Human Mutation

Characterization of spliceogenic variants located in regions linked to high levels of alternative splicing: *BRCA2* c.7976+5G>T as a case study

Journal:	*Human Mutation*
Manuscript ID	humu-2017-0494.R2
Wiley - Manuscript type:	Brief Report
Date Submitted by the Author:	04-Jun-2018
Complete List of Authors:	Montalban, Gemma; Vall d'Hebron Institut d'Oncologia, Oncogenetics Fraile-Bethencourt, Eugenia; Instituto de Biologia y Genetica Molecular Lopez-Perolio, Irene; Instituto de Investigación Sanitaria San Carlos (IdISSC), Hospital Clínico San Carlos, Madrid Pérez-Segura, Pedro; Hospital Clinico Universitario San Carlos, Molecular Oncology Laboratory CIBERONC Infante, Mar; IBGM, durán, mercedes; IBGM, ; Alonso-Cerezo, María Concepción; Hospital Universitario de la Princesa, Genética Clínica López-Fernández, Adrià; Vall d'Hebron Institut d'Oncologia, High Risk and Cancer Prevention Group Diez, Orland; Vall d'Hebron Institut d'Oncologia, Oncogenetics; Hospital Vall d'Hebron, Area of Clinical and Molecular Genetics de la Hoya, Miguel; Hospital Clinico Universitario San Carlos Velasco, Eladio; Instituto de Biología y Genética Molecular (CSIC-UVa), Laboratorio de Splicing & Cáncer (B-8) Gutiérrez-Enríquez, Sara; Vall d'Hebron Institut d'Oncologia, Oncogenetics
Key Words:	Hereditary breast and ovarian cancer, BRCA2, atypical splicing site, alternative splicing, clinical classification

Characterization of spliceogenic variants located in regions linked to high levels of alternative splicing: *BRCA2* c.7976+5G>T as a case study

Gemma Montalban[1]*, Eugenia Fraile-Bethencourt[2]*, Irene López-Perolio[3], Pedro Pérez-Segura[3], Mar Infante[4], Mercedes Durán[4], María Concepción Alonso-Cerezo[5], Adrià López-Fernández[6], Orland Diez[1,7], Miguel de la Hoya[3], Eladio A. Velasco[2]**, Sara Gutiérrez-Enríquez[1]**

*Authors contributed equally to this work

**Corresponding authors

Affiliations:

[1]Oncogenetics Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

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

[3]Molecular Oncology Laboratory CIBERONC, Hospital Clinico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Madrid, Spain

[4]Cancer Genetics, Instituto de Biología y Genética Molecular (CSIC-UVa), Valladolid, Spain

[5]Genética Clínica. Servicio Análisis Clínicos. Hospital Universitario de la Princesa, Instituto de Investigación Sanitaria Hospital Universitario de la Princesa, Madrid, Spain

[6]High Risk and Cancer Prevention Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

[7]Area of Clinical and Molecular Genetics, University Hospital of Vall d'Hebron, Barcelona, Spain

Corresponding authors:

Sara Gutiérrez-Enríquez

Oncogenetics Group, Vall d'Hebron Institute of Oncology-VHIO

Lab 2.02A, CELLEX CENTER

C/ Natzaret, 115-117

08035 Barcelona, Spain

Phone: +34 932543450 ext. 8668

E-mail: sgutierrez@vhio.net

Eladio A. 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

Phone: +34 983184829

Fax: +34 983184800

E-mail: eavelsam@ibgm.uva.es

Funding:

This work was supported by Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds: FIS PI15/00355 (to O. Diez), PI13/01711 and PI16/01218 (to S. Gutiérrez-Enríquez), FIS PI15/00059 (to M. de la Hoya), and PI13/01749 and PI17/00227 (to EA Velasco). E Fraile-Bethencourt was supported by a predoctoral fellowship from the University of Valladolid and Banco de Santander (2015-2019). EA Velasco was also funded by grant CSI090U14 from the Consejería de Educación (ORDEN EDU/122/2014), Junta de Castilla y León. S. Gutiérrez-Enríquez is supported by the Miguel Servet Progam (CP10/00617).

Abstract

Many *BRCA1* and *BRCA2* (*BRCA1/2*) genetic variants have been studied at mRNA level and linked to hereditary breast and ovarian cancer due to splicing alteration. *In silico* tools are reliable when assessing variants located in consensus splice sites, but we may identify variants in complex genomic contexts for which bioinformatics is not precise enough. In this study, we characterize *BRCA2* c.7976+5G>T variant located in intron 17 which has an atypical donor site (GC). This variant was identified in three unrelated Spanish families and we have detected exon 17 skipping as the predominant transcript occurring in carriers. We have also detected several isoforms (Δ16-18, Δ17,18, Δ18 and ▼17q^{224}) at different expression levels among carriers and controls. This study remarks the challenge of interpreting genetic variants when multiple alternative isoforms are present, and that caution must be taken when using *in silico* tools to identify potential spliceogenic variants located in GC-AG introns.

Key words: Hereditary breast and ovarian cancer, *BRCA2*, atypical splicing site, alternative splicing, clinical classification

BRCA1 (MIM# 113705) and *BRCA2* (MIM# 600185) (*BRCA1/2*) genes are associated to hereditary breast and ovarian cancer syndrome (HBOC). Pathogenic variants in *BRCA1/2* result in an increased cumulative breast cancer risk to age 80 that ranges from 61% to 79%, and in an increased risk for ovarian cancer that ranges from 11% to 53% (Kuchenbaecker et al., 2017). Genetic variants in disease-responsible genes that disrupt the splicing code have a key role in human hereditary disorders and cancer. A recent worldwide study describing the mutational spectrum of *BRCA1*/2 genes in HBOC families, identified that 10.1% of *BRCA1* and 7.6% of *BRCA2* pathogenic variants result in aberrant mRNA splicing (Rebbeck et al., 2018). Splicing mutations were traditionally considered those that affect consensus splice sites (intronic nucleotides +1 and +2 of the donor GT, and -1 and -2 of the acceptor AG sites), but other intronic and exonic nucleotides outside these regions have also been found to be highly conserved and critical for splice site selection (Cartegni et al., 2002; Manning and Cooper, 2017). More than 99% of human introns are flanked by GT-AG splice site dinucleotides and are spliced by the so-called major U2-type spliceosome. An exception to this rule are the U2-type GC-AG introns, such as *BRCA2* intron 17 (see Supp. Figure S1), comprising about 0.9% of all human splice sites (reviewed by Parada et al., 2014; Sibley et al., 2016). GC-AG introns possess weak donor sites that are compensated with strong consensus in the surrounding nucleotides of the donor, and are usually linked to alternatively spliced exons (Thanaraj and Clark, 2001; Churbanov et al., 2008; Kralovicova et al., 2011). Given the complexity of these introns, we aimed to characterize the splicing impact of *BRCA2* c.7976+5G>T variant, located at position +5 from the atypical donor site of intron 17. To our knowledge, this variant is not present in the gene-specific databases LOVD (www.lovd.nl), BRCA Exchange (brcaexchange.org) and BIC (https://research.nhgri.nih.gov/bic), ascertained by February 2018. This variant is reported twice in ClinVar database and categorized as conflicting

(https://www.ncbi.nlm.nih.gov/clinvar/), it is reported as a variant of unknown significance (VUS) in BRCA Share (www.umd.be), and it is classified as likely benign (Class-2) in a previous work published by Garibay et al., 2014.

We identified this variant in three unrelated HBOC Spanish families and initiated this collaborative study to exhaustively re-evaluate the variant across three laboratories: Hospital Universitari Vall d'Hebron (HUVH), Hospital Clínico San Carlos (HCSC) and Instituto de Biología y Genética Molecular (IBGM). Probands underwent genetic counselling and written informed consent was obtained in all cases. Family information and pedigrees are described in Supp. Figure S2.

BRCA2 c.7976+5G>T was first assessed *in silico* using Human Splicing Finder (HSF) and MaxEntScan (MES) (http://www.umd.be/HSF3/), which predicted a reduction of the native donor splice site (11.66% and 74.19%, respectively). However, when using the splicing module of Alamut software v2.10 (Interactive Biosoftware) only SSF-like computed a score, predicting a 12.8% reduction of the donor site (Supp. Table S1). Breast Cancer Genes Prior Probabilities website (http://priors.hci.utah.edu/PRIORS/index.php) also predicted a high probability of pathogenicity (0.97) due to splice site donor damage. *In vitro* characterization was performed with patient RNA and a minigene system (see Supp. Methods and Supp. Table S2 for detailed protocols used in each laboratory). HUVH samples were analyzed by RT-PCR using primers located in exons 15 and 19. We detected 5 transcripts corresponding to the reference full-length (FL) (906bp), Δ18 (551bp), Δ17 (735bp), Δ17,18 (380bp) and Δ16-18 (192bp) (Supp. Figure S3). All transcripts were detected in carrier and control samples (n=10), with the exception of Δ17 which was only present in the carrier. Capillary electrophoresis of fluorescent amplicons (CE) ruled out any expression of Δ17 in control samples, indicating that only the variant allele generates this transcript (Figure 1A). HCSC samples were analyzed using primers located in exons 16 and 19. In variant carrier, RT-PCR

assays (Figure 1B) showed transcripts corresponding to FL, Δ17, Δ17,18, and an additional peak of ≈900nt that we tentatively annotated as ▼17q^{224} (Supp. Figure S4). In controls (n=34), FL was detected in all samples, Δ17 was absent in all samples, and isoforms Δ17,18, Δ18 and the putative ▼17q^{224} were detected in 11, 3 and 6 samples, respectively. IBGM carrier was analyzed with primers located in exons 16 and 19 and RT-PCRs showed FL and Δ17 transcripts, without evidence of any other aberrant transcripts or isoforms (Figure 1C). In summary, we identified three frameshift isoforms Δ17,18 (p.Ala2603Phefs*43), Δ18 (p.Tyr2660Phefs*43) and ▼17q^{224} (p.Arg2659Argfs*3) occurring in carrier and control samples; one in-frame isoform Δ16-18 (p.Leu2540_Lys2777del) detected in carrier and control samples; and one in-frame isoform Δ17 (p.Ala2603_Arg2659del) detected only in carriers (results are summarized in Supp. Table S3). All transcripts were confirmed by Sanger sequencing with the exception of Δ16-18 and ▼17q^{224} which were imputed based on the length of the product observed.

Semi-quantitative CE analysis was performed in HUVH and HCSC samples (IBGM patient sample was not available) and showed that Δ17 represents a substantial contribution to the total splicing fraction (SF) in HUVH carrier and HCSC carrier (average SF 34.3% and 48.3%, respectively), whereas in controls was not detected (Figure 2A). Regarding alternative transcripts, notable differences were observed in isoform Δ17,18 levels between HUVH and HCSC carriers (SF 26% and 9.6%, respectively), whereas HUVH and HCSC controls had similar levels (SF 3.3% and 3.9%, respectively). Isoform Δ18 is a minor event detected only in HUVH carrier (SF 7.2%) and in HUVH/HCSC controls (SF 4.2% and 1.4%, respectively); isoform Δ16-18 was only detected in HUVH due to primer location and displayed no notable differences between carrier and controls (SF 13.1% and 11.1%, respectively); and putative isoform ▼17q^{224} was only detected in HCSC samples, although not in all RT-PCR assays, and showed higher levels in HCSC carrier (SF 17.1%) compared to controls (1.2%) (Figure

2A). Two previous studies detected Δ17,18 and Δ18 in control lymphoblastoid cell lines (LCLs) and normal breast tissue, with Δ17,18 being more abundant (Fackenthal et al., 2016; Davy et al., 2017). Normalized CE data from full-length transcript (FL) showed a 2-fold reduction in carriers compared to controls, suggesting that the variant allele is not producing FL (Figure 2B). To test this, since allele-specific assays could not be performed due to the lack of heterozygous informative loci in patient sample, the mutant allele (c.7976+5G>T) was artificially interrogated using a pSAD-derived minigene with *BRCA2* exons 14 to 20, constructed and functionally validated as previously described in Fraile-Bethencourt et al., 2017. A wild-type (wt) construct and a variant construct *BRCA2* c.7976+1G>A were used as negative and positive controls, respectively. Wt construct produced a stable canonical transcript of the expected size and structure, and variant constructs revealed a unique transcript corresponding to exon 17 skipping (Mean ± SEM of relative fluorescence: 1.045 ± 0.111 and 0.955 ± 0.020 for variant and positive control, respectively) (Figure 1D).

