   | p.[S142Rfs*13;V159Gfs*10] | VUS                             | PVS1, PS3, PM2, PP3      | Patogénica       |
| c.517-2A>G   | Intrónica                        | -     | $\Delta 7$ (r.517_631del)                                | p.G173Sfs*19              | Patogénica/<br>Prob. Patogénica | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.517-1G>A   | Intrónica                        | -     | $\Delta 7p; \Delta 7$<br>(r.[517del;517_631del])         | p.[G173Vfs*12;G173Sfs*19] | Patogénica                      | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.517G>T     | <i>Missense</i><br>(p.Gly173Cys) | -     | $\Delta 7$ (r.517_631del)                                | p.G173Sfs*19              | VUS                             | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.572A>G     | <i>Missense</i><br>(p.Asp191Gly) | 14,9% | $\Delta 7q_{60}$ (r.572_631del)                          | p.D191_Ile210del          | VUS                             | PM2, PM4, PP3            | VUS              |
| c.631G>A     | <i>Missense</i><br>(p.Val211Ile) | -     | $\Delta 7$ (r.517_631del)                                | p.G173Sfs*19              | Patogénica                      | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.631+1G>A   | Intrónica                        | -     | $\Delta 7$ (r.517_631del)                                | p.G173Sfs*19              | Patogénica/<br>Prob. Patogénica | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.631+3A>G   | Intrónica                        | 7,8%  | $\Delta 7$ (r.517_631del)                                | p.G173Sfs*19              | VUS/<br>Patogénica              | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |

|  |           |      |   |                          |                                 |                          |                  |
|--|-----------|------|---|--------------------------|---------------------------------|--------------------------|------------------|
| c.632-3C>G                                     | Intrónica | -    | ▼8p <sup>2</sup> ;Δ(6q <sub>39</sub> ,7)▼8p <sup>2</sup> (r.[631_632ins632-1_632-2;478_631delins632-1_632-2]) | p.[V211Efs*20;V160Sfs*4] | VUS                             | PVS1, PS3, PM2, PP5      | Patogénica       |
| c.632-1G>C                                     | Intrónica | -    | Δ8p <sub>4</sub> ;Δ(6q <sub>39</sub> ,7,8p <sub>4</sub> )r.[632_635del;517_635]                               | p.[V211Efs*18;G173Kfs*2] | Patogénica                      | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.681+4A>G                                     | Intrónica | -    | ▼8q <sup>4</sup> (r.681_682ins681+1_681+4)  | p.N228Vfs*11             | VUS/Prob. Patogénica/Patogénica | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.682-1G>C                                     | Intrónica | -    | Δ9 (r.682_793del)   | p.N228Dfs*12             | Prob. Patogénica                | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.793+1G>T                                     | Intrónica | -    | Δ9 (r.682_793del)   | p.N228Dfs*12             | Patogénica                      | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| <b>Variantes analizadas con el MGBR2_14-20</b> |           |      |   |                          |                                 |                          |                  |
| c.7008-3C>G                                    | Intrónica | 7,2% | Δ14 (r.7008_7435del)  | Thr2337Phefs*17          | VUS                             | PS3, PM2, PP3, PP5       | Prob. Patogénica |
| c.7008-2A>T                                    | Intrónica | -    | Δ14p10 (r.7008_7017del)   | Thr2337Asnfs*27          | Patogénica                      | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.7008-1G>A                                    | Intrónica | -    | Δ14p10 (r.7008_7017del)   | Thr2337Asnfs*27          | VUS                             | PVS1, PS3, PM2, PP3      | Patogénica       |
