Capillary Electrophoresis analysis of Conventional Splicing Assays: IARC Analytical and Clinical Classification of 31 *BRCA2* genetic variants.

Gorka Ruiz de Garibay¹, Alberto Acedo², Zaida García-Casado³, Sara Gutiérrez-Enríquez⁴, Alicia Tosar¹, Atocha Romero¹, Pilar Garre¹, Gemma Llor⁵, Mads Thomassen⁶, Orland Díez⁷, Pedro Pérez-Segura⁸, Eduardo Díaz-Rubio^{1,8}, Eladio A. Velasco², Trinidad Caldés¹, Miguel de la Hoya¹.

¹ Laboratorio de Oncología Molecular. Instituto de Investigación Sanitaria San Carlos (IdISSC), Madrid, Spain. ² Grupo de Splicing y Cáncer, Instituto de Biología y Genética Molecular (CSIC-UVA), Sanz y Forés 3, 47003 Valladolid, Spain. ³ Laboratory of Molecular Biology, Fundación Instituto Valenciano de Oncología. Valencia, Spain ⁴ Oncogenetics Group, Vall d'Hebron Institute of Oncology (VHIO), Universitat Autònoma de Barcelona, Barcelona, Spain ⁵ Unidad de Consejo Genético. Servicio de Oncología Médica. Instituto Oncológico del Vallés, Corporación Sanitaria Parc Taulí, Consorcio Sanitario de Terrassa, Barcelona, Spain. ⁶ Department of Clinical Genetics. Odense University Hospital. Soender Boulevard 29, 5000 Odense C, Denmark. ⁷ Oncogenetics Group, University Hospital Vall d'Hebron, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. ⁸ Servicio de Oncología Médica. Hospital Clínico San Carlos. Madrid. Spain.

Running Title: Analytical and Clinical Classification of *In Vitro* Splicing Assays.

Key Words: *BRCA2*, Splicing, UVs, Capillary Electrophoresis, Minigene, qPCR

Correspondence to Miguel de la Hoya, Laboratorio de Oncología Molecular. Hospital Clínico San Carlos. C/Martín Lagos s/n, Madrid 28040, Spain; mdhoya@hotmail.com

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Rare sequence variants in “high-risk” disease genes often referred as Unclassified Variants (UVs) pose a serious challenge to genetic testing. However, UVs resulting in splicing alterations can be readily assessed by *in vitro* assays. Unfortunately, analytical and clinical interpretation of these assays is often challenging. Here we explore this issue by conducting splicing assays in 31 *BRCA2* genetic variants. All variants were assessed by RT-PCR followed by capillary electrophoresis (EP) and direct sequencing. If assays did not produce clear-cut outputs (Class-2 or Class-5 according to analytical IARC guide-lines), we performed qPCR and/or Minigene assays. The latter were performed with a new splicing vector (pSAD) developed by authors of the present manuscript (Patent P201231427 CSIC). We have identified 3 clinically relevant Class-5 variants (c.682-2A>G, c.7617+1G>A, and c.8954-5A>G), and 27 analytical Class-2 variants (not inducing splicing alterations). In addition, we demonstrate that rs9534262 (c.7806-14T>C) is a *BRCA2* splicing quantitative trait loci (sQTL).

1
2
3 Germ-line inactivating mutations in *BRCA2* (MIM#600185) confer markedly increased risks of
4 breast and ovarian cancer. Arguable, *BRCA2* truncating mutations (PTCs) are considered
5 inactivating (clinically relevant) *per se* (PTCs located 3' to the last 50 nucleotides of the
6 penultimate exon representing one exception). However, a high proportion of the genetic
7 variants identified through clinical testing are not PTCs, but rather missense, synonymous or
8 intronic changes of unknown functional effects [Spurdle et al., 2012]. Several approaches have
9 been proposed to elucidate the clinical relevance (if any) of these so-called unclassified
10 variants (UVs), including association with family history, co-segregation analysis, and co-
11 occurrence with a pathogenic mutation. These and other evidences (provided that they are
12 statistically independent), can be integrated into multi-factorial models that allow likelihood
13 classification of variants [Goldgar et al., 2004; Lindor et al., 2012]. Based on posterior
14 probability outputs, the International Agency for Research on Cancer (IARC) working group on
15 UVs proposed a 5-tier classification scheme with clinical recommendations [Plon et al., 2008].
16 According to this scheme, Class-1 variants are non pathogenic, while Class-5 variants are
17 definitely pathogenic.