Given that notable differences in isoform Δ17,18 levels were observed among samples (Figure 2A), we aimed to determine whether such differences were due to technical reasons and we used digital PCR as a second approach to measure Δ17,18 levels. Data obtained was consistent with CE data, i.e., the highest value was observed in HUVH carrier (SF ~9% vs. SF ~2% observed in HCSC carrier) (Figures 2C and 2D). A slight variability was also observed among control samples, with SF levels ranging from ~0.5 to 4%. Interestingly, a recent study identified a common *BRCA2* variant c.7806-14T>C (rs9534262) which influences Δ17,18 levels by modulating exon 17 acceptor site (Garibay et al., 2014). DNA sequence analysis revealed that HUVH patient is homozygous (C/C) at this polymorphic position, whereas HCSC and IBGM are heterozygous (T/C) (data not shown). Sanger sequencing of introns surrounding exon 17 (limited to c.7806-37_7806-1 and c.7976+1_7976+25 regions) in HUVH and HCSC carriers, and whole intron 16 and 17

analysis in IBGM carrier, did not identify additional rare or common variants that could explain differential Δ17,18 isoform expression among carriers. In this regard, patient genotypes were consistent with experimental data in that the highest Δ17,18 levels were detected in the homozygous C/C patient. The influence of c.7806-14T>C on Δ17,18 levels was additionally evaluated in HCSC controls (n=34) by RT-PCR and CE, and the highest levels were again observed in C/C samples (Supp. Figure S5). Nevertheless, whether an increase of Δ17,18 is associated with our variant under scrutiny or not, would require the analysis of carriers being c.7806-14T/T to avoid its influence on Δ17,18 levels.

We also re-evaluated HUVH and HCSC RNA samples to explore whether methodological differences between laboratories could influence variability in isoform levels. To do so, carriers and controls were analysed using primers and protocols from HCSC laboratory (see Supp. Table S2). Results obtained showed that major transcriptional events were detected as previously: Δ17 was present in both carriers and absent in controls, and Δ17,18 highest levels were detected in HUVH carrier. Minor isoform Δ18 was again detected in HUVH carrier and absent in HCSC carrier, and ▼17q^{224} was not detected in either carriers (Supp. Figure S6). This data indicates that differences in minor events (Δ18 and ▼17q^{224}) are presumably due to stochastic effects during PCR amplification, but variability in major isoform Δ17,18 is likely to be linked to individual genetic features rather than methodological differences. In this analysis, polymerases and CE conditions did not seem to have an influence in isoform detection, and although new RNA samples could not be obtained to rule out any influence of RNA isolation methods, a collaborative work comparing RNA protocols for characterization of spliceogenic variants across multiple laboratories concluded that RNA extraction methods were indistinguishable (Whiley et al., 2014).

Overall, combined analysis carried out in three different laboratories provides convincing evidence that the major splicing outcome produced by *BRCA2* c.7976+5G>T variant is exon

17 skipping, which causes an in-frame deletion (r.7806_7976del171) that results in a protein lacking 57 amino acids (aa) (p.Ala2603_Arg2659del). The lost region is part of the α-helical domain (aa 2479 to 2667) of BRCA2 DNA Binding Domain (DBD), which has 30 strictly conserved residues from sea urchin to human (Supp. Figure S7). This domain enables BRCA2 binding to single-stranded and double-stranded DNA, and is essential to allow DNA repair by homologous recombination (HR) (Roy et al., 2011). Functional assays for classification of missense variants located in this region confirmed a reduction of BRCA2 HR activity for Trp2626Cys, Ile2627Phe, Leu2647Pro, Leu2653Pro and Arg2659Lys variants (Farrugia et al., 2008; Biswas et al., 2011). From these, variants Trp2626Cys, Ile2627Phe, Leu2653Pro and Arg2659Lys had been previously evaluated by multifactorial analysis and classified as pathogenic (Class 5) (Easton et al., 2007). Other variants causing exon 17 skipping (c.7976G>A, c.7976+1G>A and c.7976+3_7976+4del) have been identified in HBOC patients and reported as deleterious (Hofmann et al., 2003; Wu et al., 2005; Thirthagiri et al., 2008; Brandão et al., 2011; Fraile-Bethencourt et al., 2017). Moreover, allele-specific assessment was performed in patient RNA for variant c.7976+3_7976+4del, and only detected transcript lacking exon 17 (Brandão et al., 2011). Although Δ17 was categorized as a minor alternative splicing event occurring in control LCLs and normal breast tissue, it was not detected in whole blood control samples (Fackenthal et al., 2016). Likewise, we did not identify this transcript in our control group (n=44).

Our variant G>T is located at position +5 from intron 17 donor splice site, where a G is present in >80% of human introns (Zhang, 1998). In this particular case, *BRCA2* intron 17 has an atypical donor site GC weaker than the GT counterparts because of the +2 substitution, meaning that the rest of nucleotide positions are more conserved (Thanaraj and Clark, 2001) and that variants in any of these nucleotides may have an impact on exon recognition. However, it is important to note that not all *in silico* approaches used in this study to predict

the splicing impact of *BRCA2* c.7976+5G>T variant were able to detect its potential damage to the native donor site. Other *BRCA1/2* variants located at position +5 have been reported to induce splicing alterations, such as *BRCA1* c.5406+5G>C (exon 22 skipping) and c.5467+5G>C (exon 23 skipping), *BRCA2* c.316+5G>C (exon 3 skipping) and c.8754+5G>A (insertion of 46 nucleotides of intron 21) (Vreeswijk et al., 2009; Whiley et al., 2011; Houdayer et al., 2012; Acedo et al., 2015). Spliceogenic variants in +5 positions have also been reported for other disease-responsible genes such as *CFTR* c.2657+5G>A (exon 14b skipping) (Highsmith et al., 1997) and *MLH1* c.116+5G>C (retention of 227 intronic nucleotides) (Arnold et al., 2009). These results highlight the need to study potential splicing alterations beyond consensus positions GT-AG, and that special caution must be taken when relying on *in silico* predictions to detect potential spliceogenic variants located in GC-AG introns.

In summary, our splicing analysis performed in three independent carriers show that *BRCA2* c.7976+5G>T alleles produce a major in-frame transcript Δ17 predicted to encode a non-functional protein, and that even though variable proportions of additional transcripts (Δ16-18, Δ17,18, Δ18 and ▼17q[224]) have been detected, none of them are predicted to rescue BRCA2 functionality.

According to ACMG-AMP guidelines (Richards et al., 2015), our variant qualifies for categories 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 (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium") and PP4 ("Patient's phenotype or family history is highly specific for a disease with a single genetic etiology"). Combining these criteria, the variant is classified as likely pathogenic.

Similarly, following rigorously ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles; https://enigmaconsortium.org/) classification criteria, the variant should be classified as likely pathogenic (Class 4): "variant considered extremely likely to alter splicing based on position, and is untested for splicing aberrations using *in vitro* assays of patient RNA that assesses allele-specific transcript expression, and is predicted bioinformatically to alter use of the native donor/acceptor site, and is not predicted or known to alter production of (naturally occurring) in-frame RNA isoforms that may rescue gene functionality". However, the variant also meets the ENIGMA requirements to be considered pathogenic (Class 5): "Variant allele tested for mRNA aberrations using *in vitro* assays of patient RNA that assesses allele-specific transcript expression, and is found to produce only transcript(s) carrying a premature termination codon, or an in-frame deletion disrupting expression of one or more known clinically important residues", with the exception that our *in vitro* allele-specific analysis was performed in a minigene instead of patient RNA. Current ENIGMA guidelines do not consider construct-based mRNA assays alone as a sufficiently robust approach to be used as evidence for variant classification. However, in this study we used a validated minigene (MGBR2_ex14-20) which confers high reproducibility of splicing patterns as previously described in Fraile-Bethencourt et al., 2017. More specifically, authors analyzed variants involving exon 17 (c.7806-9T>G, c.7975A>G, c.7976G>A and c.7976G>C) using the MGBR2_ex14-20 minigene and compared the splicing patterns with patient RNA results published in previous works, and identified the same splicing results for both approaches in all cases. Furthermore, our study by semi-quantitative methods in patient RNA also supports a pathogenic role for the variant given that carriers generate a predominant aberrant transcript, with any other evidence of transcripts that could rescue protein function. In all, we consider that our results obtained with different methodologies are

in agreement and robust enough to support the classification of this variant as pathogenic (Class-5).

The clinical interpretation of inter-individual differences in isoform expression levels is challenging and whether they are true features associated with the variant of interest, or just reflect biological or technical variability, cannot be concluded from our study. Genetic variants have the potential to generate complex splicing profiles when located in genomic regions with high levels of alternative splicing, and these profiles can be even more complicated to interpret when common variants that modulate isoform levels are present. A comprehensive characterization is therefore always required in these cases, and it is worth to consider whether the individual genetic makeup may have a role in modulating alternative splicing and cancer risk.

Acknowledgments

The authors acknowledge CELLEX Foundation for providing research facilities and equipment.

Disclosure statement: The authors declare no conflict of interest.

References

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

Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA, Tavtigian S V., Goldgar DE, et al. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. Hum Mutat 30:757–770.

Biswas K, Das R, Alter BP, Kuznetsov SG, Stauffer S, North SL, Burkett S, Brody LC, Meyer S, Byrd RA, Sharan SK. 2011. A comprehensive functional characterization of BRCA2 variants associated with Fanconi anemia using mouse ES cell – based assay. Blood

118:2430–2442.

Brandão RD, Roozendaal K Van, Tserpelis D, García EG, Blok MJ. 2011. Characterisation of unclassified variants in the BRCA1/2 genes with a putative effect on splicing. Breast Cancer Res Treat 129:971–982.

Cartegni L, Chew SL, Krainer AR. 2002. Listening To Silence and Understanding Nonsense : exonic mutations that affect splicing. 3:285–298.

Churbanov A, Winters-Hilt S, Koonin E V, Rogozin IB. 2008. Accumulation of GC donor splice signals in mammals. Biol Direct 3:30.

Davy G, Rousselin A, Goardon N, Castéra L, Harter V, Legros A, Muller E, Fouillet R, Brault B, Smirnova AS, Lemoine F, la Grange P de, et al. 2017. Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. Eur J Hum Genet 25:1147–1154.

Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian S V, Monteiro ANA, Iversen ES, Couch FJ, Goldgar DE. 2007. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet 81:873–883.

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

Farrugia DJ, Agarwal MK, Pankratz VS, Deffenbaugh AM, Pruss D, Frye C, Wadum L, Johnson K, Mentlick J, Tavtigian S V., Goldgar DE, Couch FJ. 2008. Functional assays for classification of BRCA2 variants of uncertain significance. Cancer Res 68:3523–3531.

Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V, Acedo A, Sanz DJ, Velasco EA. 2017. Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. PLoS Genet 13: e1006691.