| c.7435+1G>C                                    | Intrónica | -    | ▼(E14q5) (r.7435_7435+5ins)   | Asp2479Glyfs*4           | Patogénica                      | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |
| c.7436-2A>T                                    | Intrónica | -    | Δ15p13 (r.7436_7448del)   | Asp2479Valfs*41          | Patogénica                      | PVS1, PS3, PM2, PP3, PP5 | Patogénica       |

|             |                                   |      |   |                                |            |                                |                     |
|-------------|-----------------------------------|------|---|--------------------------------|------------|--------------------------------|---------------------|
| c.7436-1G>A | Intrónica                         | -    | Δ15p13<br>(r.7436_7448del)                          | Asp2479Valfs*41                | VUS        | PVS1, PS3,<br>PM2, PP3         | Patogénica          |
| c.7617+1G>A | Intrónica                         | -    | Δ15 (r.7436_7617del)                                | Asp2479Alafs*8                 | Patogénica | PVS1, PS3,<br>PM2, PP3<br>PP5  | Patogénica          |
| c.7617+1G>T | Intrónica                         | -    | Δ15 (r.7436_7617del)                                | Asp2479Alafs*8                 | Patogénica | PVS1, PS3,<br>PM2, PP3<br>PP5  | Patogénica          |
| c.7617+2T>G | Intrónica                         | -    | Δ15 (r.7436_7617del)                                | Asp2479Alafs*8                 | Patogénica | PVS1, PS3,<br>PM2, PP3<br>PP5  | Patogénica          |
| c.7618-2A>T | Intrónica                         | -    | Δ16p44<br>(r.7618_7661del)                          | p.L2540Qfs*11                  | Patogénica | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| c.7618-2A>G | Intrónica                         | -    | Δ16p44<br>(r.7618_7661del)                          | p.L2540Qfs*11                  | Patogénica | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| c.7618-1G>A | Intrónica                         | -    | Δ16p44<br>(r.7618_7661del)                          | p.L2540Qfs*11                  | Patogénica | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| c.7618-1G>C | Intrónica                         | -    | Δ16p44<br>(r.7618_7661del)                          | p.L2540Qfs*11                  | Patogénica | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| c.7805G>C   | <i>Missense</i><br>(p.Arg2602Thr) | -    | Δ16;Δ16q100<br>(r.[7618_7805del;<br>7706_7805del])  | p.[L2540Gfs*4;<br>K2570Lfs*45] | Patogénica | PS3, PM2,<br>PP3, PP5          | Prob.<br>Patogénica |
| c.7805+1G>A | Intrónica                         | -    | Δ16, Δ16q100<br>(r.[7618_7805del;<br>7706_7805del]) | p.[L2540Gfs*4;<br>K2570Lfs*45] | Patogénica | PVS1, PS3,<br>PM2, PP3,<br>PP5 | Patogénica          |
| c.7805+3A>C | Intrónica                         | 7,6% | Δ16;Δ16q100<br>(r.[7618_7805del;<br>7706_7805del])  | p.[L2540Gfs*4;<br>K2570Lfs*45] | VUS        | PS3, PM2,<br>PP3, PP5          | Prob.<br>Patogénica |

|                  |                                   |   |  |   |                    |                          |                  |
|------------------|-----------------------------------|---|--|---|--------------------|--------------------------|------------------|
| c.7806-9T>G      | Intrónica                         | - | Δ17;▼17p8 ;Δ17p69<br>(r.[7806_7976del;7805_7806ins7806-8_7806-1;7806_7874del]) | p.[A2603_R2659del;R2602Sfs*49;A2603_R2625del] | VUS                | PS3, PM2, PM4, PP3       | Prob. Patogénica |
| c.7806-2A>G      | Intrónica                         | - | Δ17;Δ17p20;Δ17p69<br>(r.[7806_7976del;7806_7825del;7806_7874del,])             | p.[A2603_R2659del;A2603Cfs*8;A2603_R2625del]  | Patogénica         | PVS1, PS3, PM2, PM4, PP3 | Patogénica       |