18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34 *In vitro* assays assessing the effect of UVs on splicing (or protein activity) can be highly
35 informative in the classification process, especially if multi-factorial likelihood classification is
36 not feasible. In particular, *in vitro* splicing assays are widely used in the clinical setting [Walker
37 et al, 2013], probably reflecting the fact that general approaches to predict and verify splicing
38 alterations are relatively straightforward. Unfortunately, interpretation of results is often more
39 challenging [Spurdle et al., 2008; Walker et al., 2013]. The IARC Working Group on UVs has
40 developed guide-lines for the prediction and assessment of splicing alterations [Spurdle et al.,
41 2008]. Their authors adapted the IARC 5-tier clinical classification scheme (originally intended
42 for clinical classification of multi-factorial model outputs) to the particular case of interpreting *in*
43 *vitro* splicing results [Spurdle et al., 2008]. Due to the specific characteristics of these *in vitro*
44 assays (such as absence of numerical outputs) major modifications were introduced, including
45 the fact that this is a 4-tier classification scheme according to which, Class-2 and Class-5
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

indicate clear-cut splicing assay outputs, the first indicating that the variant allele produces the same transcript(s) as the wild-type allele, the latter indicating that the UV allele expresses *major* PTC transcripts. Class-1 was avoided because splicing analysis outputs *per se* do not allow discarding a pathogenic role (protein function and/or other biological features might be altered). Most relevant, this 4-tier is not a *clinical*, but rather an *analytical* classification scheme. As such, *analytical* Class-2, Class-3, and Class-4 do not have specific clinical recommendations. In this regard, *analytical* Class-5 is exceptional, as *BRCA2* PTCs are considered clinically relevant *per se*. The analytical classification guide-lines have been updated recently [Walker et al., 2013]. See Supp. Table 1 for further details.

On practical grounds, quantitative (or semi-quantitative) estimation of aberrant/normal transcripts ratio (as expressed from the UV allele) is mandatory for clear-cut analytical classification [Spurdle et al., 2008; Walker et al., 2013]. Most important, two confounding effects (normal expression from the wild-type allele and selective degradation of PTC transcripts by the NMD pathway) should be mitigated in order to produce robust outputs. For instance, PTC transcripts can be stabilized by performing cell cultures in the presence of cycloheximide (or other NMD inhibitors). Interferences from the wild-type allele can be avoided by developing reporter Minigene assays. Unfortunately, these time-consuming approaches cannot be easily implemented into the routine of a Clinical Diagnostics Laboratory.

Here we show that incorporating capillary EP and direct sequencing of exonic SNPs into the analysis of otherwise conventional splicing assays (RT-PCRs of fresh-blood isolated RNAs) is able to generate clear-cut analytical outputs (so to speak, *analytical* Class-2 and Class-5 outputs according to the IARC guide-lines). However, we show as well that some variants/carriers demand more specific (and time-consuming) analytical approaches (such as dedicated mini-genes or qPCR).

As a proof of principle, we analyzed the influence of 31 *BRCA2* variants (Table 1) on splicing (see on-line methods and Supp. Table S2 for further details). *In silico* predictions are shown in Table 1 (splice sites) and Supp. Table S3 (enhancer and silencers). However, we performed *in vitro* splicing assays regardless of *in silico* predictions. Furthermore, we analyzed not only *BRCA2* UVs, but also *clinical* Class-1 and *clinical* Class-5 variants. We did it because we are interested in finding not only novel *BRCA2* splicing mutations relevant in the clinic, but also non-pathogenic splicing alterations (if any) that might contribute to elucidate the sometimes complex relationship between splicing alterations and disease. Last but not least, we are interested in evaluating whether conventional splicing assays provide clear-cut analytical outputs according to existing guide-lines.