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

Highsmith WE, Burch LH, Zhou Z, Olsen JC, Strong T V., Smith T, Friedman KJ, Silverman LM, Boucher RC, Collins FS, Knowles MR. 1997. Identification of a splice site mutation (2789 + 5 G>A) associated with small amounts of normal CFTR mRNA and mild cystic fibrosis. Hum Mutat 9:332–338.

Hofmann W, Horn D, Hüttner C, Classen E, 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.

Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P, Lefol C, Léone M, 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 33:1228–1238.

Kralovicova J, Hwang G, Asplund AC, Churbanov A, Smith CIE, Vorechovsky I. 2011. Compensatory signals associated with the activation of human GC 5′ splice sites. Nucleic Acids Res 39:7077–7091.

Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K-A, Mooij TM, Roos-Blom M-J, Jervis S, Leeuwen FE van, Milne RL, Andrieu N, Goldgar DE, Terry MB, et al. 2017. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. Jama 317:2402–2416.

Manning KS, Cooper TA. 2017. The roles of RNA processing in translating genotype to phenotype. Nat Rev Mol Cell Biol 18:102–114.

Parada GE, Munita R, Cerda CA, Gysling K. 2014. A comprehensive survey of non-canonical splice sites in the human transcriptome. Nucleic Acids Res 42:10564–10578.

Rebbeck TR, Friebel TM, Friedman E, Hamann U, Huo D, Kwong A, Olah E, Olopade OI, Solano AR, Teo S-H, Thomassen M, Weitzel JN, et al. 2018. Mutational Spectrum in a Worldwide Study of 29,700 Families with *BRCA1* or *BRCA2* Mutations. Hum Mutat [Epub ahead of print].

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–

423.

Roy R, Chun J, Powell SN. 2011. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat Rev Cancer 12:68–78.

Sibley CR, Blazquez L, Ule J. 2016. Lessons from non-canonical splicing. Nat Rev Genet 17:407–421.

Thanaraj TA and CF. 2001. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res 29:2581–2593.

Thirthagiri E, Lee SY, Kang P, Lee DS, Toh GT, Selamat S, Yoon S-Y, Taib N a M, Thong MK, Yip CH, Teo SH. 2008. Evaluation of BRCA1 and BRCA2 mutations and risk-prediction models in a typical Asian country (Malaysia) with a relatively low incidence of breast cancer. Breast Cancer Res 10:R59.

Vreeswijk MPG, Kraan JN, Klift HM Van Der, Vink GR, Cornelisse CJ, Wijnen JT, Bakker E, Asperen CJ Van, Devilee P. 2009. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. Hum Mutat 30:107–114.

Whiley PJ, Guidugli L, Walker LC, Healey S, Thompson BA, Lakhani SR, Silva LM Da, Investigators Kc, Tavtigian S V., Goldgar DE, Brown MA, Couch FJ, 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.

Whiley PJ, La Hoya M De, Thomassen M, Becker A, Brandão R, Pedersen IS, Montagna M, Menéndez M, Quiles F, Gutiérrez-Enríquez S, Leeneer K De, Tenés A, et al. 2014. Comparison of mRNA splicing assay protocols across multiple laboratories: Recommendations for best practice in standardized clinical testing. Clin Chem 60:341–352.

Wu K, Hinson SR, Ohashi A, Farrugia D, Wendt P, Tavtigian S V, Deffenbaugh A, Goldgar D, Couch FJ. 2005. Functional Evaluation and Cancer Risk Assessment of BRCA2 Unclassified Variants. Cancer Res 65:417–426.

Zhang MQ. 1998. Statistical features of human exons and their flanking regions. Hum Mol Genet 7:919–932.

Figure Legends:

Figure 1. Capillary electrophoresis (CE) and Sanger electropherograms from *BRCA2* c.7976+5G>T analysis in HUVH, HCSC and IBGM samples. A. In HUVH samples, exon 17 skipping (in red) was only detected in variant carrier. Full-length (FL) transcript and isoforms Δ18, Δ17,18 and Δ16-18 were detected in carrier and controls but at different expression levels. **B.** CE from HCSC samples showed exon 17 skipping as a major splicing alteration, and also detected isoforms Δ17,18, Δ18 and ▼17q^{224}. **C.** IBGM carrier also showed exon 17 skipping as a major aberrant transcript, without the presence of any other isoforms. **D.** CE from minigene assays only detected Δ17 in c.7976+5G>T and c.7976+1G>A constructs.

Figure 2. Semi-quantitative and quantitative analysis of *BRCA2* transcripts in HUVH and HCSC samples. A. Splicing fraction (SF) mean of each transcript detected in carriers and controls. In both carriers, exon 17 skipping was the predominant splicing event, and additional isoforms were identified at different expression levels. **B.** Semi-quantitative analysis of full-length transcript using normalized CE data. HUVH samples (carrier and controls) were assessed in three independent RT-PCR experiments; HCSC carrier was analyzed in 5 independent RT-PCRs and HCSC controls (n=34) were assessed in one RT-PCR experiment. The grid line represents the average of normalized data from control samples (y=1) and can be used as a reference to compare FL expression levels between samples**. C.** Semi-quantitative analysis of isoform Δ17,18 using normalized CE data. The highest levels were observed in HUVH carrier. **D.** Quantitative analysis of Δ17,18 with digital PCR. The highest levels were again observed in HUVH carrier. Mean and ± SEM are shown in all semi-quantitative and quantitative experiments.

Characterization of spliceogenic variants located in regions linked to high levels of alternative splicing: *BRCA2* c.7976+5G>T as a case study

Gemma Montalban[1]*, Eugenia Fraile-Bethencourt[2]*, Irene López-Perolio[3], Pedro Pérez-Segura[3], Mar Infante[4], Mercedes Durán[4], María Concepción Alonso-Cerezo[5], Adrià López-Fernández[6], Orland Diez[1,7], Miguel de la Hoya[3], Eladio A. Velasco[2]**, Sara Gutiérrez-Enríquez[1]**

*Authors contributed equally to this work

**Corresponding authors

Affiliations:

[1]Oncogenetics Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

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

[3]Molecular Oncology Laboratory CIBERONC, Hospital Clinico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Madrid, Spain

[4]Cancer Genetics, Instituto de Biología y Genética Molecular (CSIC-UVa), Valladolid, Spain

[5]Genética Clínica. Servicio Análisis Clínicos. Hospital Universitario de la Princesa, Instituto de Investigación Sanitaria Hospital Universitario de la Princesa, Madrid, Spain

[6]High Risk and Cancer Prevention Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

[7]Area of Clinical and Molecular Genetics, University Hospital of Vall d'Hebron, Barcelona, Spain

Corresponding authors:

Sara Gutiérrez-Enríquez

Oncogenetics Group, Vall d'Hebron Institute of Oncology-VHIO

Lab 2.02A, CELLEX CENTER

C/ Natzaret, 115-117

08035 Barcelona, Spain

Phone: +34 932543450 ext. 8668

E-mail: sgutierrez@vhio.net

Eladio A. 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

Phone: +34 983184829

Fax: +34 983184800

E-mail: eavelsam@ibgm.uva.es

Funding:

This work was supported by Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds: FIS PI15/00355 (to O. Diez), PI13/01711 and PI16/01218 (to S. Gutiérrez-Enríquez), FIS PI15/00059 (to M. de la Hoya), and PI13/01749 and PI17/00227 (to EA Velasco). E Fraile-Bethencourt was supported by a predoctoral fellowship from the University of Valladolid and Banco de Santander (2015-2019). EA Velasco was also funded by grant CSI090U14 from the Consejería de Educación (ORDEN EDU/122/2014), Junta de Castilla y León. S. Gutiérrez-Enríquez is supported by the Miguel Servet Progam (CP10/00617).

Abstract

Many *BRCA1* and *BRCA2* (*BRCA1/2*) genetic variants have been studied at mRNA level and linked to hereditary breast and ovarian cancer due to splicing alteration. *In silico* tools are reliable when assessing variants located in consensus splice sites, but we may identify variants in complex genomic contexts for which bioinformatics is not precise enough. In this study, we characterize *BRCA2* c.7976+5G>T variant located in intron 17 which has an atypical donor site (GC). This variant was identified in three unrelated Spanish families and we have detected exon 17 skipping as the predominant transcript occurring in carriers. We have also detected several isoforms (Δ16-18, Δ17,18, Δ18 and ▼17q^{224}) at different expression levels among carriers and controls. This study remarks the challenge of interpreting genetic variants when multiple alternative isoforms are present, and that caution must be taken when using *in silico* tools to identify potential spliceogenic variants located in GC-AG introns.

Key words: Hereditary breast and ovarian cancer, *BRCA2*, atypical splicing site, alternative splicing, clinical classification

BRCA1 (MIM# 113705) and *BRCA2* (MIM# 600185) (*BRCA1/2*) genes are associated to hereditary breast and ovarian cancer syndrome (HBOC). Pathogenic variants in *BRCA1/2* result in an increased cumulative breast cancer risk to age 80 that ranges from 61% to 79%, and in an increased risk for ovarian cancer that ranges from 11% to 53% (Kuchenbaecker et al., 2017). Genetic variants in disease-responsible genes that disrupt the splicing code have a key role in human hereditary disorders and cancer. A recent worldwide study describing the mutational spectrum of *BRCA1*/2 genes in HBOC families, identified that 10.1% of *BRCA1* and 7.6% of *BRCA2* pathogenic variants result in aberrant mRNA splicing (Rebbeck et al., 2018). Splicing mutations were traditionally considered those that affect consensus splice sites (intronic nucleotides +1 and +2 of the donor GT, and -1 and -2 of the acceptor AG sites), but other intronic and exonic nucleotides outside these regions have also been found to be highly conserved and critical for splice site selection (Cartegni et al., 2002; Manning and Cooper, 2017). More than 99% of human introns are flanked by GT-AG splice site dinucleotides and are spliced by the so-called major U2-type spliceosome. An exception to this rule are the U2-type GC-AG introns, such as *BRCA2* intron 17 (see Supp. Figure S1), comprising about 0.9% of all human splice sites (reviewed by Parada et al., 2014; Sibley et al., 2016). GC-AG introns possess weak donor sites that are compensated with strong consensus in the surrounding nucleotides of the donor, and are usually linked to alternatively spliced exons (Thanaraj and Clark, 2001; Churbanov et al., 2008; Kralovicova et al., 2011). Given the complexity of these introns, we aimed to characterize the splicing impact of *BRCA2* c.7976+5G>T variant, located at position +5 from the atypical donor site of intron 17. To our knowledge, this variant is not present in the gene-specific databases LOVD (www.lovd.nl), BRCA Exchange (brcaexchange.org) and BIC (https://research.nhgri.nih.gov/bic), ascertained by February 2018. This variant is reported twice in ClinVar database and categorized as conflicting

(https://www.ncbi.nlm.nih.gov/clinvar/), it is reported as a variant of unknown significance (VUS) in BRCA Share (www.umd.be), and it is classified as likely benign (Class-2) in a previous work published by Garibay et al., 2014.

We identified this variant in three unrelated HBOC Spanish families and initiated this collaborative study to exhaustively re-evaluate the variant across three laboratories: Hospital Universitari Vall d'Hebron (HUVH), Hospital Clínico San Carlos (HCSC) and Instituto de Biología y Genética Molecular (IBGM). Probands underwent genetic counselling and written informed consent was obtained in all cases. Family information and pedigrees are described in Supp. Figure S2.