| c.7806-1G>A      | Intrónica                         | - | Δ17p1<br>(r.7806_7807del)  | p.A2603Lfs*45                                 | Patogénica         | PVS1, PS3, PM2, PP3      | Patogénica       |
| c.7806-1G>T      | Intrónica                         | - | Δ17p20<br>(r.7806_7825del)   | p.A2603Cfs*8                                  | Patogénica         | PVS1, PS3, PM2, PP3      | Patogénica       |
| c.7806insAG      | Intrónica                         | - | ▼17p2<br>(r.7805_7806insAG)  | p.A2603Gfs*46                                 | Patogénica         | PVS1, PS3, PM2, PP3      | Patogénica       |
| c.7976G>C        | <i>Missense</i><br>(p.Arg2659Thr) | - | Δ17 (r.7806_7976del)   | p.A2603_R2659del                              | VUS                | PS3, PM2, PP3            | Prob. Patogénica |
| c.7976G>A        | <i>Missense</i><br>(p.Arg2659Lys) | - | Δ17 (r.7806_7976del)   | p.A2603_R2659del                              | VUS                | PS3, PP3, PP5            | Prob. Patogénica |
| c.7976+1G>A      | Intrónica                         | - | Δ17 (r.7806_7976del)   | p.A2603_R2659del                              | VUS                | PVS1, PS3, PM2, PP3      | Patogénica       |
| c.7976+5G>T      | Intrónica                         | - | Δ17 (r.7806_7976del)   | p.A2603_R2659del                              | VUS                | PS3, PM2, PP4            | Prob. Patogénica |
| c.7977-7C>G      | Intrónica                         | - | ▼18p6;Δ18<br>(r.[7976_7977ins6,7977_8331del])                                  | p.[Y2658_R2659insSF;Y2660Ffs*43]              | VUS                | PS3, PM2, PM4, PP3       | Prob. Patogénica |
| c.7977-3_7978del | Intrónica                         | - | Δ18 (r.7977_8331del)   | p.Y2660Ffs*43                                 | VUS                | PVS1, PS3, PM2, PP3      | Patogénica       |
| c.7977-2A>T      | Intrónica                         | - | Δ18 (r.7977_8331del)   | p.Y2660Ffs*43                                 | VUS                | PVS1, PS3, PM2, PP3      | Patogénica       |
| c.7977-1G>C      | Intrónica                         | - | Δ18 (r.7977_8331del)   | p.Y2660Ffs*43                                 | VUS/<br>Patogénica | PVS1, PS3, PM2, PP3      | Patogénica       |

|             |                                   |       |                             |               |            |                     |                     |
|-------------|-----------------------------------|-------|-----------------------------|---------------|------------|---------------------|---------------------|
| c.7977-1G>T | Intrónica                         | -     | Δ18 (r.7977_8331del)        | p.Y2660Ffs*43 | VUS        | PVS1, PS3, PM2, PP3 | Patogénica          |
| c.7985C>G   | <i>Missense</i><br>(p.Thr2662Arg) | -     | Δ18 (r.7977_8331del)        | p.Y2660Ffs*43 | VUS        | PS3, PM2, PP3       | Prob.<br>Patogénica |
| c.8009C>A   | <i>Nonsense</i><br>(p.Ser2670X)   | 4,0%  | Δ18 (r.7977_8331del)        | p.Y2660Ffs*43 | Patogénica | PVS1, PS3, PM2, PP3 | Patogénica          |
| c.8023A>G   | <i>Missense</i><br>(p.Ile2675Val) | -     | Δ18q309<br>(r.8023_8331del) | p.I2675_K2777 | Patogénica | PS3, PM2, PP3, PP5  | Prob.<br>Patogénica |
| c.8035G>T   | <i>Missense</i><br>(p.Asp2679Tyr) | 2,4%  | Δ18q298<br>(r.8034_8331del) | p.D2679Ffs*43 | Patogénica | PS3, PM2, PP3       | Prob.<br>Patogénica |
| c.8331G>A   | Sinónima<br>(p.Lys2777Lys)        | 40,7% | Δ18 (r.7977_8331del)        | p.Y2660Ffs*43 | VUS        | PS3, PM2, PP3, PP5  | Prob.<br>Patogénica |
| c.8331+1G>T | Intrónica                         | -     | Δ18 (r.7977_8331del)        | p.Y2660Ffs*43 | Patogénica | PVS1, PS3, PM2      | Patogénica          |
| c.8331+2T>C | Intrónica                         | -     | Δ18 (r.7977_8331del)        | p.Y2660Ffs*43 | Patogénica | PVS1, PS3, PM2      | Patogénica          |