As far as we know, *BRCA2* c.682-2A>G is a novel variant not previously reported. Capillary EP analyses in carriers (33 cycles RT-PCR assays) detected (in addition to the full-length) a 264nt peak representing roughly 10% of the overall signal (Figure 1A). Interestingly, 45 cycles RT-PCR assays revealed five additional peaks (Figure 1B). Size-callings were compatible with exon 9 skipping (264nt peak), and activation of five cryptic acceptor sites (CASs) located in intron 8 and exon 9 (Figures 1B and 1C). All CASs were supported by *in silico* predictions (Figure 1C). Direct sequencing demonstrated exon 9 skipping, and CAS4 activation (Figure 1D). Other CAS activation events were not sequenced due to the low abundance of the corresponding RT-PCR products. These should be considered putative CAS activation events inferred from capillary EP size-calling. Similar peak patterns were observed in four additional variant carriers (Figure 1E), but not in control samples (data not shown). Importantly, all aberrant splicing events produce PTC transcripts. Segregation analysis together with direct RT-PCR sequencing at informative *loci* (rs9534262 and rs1799955) demonstrated that full-length *BRCA2* expression was wild-type allele specific (Supp. Figure S1). Taking together, the data qualifies for inclusion into *analytical* Class-5 (see Supp. Table 1).

BRCA2 c.7617+1G>A has been reported previously in Spanish [Gutierrez-Enriquez et al., 2009] and Danish [Thomassen et al., 2011] hereditary breast and ovarian cancer families.

Both studies detected exon 15 skipping (a PTC event) in variant carriers, but bi-allelic wild-type expression was not investigated (*analytical* Class-4 outputs). Yet, co-segregation studies performed in Danish families supported a pathogenic role for this variant (*clinical* Class-5 variant). We have performed splicing assays in 6 *BRCA2* c.7617+1G>A carriers. As shown in Supp. Figure S2, capillary EP analysis (33 cycles RT-PCR) detected, in addition to the full-length, a 182nt shorter peak compatible with exon15 skipping. Direct sequencing confirmed the skipping event (Supp. Figure S2). Capillary EP data (semi-quantitative) shows roughly identical expression levels of full-length and $\Delta 15$ transcripts. The data is compatible with the UV allele causing complete (or near complete) exon skipping, but does not prove it. This is so because capillary EP estimates (at best) the ratio between steady-state levels (accumulation) of wild-type and variant transcripts in carriers. This in turn might reflect a combination of multiple underlying processes, including reduced (or unusually higher) transcription rate from the UV allele, instable (or unusually stable) UV transcripts, and wild-type transcripts arising from the UV allele. Yet, segregation analysis together with direct RT-PCR sequencing at one informative *locus* (rs1799955) revealed that full-length *BRCA2* expression was indeed wild-type allele specific (Supp. Figure S2). Overall, the data qualifies for inclusion into *analytical* Class-5 (see Supp. Table 1). We have identified eight Spanish families carrying *BRCA2* c.7617+1G>A (Supp. Figure S3). Genotyping of STRs and SNPs markers detected a common haplotype in carriers from all eight families, suggesting a founder effect (Supp. Table S4).

BRCA2 c.8954-5A>G has been identified recently in Spanish Hereditary Breast and Ovarian Cancer families [de Juan Jimenez et al., 2013], but splicing defects (if any) were not investigated. The change is reported in the Universal Mutation Database (UMD) as a *clinical* Class-3 variant (<http://www.umd.be/BRCA2/>). We have performed splicing assays in two *BRCA2* c.8954-5A>G carriers. Agarose gel electrophoresis showed a single RT-PCR product both in controls and carriers (data not shown), but capillary EP detected an additional peak (+4nt) in carriers (Supp. Figure S4). Direct sequencing revealed a PTC transcript in which the last 4nt of intron 22 have been incorporated into *BRCA2* exon 23 (Supp. Figure S4). Measured by capillary EP, the aberrant mRNA represents roughly 30% of the overall signal.