BRCA2 c.7976+5G>T was first assessed *in silico* using Human Splicing Finder (HSF) and MaxEntScan (MES) (http://www.umd.be/HSF3/), which predicted a reduction of the native donor splice site (11.66% and 74.19%, respectively). However, when using the splicing module of Alamut software v2.10 (Interactive Biosoftware) only SSF-like computed a score, predicting a 12.8% reduction of the donor site (Supp. Table S1). Breast Cancer Genes Prior Probabilities website (http://priors.hci.utah.edu/PRIORS/index.php) also predicted a high probability of pathogenicity (0.97) due to splice site donor damage. *In vitro* characterization was performed with patient RNA and a minigene system (see Supp. Methods and Supp. Table S2 for detailed protocols used in each laboratory). HUVH samples were analyzed by RT-PCR using primers located in exons 15 and 19. We detected 5 transcripts corresponding to the reference full-length (FL) (906bp), Δ18 (551bp), Δ17 (735bp), Δ17,18 (380bp) and Δ16-18 (192bp) (Supp. Figure S3). All transcripts were detected in carrier and control samples (n=10), with the exception of Δ17 which was only present in the carrier. Capillary electrophoresis of fluorescent amplicons (CE) ruled out any expression of Δ17 in control samples, indicating that only the variant allele generates this transcript (Figure 1A). HCSC samples were analyzed using primers located in exons 16 and 19. In variant carrier, RT-PCR

assays (Figure 1B) showed transcripts corresponding to FL, Δ17, Δ17,18, and an additional peak of ≈900nt that we tentatively annotated as ▼17q^{224} (Supp. Figure S4). In controls (n=34), FL was detected in all samples, Δ17 was absent in all samples, and isoforms Δ17,18, Δ18 and the putative ▼17q^{224} were detected in 11, 3 and 6 samples, respectively. IBGM carrier was analyzed with primers located in exons 16 and 19 and RT-PCRs showed FL and Δ17 transcripts, without evidence of any other aberrant transcripts or isoforms (Figure 1C). In summary, we identified three frameshift isoforms Δ17,18 (p.Ala2603Phefs*43), Δ18 (p.Tyr2660Phefs*43) and ▼17q^{224} (p.Arg2659Argfs*3) occurring in carrier and control samples; one in-frame isoform Δ16-18 (p.Leu2540_Lys2777del) detected in carrier and control samples; and one in-frame isoform Δ17 (p.Ala2603_Arg2659del) detected only in carriers (results are summarized in Supp. Table S3). All transcripts were confirmed by Sanger sequencing with the exception of Δ16-18 and ▼17q^{224} which were imputed based on the length of the product observed.

Semi-quantitative CE analysis was performed in HUVH and HCSC samples (IBGM patient sample was not available) and showed that Δ17 represents a substantial contribution to the total splicing fraction (SF) in HUVH carrier and HCSC carrier (average SF 34.3% and 48.3%, respectively), whereas in controls was not detected (Figure 2A). Regarding alternative transcripts, notable differences were observed in isoform Δ17,18 levels between HUVH and HCSC carriers (SF 26% and 9.6%, respectively), whereas HUVH and HCSC controls had similar levels (SF 3.3% and 3.9%, respectively). Isoform Δ18 is a minor event detected only in HUVH carrier (SF 7.2%) and in HUVH/HCSC controls (SF 4.2% and 1.4%, respectively); isoform Δ16-18 was only detected in HUVH due to primer location and displayed no notable differences between carrier and controls (SF 13.1% and 11.1%, respectively); and putative isoform ▼17q^{224} was only detected in HCSC samples, although not in all RT-PCR assays, and showed higher levels in HCSC carrier (SF 17.1%) compared to controls (1.2%) (Figure

2A). Two previous studies detected Δ17,18 and Δ18 in control lymphoblastoid cell lines (LCLs) and normal breast tissue, with Δ17,18 being more abundant (Fackenthal et al., 2016; Davy et al., 2017). Normalized CE data from full-length transcript (FL) showed a 2-fold reduction in carriers compared to controls, suggesting that the variant allele is not producing FL (Figure 2B). To test this, since allele-specific assays could not be performed due to the lack of heterozygous informative loci in patient sample, the mutant allele (c.7976+5G>T) was artificially interrogated using a pSAD-derived minigene with *BRCA2* exons 14 to 20, constructed and functionally validated as previously described in Fraile-Bethencourt et al., 2017. A wild-type (wt) construct and a variant construct *BRCA2* c.7976+1G>A were used as negative and positive controls, respectively. Wt construct produced a stable canonical transcript of the expected size and structure, and variant constructs revealed a unique transcript corresponding to exon 17 skipping (Mean ± SEM of relative fluorescence: 1.045 ± 0.111 and 0.955 ± 0.020 for variant and positive control, respectively) (Figure 1D).

Given that notable differences in isoform Δ17,18 levels were observed among samples (Figure 2A), we aimed to determine whether such differences were due to technical reasons and we used digital PCR as a second approach to measure Δ17,18 levels. Data obtained was consistent with CE data, i.e., the highest value was observed in HUVH carrier (SF ~9% vs. SF ~2% observed in HCSC carrier) (Figures 2C and 2D). A slight variability was also observed among control samples, with SF levels ranging from ~0.5 to 4%. Interestingly, a recent study identified a common *BRCA2* variant c.7806-14T>C (rs9534262) which influences Δ17,18 levels by modulating exon 17 acceptor site (Garibay et al., 2014). DNA sequence analysis revealed that HUVH patient is homozygous (C/C) at this polymorphic position, whereas HCSC and IBGM are heterozygous (T/C) (data not shown). Sanger sequencing of introns surrounding exon 17 (limited to c.7806-37_7806-1 and c.7976+1_7976+25 regions) in HUVH and HCSC carriers, and whole intron 16 and 17

analysis in IBGM carrier, did not identify additional rare or common variants that could explain differential Δ17,18 isoform expression among carriers. In this regard, patient genotypes were consistent with experimental data in that the highest Δ17,18 levels were detected in the homozygous C/C patient. The influence of c.7806-14T>C on Δ17,18 levels was additionally evaluated in HCSC controls (n=34) by RT-PCR and CE, and the highest levels were again observed in C/C samples (Supp. Figure S5). Nevertheless, whether an increase of Δ17,18 is associated with our variant under scrutiny or not, would require the analysis of carriers being c.7806-14T/T to avoid its influence on Δ17,18 levels.

We also re-evaluated HUVH and HCSC RNA samples to explore whether methodological differences between laboratories could influence variability in isoform levels. To do so, carriers and controls were analysed using primers and protocols from HCSC laboratory (see Supp. Table S2). Results obtained showed that major transcriptional events were detected as previously: Δ17 was present in both carriers and absent in controls, and Δ17,18 highest levels were detected in HUVH carrier. Minor isoform Δ18 was again detected in HUVH carrier and absent in HCSC carrier, and ▼17q^{224} was not detected in either carriers (Supp. Figure S6). This data indicates that differences in minor events (Δ18 and ▼17q^{224}) are presumably due to stochastic effects during PCR amplification, but variability in major isoform Δ17,18 is likely to be linked to individual genetic features rather than methodological differences. In this analysis, polymerases and CE conditions did not seem to have an influence in isoform detection, and although new RNA samples could not be obtained to rule out any influence of RNA isolation methods, a collaborative work comparing RNA protocols for characterization of spliceogenic variants across multiple laboratories concluded that RNA extraction methods were indistinguishable (Whiley et al., 2014).

Overall, combined analysis carried out in three different laboratories provides convincing evidence that the major splicing outcome produced by *BRCA2* c.7976+5G>T variant is exon

17 skipping, which causes an in-frame deletion (r.7806_7976del171) that results in a protein lacking 57 amino acids (aa) (p.Ala2603_Arg2659del). The lost region is part of the α-helical domain (aa 2479 to 2667) of BRCA2 DNA Binding Domain (DBD), which has 30 strictly conserved residues from sea urchin to human (Supp. Figure S7). This domain enables BRCA2 binding to single-stranded and double-stranded DNA, and is essential to allow DNA repair by homologous recombination (HR) (Roy et al., 2011). Functional assays for classification of missense variants located in this region confirmed a reduction of BRCA2 HR activity for Trp2626Cys, Ile2627Phe, Leu2647Pro, Leu2653Pro and Arg2659Lys variants (Farrugia et al., 2008; Biswas et al., 2011). From these, variants Trp2626Cys, Ile2627Phe, Leu2653Pro and Arg2659Lys had been previously evaluated by multifactorial analysis and classified as pathogenic (Class 5) (Easton et al., 2007). Other variants causing exon 17 skipping (c.7976G>A, c.7976+1G>A and c.7976+3_7976+4del) have been identified in HBOC patients and reported as deleterious (Hofmann et al., 2003; Wu et al., 2005; Thirthagiri et al., 2008; Brandão et al., 2011; Fraile-Bethencourt et al., 2017). Moreover, allele-specific assessment was performed in patient RNA for variant c.7976+3_7976+4del, and only detected transcript lacking exon 17 (Brandão et al., 2011). Although Δ17 was categorized as a minor alternative splicing event occurring in control LCLs and normal breast tissue, it was not detected in whole blood control samples (Fackenthal et al., 2016). Likewise, we did not identify this transcript in our control group (n=44).

Our variant G>T is located at position +5 from intron 17 donor splice site, where a G is present in >80% of human introns (Zhang, 1998). In this particular case, *BRCA2* intron 17 has an atypical donor site GC weaker than the GT counterparts because of the +2 substitution, meaning that the rest of nucleotide positions are more conserved (Thanaraj and Clark, 2001) and that variants in any of these nucleotides may have an impact on exon recognition. However, it is important to note that not all *in silico* approaches used in this study to predict

the splicing impact of *BRCA2* c.7976+5G>T variant were able to detect its potential damage to the native donor site. Other *BRCA1/2* variants located at position +5 have been reported to induce splicing alterations, such as *BRCA1* c.5406+5G>C (exon 22 skipping) and c.5467+5G>C (exon 23 skipping), *BRCA2* c.316+5G>C (exon 3 skipping) and c.8754+5G>A (insertion of 46 nucleotides of intron 21) (Vreeswijk et al., 2009; Whiley et al., 2011; Houdayer et al., 2012; Acedo et al., 2015). Spliceogenic variants in +5 positions have also been reported for other disease-responsible genes such as *CFTR* c.2657+5G>A (exon 14b skipping) (Highsmith et al., 1997) and *MLH1* c.116+5G>C (retention of 227 intronic nucleotides) (Arnold et al., 2009). These results highlight the need to study potential splicing alterations beyond consensus positions GT-AG, and that special caution must be taken when relying on *in silico* predictions to detect potential spliceogenic variants located in GC-AG introns.

In summary, our splicing analysis performed in three independent carriers show that *BRCA2* c.7976+5G>T alleles produce a major in-frame transcript Δ17 predicted to encode a non-functional protein, and that even though variable proportions of additional transcripts (Δ16-18, Δ17,18, Δ18 and ▼17q^{224}) have been detected, none of them are predicted to rescue BRCA2 functionality.

According to ACMG-AMP guidelines (Richards et al., 2015), our variant qualifies for categories 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 (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium") and PP4 ("Patient's phenotype or family history is highly specific for a disease with a single genetic etiology"). Combining these criteria, the variant is classified as likely pathogenic.

Similarly, following rigorously ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles; https://enigmaconsortium.org/) classification criteria, the variant should be classified as likely pathogenic (Class 4): "variant considered extremely likely to alter splicing based on position, and is untested for splicing aberrations using *in vitro* assays of patient RNA that assesses allele-specific transcript expression, and is predicted bioinformatically to alter use of the native donor/acceptor site, and is not predicted or known to alter production of (naturally occurring) in-frame RNA isoforms that may rescue gene functionality". However, the variant also meets the ENIGMA requirements to be considered pathogenic (Class 5): "Variant allele tested for mRNA aberrations using *in vitro* assays of patient RNA that assesses allele-specific transcript expression, and is found to produce only transcript(s) carrying a premature termination codon, or an in-frame deletion disrupting expression of one or more known clinically important residues", with the exception that our *in vitro* allele-specific analysis was performed in a minigene instead of patient RNA. Current ENIGMA guidelines do not consider construct-based mRNA assays alone as a sufficiently robust approach to be used as evidence for variant classification. However, in this study we used a validated minigene (MGBR2_ex14-20) which confers high reproducibility of splicing patterns as previously described in Fraile-Bethencourt et al., 2017. More specifically, authors analyzed variants involving exon 17 (c.7806-9T>G, c.7975A>G, c.7976G>A and c.7976G>C) using the MGBR2_ex14-20 minigene and compared the splicing patterns with patient RNA results published in previous works, and identified the same splicing results for both approaches in all cases. Furthermore, our study by semi-quantitative methods in patient RNA also supports a pathogenic role for the variant given that carriers generate a predominant aberrant transcript, with any other evidence of transcripts that could rescue protein function. In all, we consider that our results obtained with different methodologies are

in agreement and robust enough to support the classification of this variant as pathogenic (Class-5).

The clinical interpretation of inter-individual differences in isoform expression levels is challenging and whether they are true features associated with the variant of interest, or just reflect biological or technical variability, cannot be concluded from our study. Genetic variants have the potential to generate complex splicing profiles when located in genomic regions with high levels of alternative splicing, and these profiles can be even more complicated to interpret when common variants that modulate isoform levels are present. A comprehensive characterization is therefore always required in these cases, and it is worth to consider whether the individual genetic makeup may have a role in modulating alternative splicing and cancer risk.

Acknowledgments

The authors acknowledge CELLEX Foundation for providing research facilities and equipment.

Disclosure statement: The authors declare no conflict of interest.

References

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

Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA, Tavtigian S V., Goldgar DE, et al. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. Hum Mutat 30:757–770.

Biswas K, Das R, Alter BP, Kuznetsov SG, Stauffer S, North SL, Burkett S, Brody LC, Meyer S, Byrd RA, Sharan SK. 2011. A comprehensive functional characterization of BRCA2 variants associated with Fanconi anemia using mouse ES cell – based assay. Blood

118:2430–2442.

Brandão RD, Roozendaal K Van, Tserpelis D, García EG, Blok MJ. 2011. Characterisation of unclassified variants in the BRCA1/2 genes with a putative effect on splicing. Breast Cancer Res Treat 129:971–982.

Cartegni L, Chew SL, Krainer AR. 2002. Listening To Silence and Understanding Nonsense : exonic mutations that affect splicing. 3:285–298.

Churbanov A, Winters-Hilt S, Koonin E V, Rogozin IB. 2008. Accumulation of GC donor splice signals in mammals. Biol Direct 3:30.

Davy G, Rousselin A, Goardon N, Castéra L, Harter V, Legros A, Muller E, Fouillet R, Brault B, Smirnova AS, Lemoine F, la Grange P de, et al. 2017. Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. Eur J Hum Genet 25:1147–1154.

Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian S V, Monteiro ANA, Iversen ES, Couch FJ, Goldgar DE. 2007. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet 81:873–883.

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

Farrugia DJ, Agarwal MK, Pankratz VS, Deffenbaugh AM, Pruss D, Frye C, Wadum L, Johnson K, Mentlick J, Tavtigian S V., Goldgar DE, Couch FJ. 2008. Functional assays for classification of BRCA2 variants of uncertain significance. Cancer Res 68:3523–3531.

Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V, Acedo A, Sanz DJ, Velasco EA. 2017. Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. PLoS Genet 13: e1006691.

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

Formatted: Highlight

Formatted: Highlight

Highsmith WE, Burch LH, Zhou Z, Olsen JC, Strong T V., Smith T, Friedman KJ, Silverman LM, Boucher RC, Collins FS, Knowles MR. 1997. Identification of a splice site mutation (2789 + 5 G>A) associated with small amounts of normal CFTR mRNA and mild cystic fibrosis. Hum Mutat 9:332–338.

Hofmann W, Horn D, Hüttner C, Classen E, 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.

Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P, Lefol C, Léone M, 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 33:1228–1238.

Kralovicova J, Hwang G, Asplund AC, Churbanov A, Smith CIE, Vorechovsky I. 2011. Compensatory signals associated with the activation of human GC 5′ splice sites. Nucleic Acids Res 39:7077–7091.

Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K-A, Mooij TM, Roos-Blom M-J, Jervis S, Leeuwen FE van, Milne RL, Andrieu N, Goldgar DE, Terry MB, et al. 2017. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. Jama 317:2402–2416.

Manning KS, Cooper TA. 2017. The roles of RNA processing in translating genotype to phenotype. Nat Rev Mol Cell Biol 18:102–114.

Parada GE, Munita R, Cerda CA, Gysling K. 2014. A comprehensive survey of non-canonical splice sites in the human transcriptome. Nucleic Acids Res 42:10564–10578.

Rebbeck TR, Friebel TM, Friedman E, Hamann U, Huo D, Kwong A, Olah E, Olopade OI, Solano AR, Teo S-H, Thomassen M, Weitzel JN, et al. 2018. Mutational Spectrum in a Worldwide Study of 29,700 Families with *BRCA1* or *BRCA2* Mutations. Hum Mutat [Epub ahead of print].

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–

423.

Roy R, Chun J, Powell SN. 2011. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat Rev Cancer 12:68–78.

Sibley CR, Blazquez L, Ule J. 2016. Lessons from non-canonical splicing. Nat Rev Genet 17:407–421.

Thanaraj TA and CF. 2001. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res 29:2581–2593.

Thirthagiri E, Lee SY, Kang P, Lee DS, Toh GT, Selamat S, Yoon S-Y, Taib N a M, Thong MK, Yip CH, Teo SH. 2008. Evaluation of BRCA1 and BRCA2 mutations and risk-prediction models in a typical Asian country (Malaysia) with a relatively low incidence of breast cancer. Breast Cancer Res 10:R59.

Vreeswijk MPG, Kraan JN, Klift HM Van Der, Vink GR, Cornelisse CJ, Wijnen JT, Bakker E, Asperen CJ Van, Devilee P. 2009. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. Hum Mutat 30:107–114.

Whiley PJ, Guidugli L, Walker LC, Healey S, Thompson BA, Lakhani SR, Silva LM Da, Investigators Kc, Tavtigian S V., Goldgar DE, Brown MA, Couch FJ, 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.

Whiley PJ, La Hoya M De, Thomassen M, Becker A, Brandão R, Pedersen IS, Montagna M, Menéndez M, Quiles F, Gutiérrez-Enríquez S, Leeneer K De, Tenés A, et al. 2014. Comparison of mRNA splicing assay protocols across multiple laboratories: Recommendations for best practice in standardized clinical testing. Clin Chem 60:341–352.

Wu K, Hinson SR, Ohashi A, Farrugia D, Wendt P, Tavtigian S V, Deffenbaugh A, Goldgar D, Couch FJ. 2005. Functional Evaluation and Cancer Risk Assessment of BRCA2 Unclassified Variants. Cancer Res 65:417–426.

Zhang MQ. 1998. Statistical features of human exons and their flanking regions. Hum Mol Genet 7:919–932.

Figure Legends:

Figure 1. Capillary electrophoresis (CE) and Sanger electropherograms from *BRCA2* c.7976+5G>T analysis in HUVH, HCSC and IBGM samples. A. In HUVH samples, exon 17 skipping (in red) was only detected in variant carrier. Full-length (FL) transcript and isoforms Δ18, Δ17,18 and Δ16-18 were detected in carrier and controls but at different expression levels. **B.** CE from HCSC samples showed exon 17 skipping as a major splicing alteration, and also detected isoforms Δ17,18, Δ18 and ▼17q^{224}. **C.** IBGM carrier also showed exon 17 skipping as a major aberrant transcript, without the presence of any other isoforms. **D.** CE from minigene assays only detected Δ17 in c.7976+5G>T and c.7976+1G>A constructs.

Figure 2. Semi-quantitative and quantitative analysis of *BRCA2* transcripts in HUVH and HCSC samples. A. Splicing fraction (SF) mean of each transcript detected in carriers and controls. In both carriers, exon 17 skipping was the predominant splicing event, and additional isoforms were identified at different expression levels. **B.** Semi-quantitative analysis of full-length transcript using normalized CE data. HUVH samples (carrier and controls) were assessed in three independent RT-PCR experiments; HCSC carrier was analyzed in 5 independent RT-PCRs and HCSC controls (n=34) were assessed in one RT-PCR experiment. The grid line represents the average of normalized data from control samples (y=1) and can be used as a reference to compare FL expression levels between samples**. C.** Semi-quantitative analysis of isoform Δ17,18 using normalized CE data. The highest levels were observed in HUVH carrier. **D.** Quantitative analysis of Δ17,18 with digital PCR. The highest levels were again observed in HUVH carrier. Mean and ± SEM are shown in all semi-quantitative and quantitative experiments.

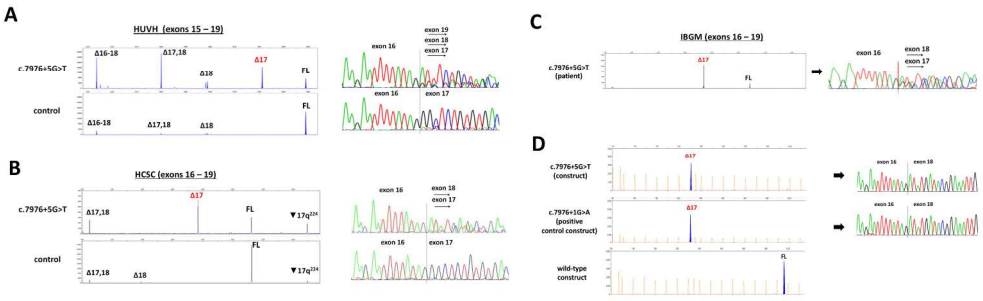


Figure 1. Capillary electrophoresis (CE) and Sanger electropherograms from *BRCA2* c.7976+5G>T analysis in HUVH, HCSC and IBGM samples. A. In HUVH samples, exon 17 skipping (in red) was only detected in variant carrier. Full-length (FL) transcript and isoforms Δ18, Δ17,18 and Δ16-18 were detected in carrier and controls but at different expression levels. B. CE from HCSC samples showed exon 17 skipping as a major splicing alteration, and also detected isoforms Δ17,18, Δ18 and ▼17q^{224}. C. IBGM carrier also showed exon 17 skipping as a major aberrant transcript, without the presence of any other isoforms. D. CE from minigene assays only detected Δ17 in c.7976+5G>T and c.7976+1G>A constructs.

1057x352mm (72 x 72 DPI)

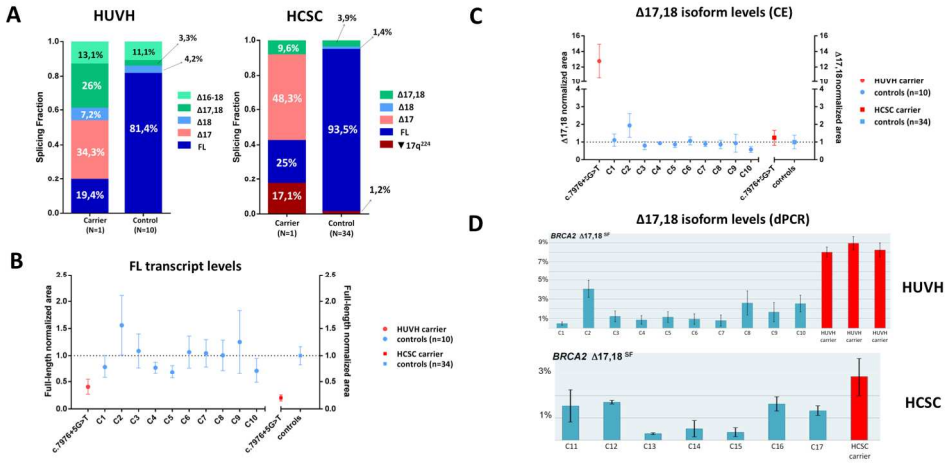


Figure 2. Semi-quantitative and quantitative analysis of *BRCA2* transcripts in HUVH and HCSC samples. A. Splicing fraction (SF) mean of each transcript detected in carriers and controls. In both carriers, exon 17 skipping was the predominant splicing event, and additional isoforms were identified at different expression levels. B. Semi-quantitative analysis of full-length transcript using normalized CE data. HUVH samples (carrier and controls) were assessed in three independent RT-PCR experiments; HCSC carrier was analyzed in 5 independent RT-PCRs and HCSC controls (n=34) were assessed in one RT-PCR experiment. The grid line represents the average of normalized data from control samples (y=1) and can be used as a reference to compare FL expression levels between samples. C. Semi-quantitative analysis of isoform Δ17,18 using normalized CE data. The highest levels were observed in HUVH carrier. D. Quantitative analysis of Δ17,18 with digital PCR. The highest levels were again observed in HUVH carrier. Mean and ± SEM are shown in all semi-quantitative and quantitative experiments.