The data is compatible with wild-type expression arising from the UV allele, but does not prove it (as discussed above). Since bi-allelic wild-type expression was not tested due to lack of informative *loci*, it is not possible to classify the output into any specific analytical category (Class-4 or Class-5?) according to the 2008 guide-lines. However, the very same output qualifies for *analytical* Class-4 according to the 2013 guide-lines (see Supp. Table 1). Interestingly, HSF analysis predicted a *de novo* acceptor site at position c.8954-4 without disrupting exon 23 native acceptor site (Table 1), thus suggesting a mechanism through which c.8954-5A>G might indeed contribute to normal splicing. To investigate wild-type expression arising from the UV allele, we used a reporter Minigene assay (see on-line methods). The analysis revealed that c.8954-5G does not produce full-length mRNA (Supp. Figure S5). As a whole, our data qualifies for inclusion into *analytical* Class-5. We have identified four Spanish families carrying this mutation (Supp Figure S6). Haplotype analysis identified a 1.28Mb shared haplotype in all carriers, suggesting a founder effect (Supp. Table S4).

Similar assays performed in 23 additional *BRCA2* variants did not detect aberrant transcripts, despite the fact that we used high sensitivity capillary EP analysis. *BRCA2* bi-allelic expression was confirmed in all cases (Supp. Figures S7 to S18). Taking together, all 23 studies qualify for inclusion into *analytical* Class-2 (see Supp. Table 1). Yet, it is important to remark that *analytical* Class-2 *per se* do not have a definite role in the clinical classification of *BRCA2* variants.

Not surprisingly, splicing analysis tends to be overlooked in *BRCA2* PTC mutation carriers. However, these (as other mutations) might induce splicing alterations (perhaps even regenerating a full-length open reading frame). This phenomenon is probably unlikely, but is certainly real. For instance, it has been reported that c.145G>T (p.Glu49Ter), a PTC mutation located in *BRCA2* exon 3, enhances exon 3 skipping (an in-frame splicing isoform), illustrating a complex interplay between RNA and protein effects of potential clinical interest [Sanz et al., 2010]. We have included in our analysis five *BRCA2* variants predicted to express PTC transcripts, including four pathogenic (*clinical* Class-5) mutations (c.1773_1776del, c.3264dupT, c.8010_8032del22, and c.9026_9030del) and one benign (*clinical* Class-1)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

change (c.9976A>T). However, capillary EP analysis of RT-PCR assays (agarose in the case of c.3264dupT) did not detect splicing alterations. With the single exception of c.9026_9030del, PTC transcripts were detected at similar levels as wild-type transcripts, suggesting (albeit not proving) that they are not targeted by the NMD pathway (Supp. Figures S19 and S20). Since wild-type expression arising from PTC alleles can be discarded *per se*, all five splicing results qualify for *analytical* Class-5 according to existing guide-lines (Supp. Table 1), regardless of the fact that none of these variants affects splicing.

We have identified five common (MAF>0.1) *BRCA2* variants in our cohort (Table 1). *In silico* analyses did not predict splicing alterations, with the single exception that (according to MaxEnt), c.7806-14T>C (rs9534262) produces a slight reduction of the exon17 acceptor site score (Table 1). Interestingly, two “naturally occurring” alternative splicing events ($\Delta 17,18$ and $\Delta 18$) have been described in this region [Walker et al., 2010]. Capillary EP did not detect novel transcripts in c.7806-14T>C carriers, but suggested that *BRCA2* $\Delta 17,18$ was slightly up-regulated (data not shown). Hence, we decided to explore the role of c.7806-14T>C in alternative splicing regulation by qPCR (see on-line Methods). Quantification experiments were performed in 7 T/T, 20 T/C, and 11 C/C carriers. The expression level of *BRCA2* $\Delta 17,18$ (relative to overall *BRCA2* expression level) was low ($\Delta C_T = 5.19 \pm 0.59$), reflecting the fact that *BRCA2* $\Delta 17,18$ is a minor alternative splicing isoform. Yet, the ANOVA test identified differences associated with each genotype ($p=0.006$). Post-hoc test demonstrated higher levels of *BRCA2* $\Delta 17,18$ in C/C carriers compared with T/T carriers ($p=0.004$) (Supp. Figure S21). To confirm that the C allele is indeed functionally related, and not just in linkage disequilibrium with a true splicing regulatory element, we performed further analysis in a reporter minigene (exons 16 to 20 of *BRCA2*) based on the new splicing vector pSAD® (Consejo Superior de Investigaciones Científicas). In this model, the c.7806-14C reporter minigene showed relative increased levels of both $\Delta 17,18$ and $\Delta 18$ isoforms if compared with c.7806-14T reporter minigene (Supp. Figure S22). Overall, the data indicates that rs9534262 is a bonafide *BRCA2* splicing quantitative trait locus (sQTL). Since *BRCA2* $\Delta 17,18$ and *BRCA2* $\Delta 18$ are PTC splicing isoforms, we hypothesized that rs9534262 might be associated