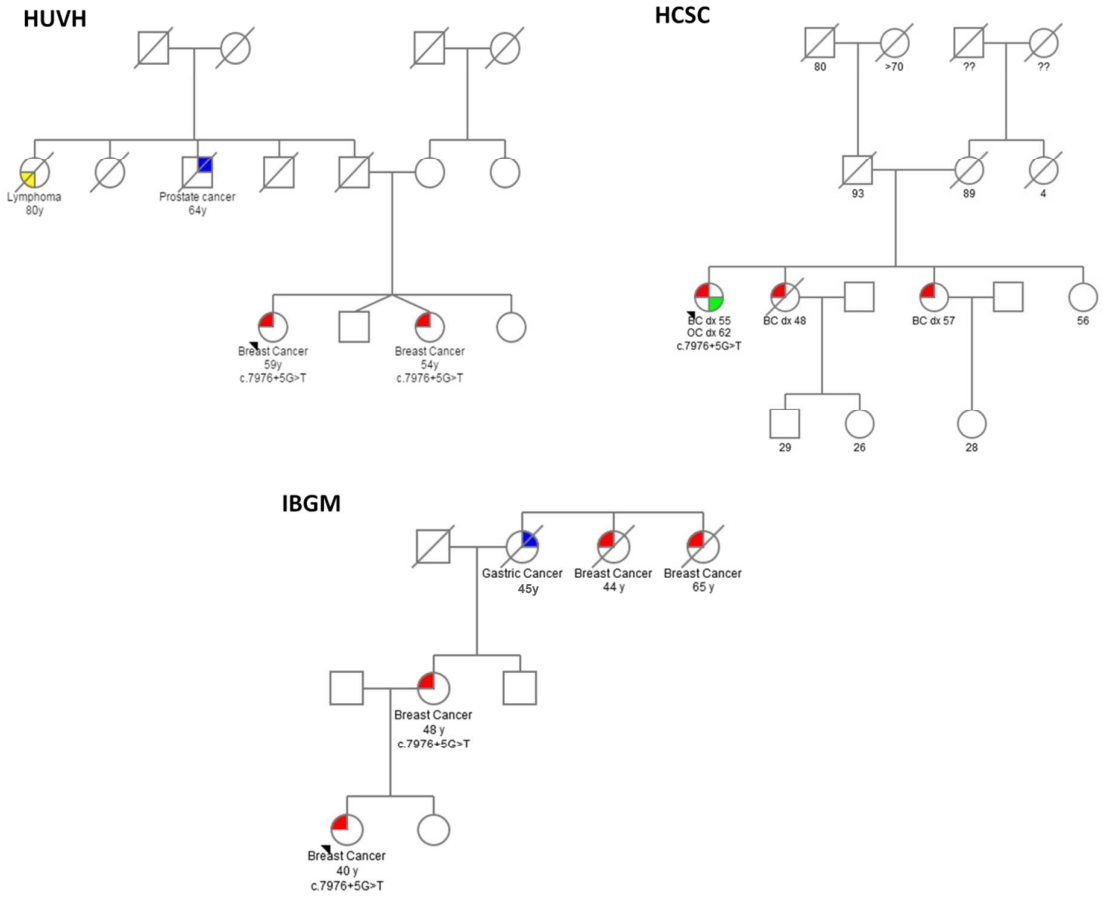
338x166mm (150 x 150 DPI)

exons 2-26

exon 17

Position Exon/Intron	A	C	G	T
-3	0,250	0,375	0,125	0,250
-2	**0,708**	0,125	0,042	0,125
-1	0,000	0,000	**0,958**	0,048
+1	0,000	0,000	**1,000**	0,000
+2	0,000	0,000	0,000	**1,000**
+3	**0,958**	0,000	0,042	0,000
+4	**0,625**	0,125	0,000	0,250
+5	0,208	0,042	**0,542**	0,208
+6	0,375	0,042	0,167	0,417

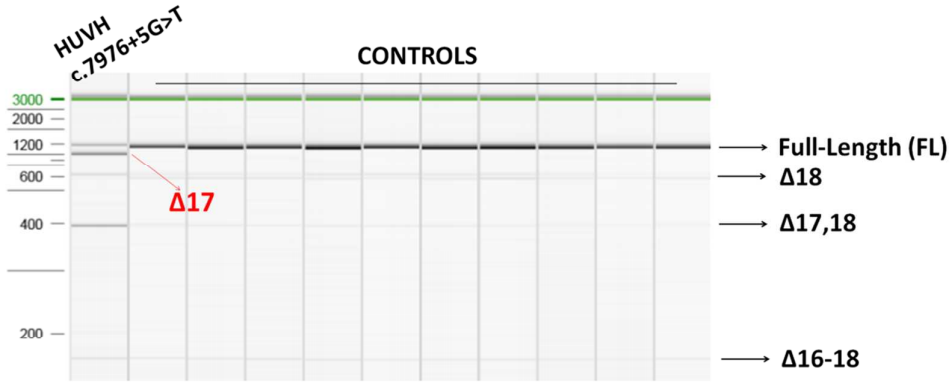
Supp. Figure S1. Pictograms of *BRCA2* donor splice sites (5') and nucleotide frequency at each position. Upper figures represent donor splice sites of exons 2-26 (except exon 17), and exon 17. Letter size represents the frequency of the nucleotide, which is summarized in the table (exon 17 not included). Discontinue black line indicates exon-intron limit. Pictograms were built using http://genes.mit.edu/pictogram.html



Supp. Figure S2. Family pedigrees and clinical information. Red coloured symbols indicate individuals affected with breast cancer (BC) and green coloured indicate ovarian cancer (OC). Other types of cancer (prostate, gastric and lymphoma) are represented in colours blue and yellow. Probands are marked by arrowheads. Pedigrees were built using the progeny free online pedigree tool (https://pedigree.progenygenetics.com). HUVH family has a southern Spain origin. The proband was diagnosed with breast cancer (BC) (right breast: infiltrating ductal carcinoma, triple negative ER-, PgR-, HER2-) at the age of 59. Her sister was also diagnosed at 54 with two BC (right breast: an infiltrating ductal carcinoma and a ductal carcinoma in situ), and their paternal uncle died at 64 of prostate cancer. HCSC family comes from north-western area of Spain (Castilla y León). The proband was diagnosed at 55

with BC (left breast: ductal carcinoma in situ, triple negative) and with ovarian cancer (OC) at age 62. Two out of her three sisters were also diagnosed with BC at ages of 48 and 57, respectively. The family analysed at IBGM is also from Castilla y León and the proband was diagnosed with a hormone-receptor-positive BC at 40 years of age. Her mother and two of her maternal great aunts were diagnosed with BC at 48, 44 and 65 years of age, respectively. Her maternal grandmother was diagnosed with gastric cancer at 45 years of age.

In HUVH family, DNA genetic testing identified *BRCA2* c.7976+5G>T variant in the proband and her sister, both affected with BC. In IBGM family, the variant was detected in the proband and her mother.



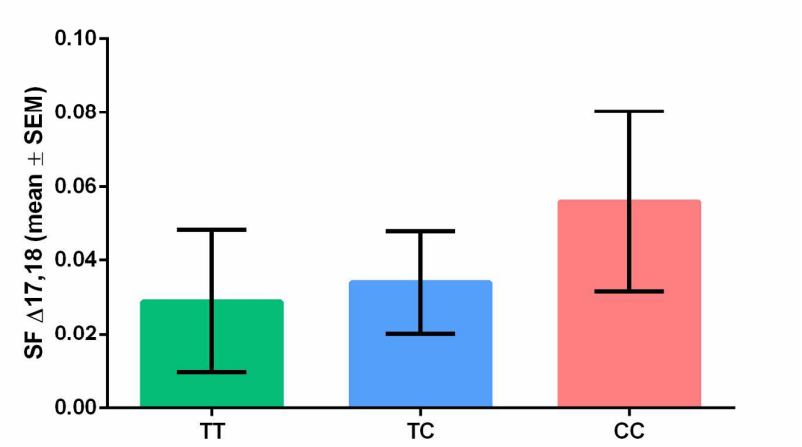
Supp. Figure S3. Splicing analysis of *BRCA2* c.7976+5G>T variant in HUVH samples. QIAxcel electrophoresis from HUVH carrier and 10 control samples. Black arrows indicate the full-length (FL) transcript (906bp) and isoforms Δ18 (551bp), Δ17,18 (380bp) and Δ16-18 (192bp), all present in variant carrier and controls but at different expression levels. The red arrow highlights the transcript Δ17 (735bp) only present in the variant carrier.

TGCTAATAGATGCCTAAGCCCAGAAAGGGTGCTTCTTCAACTAAAATACAGgcaagtttaaagcattacattacgtaatc
atatacggcagtatggttaaggtttctgtgtagtctgtgacttccatgtcaaaatgttgcacaagccagttgtcagtgacagttgccatcccacact
gctgttctcctgtcatccctagccccatttaagagagatcacacattcatgcattgcttgcttccctctttccccacccccctccttaacctctt**gatgt**
atgagaagaatatgagttactaatttgatccactatttggggattgctaataaagcatttttgcattttatttttgcttttaaaaataattgatatttt
aacaatatgaaacaatatattcctagctacaaaatttttaattctcagtatttcttagataaattcagtttttattctcagttattcagtgacttgtttaa
acagtggaattctagagtcacacttcctaaaatatgcatttttgttttcacttttag**ATATGATACGGAAATTGATAGAAGCAGAAGA**
TCGGCTATAAAAAAGATAATGGAAAGGGA

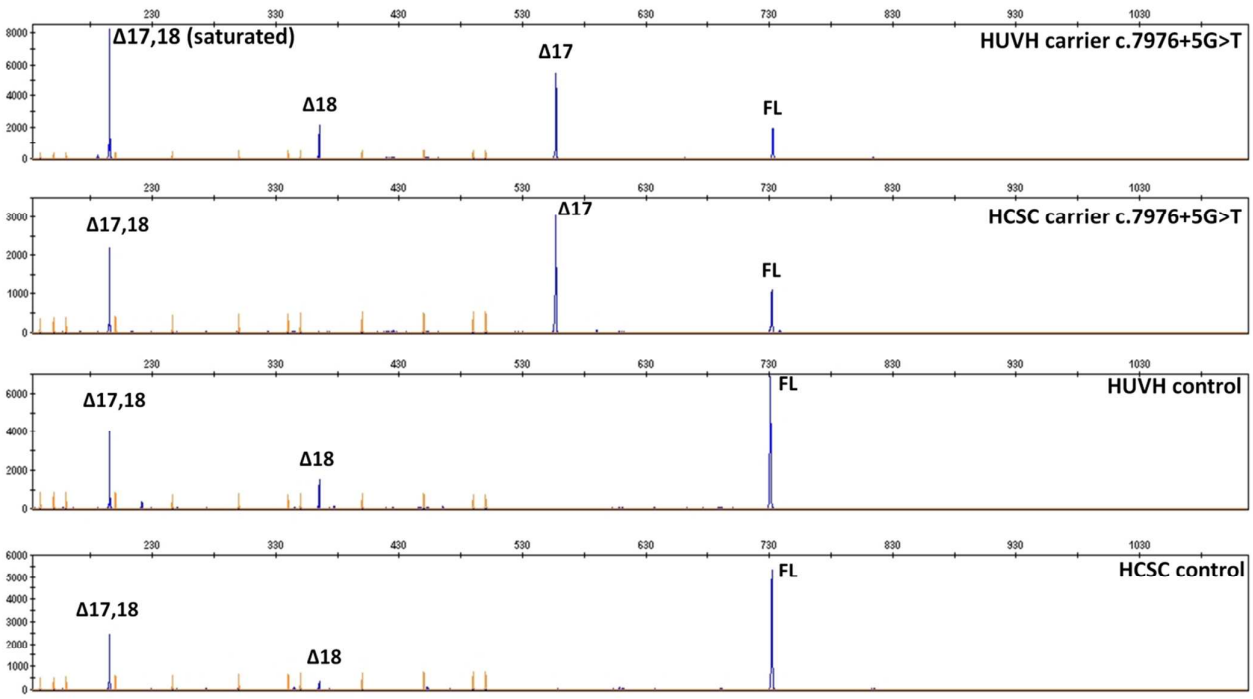
595nt

Donor site	MaxEnt Score (5' Motif)	Seq	Predicted Size (nt) RT-PCR E16-E19	Designation
Exon 17 native site	3.1	CAGgcaagt	706	FL
Exon 17 variant site	0.8	CAGgcaatt	535	Δ17
Cryptic site 1	1.15	GCAgtatgg	745	-
Cryptic site 2	0.58	TGgttaagg	751	-
Cryptic site 3	0.58	AAGgtttct	756	-
Cryptic site 4	**2.41**	**GATgtatga**	**930**	**▼17q^{224}**

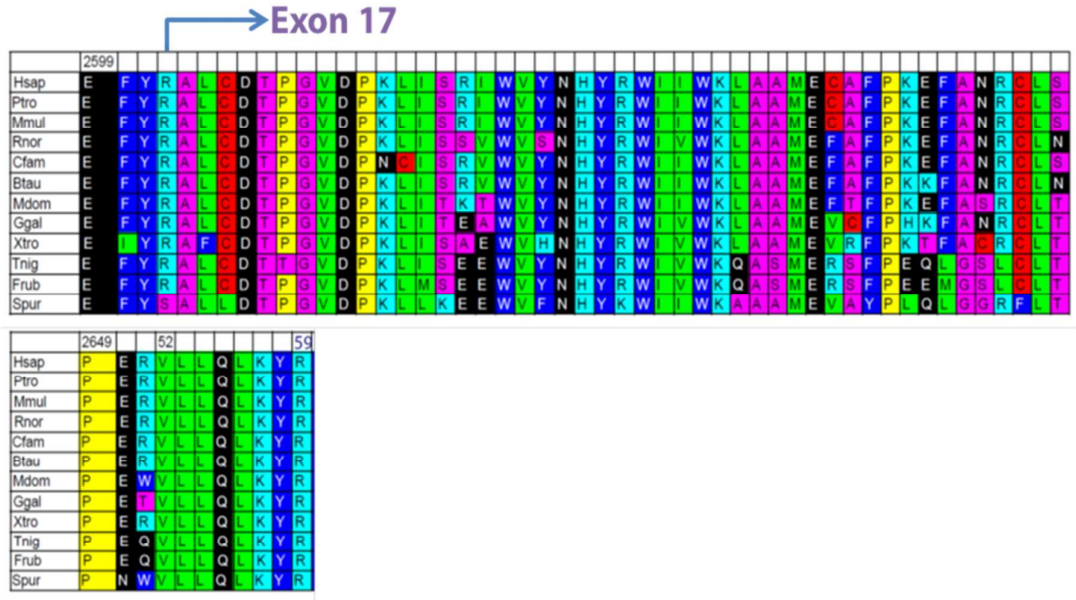
Supp. Figure S4. Cryptic splice sites analysis with MaxEnt tool. A. The 595nt sequence indicated above (which include 51nt at the 3' end of exon 17, the complete 485nt of intron 17, and 60nt at the 5' end of exon 18) were analyzed with MaxEnt (http://www.umd.be/HSF3/). Apart from the native donor site (3.1), the analysis identified four potential cryptic donor sites, from which only the cryptic sites 1 (1.15) and 4 (2.41) score above the variant allele donor site (0.8). The latter predicts that cryptic sites 1 and 4 might be preferentially used in variant alleles. RT-PCR experiment covering exons 16-19 in HCSC samples (Figure 1B) detected peaks compatible with the use of cryptic site 4 (▼17q^{224}) both in carrier and controls (Figure 1B) albeit the average signal appears to be higher in the carrier (Figure 2A). The use of putative cryptic sites 1, 2 and 3 was not detected in either carriers or controls.



Supp. Figure S5. Isoform Δ17,18 levels measured in 34 controls genotyped at position c.7806-14T>C. Healthy controls used for variant analysis in HCSC laboratory were genotyped at polymorphic position c.7806-14T>C (10 T/T, 14 T/C and 10 C/C), known to modulate Δ17,18 levels. Semi-quantitative CE data (Mean ± SEM) shows higher levels of Δ17,18 in CC samples, consistent with previously published results in de Garibay et al., 2014.



Supp. Figure S6. CE analysis of HUVH and HCSC samples. Both carriers and two controls were analyzed using the same primer pair, cycling conditions and electrophoresis protocol from HCSC laboratory. Major transcripts were again detected (Δ17 and Δ17,18), whereas minor transcripts (Δ18 and ▼17q[224]) were not detected in all RT-PCR assays as previously observed in when samples were analyzed independently in each laboratory (see Figures 1A and 1B, and Supp. Table S3).



Supp. Figure S7. BRCA2 protein alignment. Alignment of BRCA2 region encoded by exon17 showing more than 50% residues conserved.

Supp. Table S1. *In silico* analysis of *BRCA2* c.7976+5G>T variant.

		Wild-type sequence	Variant sequence	Native donor strength reduction (%)
HSF3 website	**MES [0-12]**	3.10	0.8	74.19
	HSF [0-100]	100	88.34	11.66
Alamut biosoftware v2.10	**SSF-like [0-100]**	100	87.25	12.8
	MES [0-12]	-	-	-
	NNSPLICE [0-1]	-	-	-
	GeneSplicer [0-24]	-	-	-
	HSF [0-100]	-	-	-

Bioinformatics predictions were obtained from Human Splicing Finder (HSF) and MaxEntScan (MES) through Human Splicing Finder v3.0 website (http://www.umd.be/HSF3/). We also used the splicing module of Alamut biosoftware v2.10 (Interactive Biosoftware) which includes SpliceSiteFinder-like (SSF), MaxEntScan (MES), NNSPLICE, GeneSplicer and Human Splicing Finder (HSF) tools. Score ranges are detailed between brackets. Both MES and HSF tools predicted a reduction of the native donor site, however when using Alamut only SSF was able to compute a score.

Supp. Table S2. Summary of the methodology used in each participant laboratory to characterize *BRCA2* c.7976+5G>T variant.

Laboratory	RNA source	Samples	RNA isolation	cDNA synthesis	PCR polymerase	PCR primers	FL size	PCR conditions	Detection and Quantification	CE conditions	CE size marker
HUVH	5 mL whole blood	c.7976+5G>T carrier 10 wt controls	Trizol reagent (ThermoFisher) + Ambion DNA-free kit (Life Technologies)	**Protocol:** PrimeScript RT Reagent Kit (Takara) **Primers**: oligodT + random **RNA amount:** 500ng	EcoTaq DNA Polymerase (Ecogen)	**exon 15F(FAM) -** CCACTCTGCCTCGAATCTCT **exon 19R -** CACAACCAACATTTCCTCCA	906bp	x35 cycles 1min extension time	QIAxcel Sanger sequencing Semi-quantitative CE	**Array length:** 36cm **Electrophoresis protocol:** temperature 60ºC, 15sec injection at 1.2KVolts **Run protocol**: 2000sec at 12Kvolts	GeneScan 1000 ROX dye (ThermoFisher)
HCSC	1 mL whole blood	c.7976+5G>T carrier 34 wt controls	Magna Pure RNA kit (Roche) + QIAmp RNA blood Mini kit (QIAGEN)	**Protocol:** PrimeScript RT Reagent Kit (Takara) **Primers:** oligodT + random **RNA amount**: 800ng	FastStart Taq (Roche Diagnostics)	**exon 16F (FAM) -** GAGTCTTTTCAGTTTCACACTG **exon 19R -** CAAGTTTGGTATACCAGCGAG	700bp	x35 cycles 1min extension time	Agarose gel Sanger sequencing Semi-quantitative CE Digital PCR	**Array length**: 50cm **Electrophoresis protocol**: temperature 60ºC, 15sec injection at 1.2KVolts **Run protocol:** 1800sec at 15Kvolts	GeneScan 500 LIZ dye (ThermoFisher)
IBGM	1.5 mL whole blood	c.7976+5G>T carrier wt control	QIAamp RNA blood mini kit (QIAGEN)	**Protocol:** Transcriptor first strand cDNA synthesis kit (Roche) **Primers:** oligodT + random hexamers **RNA amount:** 200ng	GoTaq Hot Start DNA polymerase (Promega)	**exon 16F (FAM) -** TATGGACTGGAAAAGGAATAC **exon 19R -** AGCGATGATAAGGGCAGAG	706bp	x35 cycles 1min extension time	Agarose gel Sanger sequencing	**Array length**: 50cm **Electrophoresis protocol**: temperature 60ºC, 15sec injection at 1.2KVolts **Run protocol:** 1800sec at 15Kvolts	GeneScan 500 LIZ dye (ThermoFisher)
	MCF-7 cells (minigene exons 14-20)	c.7976+5G>T construct c.7976+1G>A construct (positive control) wt construct	Genematrix Universal RNA Purification kit (EURx)	**Protocol:** RevertAid H Minus First Strand cDNA synthesis kit (Life Technologies) **Primers:** gene specific primer **RT-pSPL3-RV** – TGAGGAGTGAATTGGTCGAA **RNA amount:** 400ng	Platinum Taq DNA polymerase (Life Technologies)	**SD6-PSPL3_RTFW -** TCACCTGGACAACCTCAAAG **RTpSAD-RV-** GGAGTGAATTGGTCGAATG	1806bp	x35 cycles 1min extension time	Sanger sequencing	-	-
						RTBR2_ex16FW - TATGGACTGGAAAAGGAATAC **FAM-RT-psPL3-RV-** TGAGGAGTGAATTGGTCGAA	1015bp	x26 cycles 1min extension time	Semi-quantitative CE	**Array length**: 50cm **Electrophoresis protocol**: temperature 60ºC, 15sec injection at 1.6KVolts **Run protocol:** 1600sec at 15Kvolts	GeneScan 1200 LIZ dye (ThermoFisher)

wt = wild-type; FL = Full-length; CE = capillary electrophoresis

Supp. Table S3. RNA results obtained from the characterization of *BRCA2* c.7976+5G>T variant in each laboratory.

Laboratory	RNA source	Sample	PCR fragment	*BRCA2* transcripts detected[a]					
				FL (reference transcript)	Δ17 r.7806_7976del171 (p.Ala2603_Arg2659del)	Δ18 r.7977_8331del355 (p.Tyr2660Phefs*43)	Δ17-18 r.7806_8331del526 (p.Ala2603Phefs*43)	Δ16-18* r.7618_8331del714 (p.Leu2540_Lys2777del)	▼17q224* r.7976_7977ins 7976+1_7976+224 (p.Arg2659Argfs*3)
HUVH	blood	c.7976+5G>T carrier	exons 15 - 19	+	+	+	+	+	-
		10 controls		+	-	+	+	+	-
HCSC	blood	c.7976+5G>T carrier	exons 16 - 19	+	+	-	+	na	+
		34 controls		+	-	+ (3 controls)	+ (11 controls)	na	+ (6 controls)
IBGM	blood	c.7976+5G>T carrier	exons 16 - 19	+	+	-	-	na	-
		1 control		+	-	-	-	na	-
	minigene	c.7976+5G>T construct	exon 16 - vector	-	+	-	-	-	-
		c.7976+1G>T construct		-	+	-	-	-	-
		wt construct		+	-	-	-	-	-

wt = wild-type; FL = full-length; na = non-applicable

[a] RNA and Protein annotations follow HGVS guidelines (den Dunnen et al., 2016). Transcripts were annotated according to *BRCA2* GeneBank sequence NM_000059.3

*Isoforms imputed by length of product observed in capillary electrophoresis experiments

Supplementary Methods

Two main strategies were followed to analyse the splicing impact of *BRCA2* c7976+5G>T variant: analysis of patient mRNA and development of a minigene assay. All transcripts identified were described according to *BRCA2* GeneBank sequence NM_000059.3 and nomenclature guidelines from Human Genome Variation Society (HGVS) (den Dunnen et al., 2016).

***In silico* analysis**

We obtained bioinformatics predictions from Human Splicing Finder (HSF) and MaxEntScan (MES) algorithms which can be found on Human Splicing Finder v3.0 website (http://www.umd.be/HSF3/). Breast Cancer Genes Prior Probabilities website (http://priors.hci.utah.edu/PRIORS/index.php) was used to obtain a probability of pathogenicity for our variant based on disruption of a canonical splice site. Tools included in Alamut Visual 2.10 Software (http://www.interactive-biosoftware.com/alamut-visual/) were also used (SpliceSiteFinder-like, MES, NNSplice, GeneSplicer and HSF).

Patient whole blood mRNA analysis

HUVH: RNA samples from variant carrier and 10 controls were isolated from 5mL of peripheral blood using Trizol reagent (Invitrogen). DNase treatment was performed using the Ambion®DNA-free™ DNase Treatment and Removal Reagents (Life Technologies) following manufacturer's protocol. A total of 500ng of RNA were reverse transcribed using PrimeScript RT reagent kit (Takara Biotechnology) with both random and oligo-dT primers. PCR primers were designed to amplify at least one whole exon 5' and 3' flanking the variant of interest (see Supp. Table S1 for primers). PCR was performed using EcoTaq DNA Polymerase (Ecogen) under the following cycling conditions: denaturing step of 5min at 95ºC; 35 cycles consisting of 15sec at 95ºC, 15sec at 56ºC and 1min at 72ºC; and a final

elongation step of 10min at 72ºC. RT-PCR products were qualitatively assessed by capillary electrophoresis in a QIAxcel instrument using QIAxcel DNA High Resolution Kit (QIAGEN), and bidirectionally sequenced using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).

HCSC: RNA samples from variant carrier and controls (n=34) were isolated from 1 mL of peripheral blood using a MagnaPure Compact workstation and Magna Pure RNA kits according to manufacturer's instructions (Roche Diagnostics). Approximately 800ng of RNA (as measured in a Nanodrop ND-1000 spectrophotometer) were reverse transcribed using PrimeScript RT reagent kit (Takara Biotechnology) using a mixture of random and oligo-dT primers. PCR primers were designed to amplify at least one whole exon 5' and 3' flanking the variant of interest (see Supp. Table S1). PCR was performed with FastStart Taq (Roche Diagnostics) under the following cycling conditions: denaturing step at 95º for 10 min, followed by 35 cycles consisting of 95º for 30 sec, 58 º for 30 sec, and 72º for 60 sec; and a final extension step at 72º for 7 min. Products were assessed in a conventional ethidium-bromide stained agarose gel. Cycle sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems).

IBGM: Proband's RNA was extracted from 1.5 mL of blood with the QIAamp RNA blood mini kit (QIAGEN). Approximately 200 ng of RNA were retrotranscribed using Transcriptor first strand cDNA synthesis kit (Roche Applied Science) with a mixture of oligo-dT and random hexamers. Subsequently 2-5 µL of cDNA were amplified with GoTaq Hot Start DNA polymerase using primers located at exons 16 and 19 (see Supp. Table S1), under the following conditions: denaturing step at 94º for 2 min, followed by 35 cycles consisting of 94º for 30 sec, 58 º for 30 sec, and 72º for 60 sec; and a final extension step at 72º for 5 min.

RT-PCR products were run in conventional 1.2% agarose gel stained with ethidium-bromide and sequenced with BigDye Terminator v3.1 in an ABI 3100 Genetic Analyzer (Applied Biosystems).

Semi-quantitative Analysis by Capillary Electrophoresis (CE)

To characterize the putative effect of *BRCA2* c.7976+5G>T variant on splicing, we have analyzed the *BRCA2* alternative splicing landscape at the vicinity of exon 17 using two different FAM-labeled primer pairs (see Supp. Table S1). Fluorescent PCRs were performed with FastStart Taq (Roche Diagnostics) or EcoTaq (Ecogen) in a final volume of 25 μL that contained 1-2 μL of cDNA. Samples were denatured at 95º for 5-10 min, followed by 35 cycles consisting of 95º for 30 sec, 58º for 30 sec, and 72º for 60 sec, and final extension step at 72º for 7 min. RT-PCR products were first assessed in conventional ethidium-bromide stained gels or in a QIAxcel (QIAGEN) instrument. After visual inspection, PCRs were diluted (1:1-1:20) and 1 microliter of the diluted PCR was analyzed by CE on an ABI3130xl Genetic Analyzer (Applied Biosystems). GeneScan 500 LIZ or GeneScan 1000 ROX (Applied Biosystems) were used as internal size-standards. Detailed electrophoresis conditions are described in Supp. Table S1. Size-calling and splicing products visualisation was performed using Genemapper software v5.0 (Applied Biosystems).

In all semi-quantitative CE experiments, peak areas from the full-length (FL) transcript and the splicing products detected were used to determine their relative abundance to the overall *BRCA2* expression. To do so, we estimated the splicing fraction (SF) of each transcript, defined as the ratio between the peak area of the individual transcript and the Σ of all peak areas (all transcripts) detected. In a second approach, we normalized FL CE data by dividing each FL peak area with the mean of controls FL peak areas, which were used as the reference group. The same approach was used to determine Δ17,18 relative levels. Only peaks above

100 RFUs (relative fluorescent units) were considered and electropherograms with saturated peaks were discarded. Data graphs were built using GraphPad Prism 7 software.

Splicing analysis by digital PCR (dPCR)

dPCR experiments were performed on a QuantStudio 3D Digital PCR 20K platform according to the manufacturer's instructions (Applied Biosystem, Foster City, CA). To detect *BRCA2* Δ17,18 transcripts, we used a FAM-labelled custom designed TaqMan assay (Applied Biosystems) specific for exons 16-19 junction (5'-GAAGAATTTTATAG-ATTTCTGCTAA-3'). As a proxy for *BRCA2* overall expression, we used a 2'-chloro-7'phenyl-1,4-dichloro-6-carboxy-fluorescein labeled (VIC-labelled) pre-designed TaqMan assay (Applied Biosystems, Hs00609073_m1) specific for the exons 26-27 junction (5'-GAAACAAGCTTCTG-ATGTCTTCTCC-3'). All relative quantification experiments were performed combining both assays in individual chips. dPCR chips were analyzed in the QuantStudio 3D Analysis Suit Cloud Software v2.0 (Applied Biosystem, Foster City, CA), defining FAM as Target. Default settings were used in all cases. After reviewing quality authomatic assesment of the chip quality by the software, only green (data meets all quality thresholds, review of the analysis result not required) and yellow flag chips (data meets all quality thresholds, but manual inspection is recommended) were considered for further analyses. Subsequently, data was exported to an excel file to calculate the FAM/VIC ratio We consider the FAM/VIC ratio as a direct measure of the *BRCA2* Δ17,18 contribution to the overall *BRCA2* expression (*BRCA2* Δ17,18 splicing fraction).

Minigene splicing assay

Minigene Construction and functional validation

A pSAD-derived minigene with *BRCA2* exons 14 to 20 was constructed as previously described and functionally validated (Fraile-Bethencourt et al., 2017). Intronic variant c.7976+5G>T was introduced by site-directed mutagenesis with the QuikChange Lightning kit (Agilent), using the wild-type minigene MGBR2_ex14-20 as template. Mutation c.7976+1G>A from a previous report (Fraile-Bethencourt et al., 2017) was used as positive control to compare splicing profiles of both variants.

Transfection of eukaryotic cells

Approximately $2x10^5$ MCF-7 cells (breast adenocarcinoma cell line, ATCC, LGC Standards) were grown to 90% confluency in 0.5 mL of medium (DMEM, 10% Fetal Bovine Serum, 1% glucose and 1% Penicillin/Streptomycin) in 4-well plates (Nunc). Cells were transiently transfected with 1 µg of wild type and mutant minigenes and 2 µL of Lipofectamine 2000 (Invitrogen). To inhibit nonsense-mediated decay (NMD), cells were incubated with 300 µg/mL cycloheximide (Sigma-Aldrich) for 4 hours. RNA was purified with the Genematrix Universal RNA Purification Kit (EURx) with on-column DNAse I digestion to remove genomic DNA.

RT-PCR of minigenes

Minigene RNA retrotranscription was carried out with 400 ng of RNA and RevertAid H Minus First Strand cDNA Synthesis Kit (ThermoFisher Scientific), using gene specific primer RT-pSPL3-RV (see Supp. Table S1 for primers). Samples were incubated at 42°C for 1 hour, and reactions were inactivated at 70°C for 5 min. Then 1-2 µl of cDNA was amplified with primers SD6-PSPL3_RTFW and RTpSAD-RV (Patent P201231427, CSIC) (size 1806bp) using Platinum Taq DNA polymerase (ThermoFisher Scientific) (see Supp. Table S1). Samples were denatured at 94ºC for 2 min, followed by 35 cycles consisting of 94ºC for

20sec, 58ºC for 20 secs, and 72ºC (1min/kb), and a final extension step at 72ºC for 5 min. After product purification with AccuPrep PCR Purification kit (Bioneer), sequencing reactions were performed either using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) or by the sequencing facility of Macrogen Europe (Amsterdam, The Netherlands).

Fluorescent PCRs were performed in triplicate with primers RTBR2_ex16-FW and FAM-RT-pSPL3-RV (size 1015bp) (see Supp. Table S1) using Platinum Taq DNA polymerase (Invitrogen). Aforementioned cycling conditions were used, but 26 cycles were applied. Samples were run by the Macrogen facility with Genescan Liz-1200 as Size Standard (Applied Biosystems). Runs were analyzed with the Peak Scanner software v1.0 (Applied Biosystems). Only peaks with heights ≥ 50 RFU (Relative Fluorescence Units) were taken into account.

Segregation analysis

Segregation analysis could be performed in IBGM proband's mother and in HUVH proband's sister, both affected with BC. The variant was genotyped in genomic DNA by Sanger sequencing.