with breast cancer risk. To test it, we performed case-control association studies and Kaplan-Meier survival analyses. Overall, these analyses do not support an association with risk (Supp. Figure S23).

In summary, we have performed routine *in vitro* splicing assays (blood RNA based RT-PCR experiments) in 31 *BRCA2* genetic variants, producing 29 clear-cut analytical outputs. Of these, two Class-5 (c.682-2A>G and c.7617+1G>A) are clinically relevant findings *per se*. By contrast, neither c.8954-5A>G nor c.7806-14T>C (rs9534262) routine *in vitro* splicing assays resulted in clear-cut analytical classification. The former is of particular clinical relevance, since subsequent dedicated assays (a reporter Minigene) demonstrated that *BRCA2* c.8954-5A>G is indeed a clinically relevant Class-5 variant.

Analysis of c.7806-14T>C (rs9534262) with dedicated assays (reported Minigene and qPCR) produced results that qualify for *analytical* Class 4 (2008 guide-lines), or *analytical* Class-3 (2013 guide-lines) (see Supp. Table 1). However, *BRCA2* rs9534262 is a *clinical* Class-1 variant (MAF is close to 0.5 in many populations). Similarly, splicing analysis in c.9976A>T (p.Lys3326Ter) carriers produced an *analytical* Class-5 output, despite the fact that this is a *clinical* Class-1 variant based on posterior probability analysis (<http://brca.iarc.fr>). These two examples illustrates the fact that the 4-tier classification scheme of splicing assays outputs is more related to the “analytical quality” and “interpretation” of UV splicing assays than to the actual probability of being pathogenic.

Undoubtedly, sequencing informative exonic SNPs (testing bi-allelic expression) is critical in order to produce clear-cut routine *in vitro* splicing assays. In addition, we show that incorporating capillary EP into the analysis provides highly informative analytical data (both high-resolution size-calling and high-sensitivity detecting RT-PCR products). These two features were critical in order to characterize the rather complex splicing outcome observed in *BRCA2* c.682-2A>G carriers. This variant induces the expression of several (probably six) different aberrant transcripts. As a result, the “aberrant signal” is split into different components

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(different RT-PCR products), none of which is detectable by conventional EtBr-stained agarose gel analysis (data not shown). Equally relevant, capillary EP ensures that Class-2 outputs are indeed clear-cut negative results. Capillary EP provides not only high-resolution size-calling and high-sensitivity, but also semi-quantitative information (area under the peaks) that might be useful for analytical purposes. For instance, it allowed us to identify c.7806-14T>C (rs9534262) as a candidate sQTL for further dedicated assays.

We have performed several routine *in vitro* splicing assays in carriers of UV variants with low *prior* probability (according to *in silico* predictions) of finding an alteration; a conservative approach that we nonetheless recommend in the clinical setting. Undoubtedly, *in silico* analyses might provide high accuracy predictions, but they do not detect all splicing defects, especially in variants located outside the 5' and 3' consensus sites (Houdayer et al., 2012). Yet, we think that *in silico* analyses have important roles in the clinic. For instance, they assist with prioritizing variants for dedicated splicing assays (such as reporter Minigenes) in cases where routine assays do not produce clear-cut outputs, or in cases in which routine assays are not feasible (i.e. lack of fresh blood RNA).

In conclusion, we show that incorporating capillary EP and direct sequencing of exonic informative SNPs to, otherwise, conventional fresh blood-based RT-PCR splicing assays often provides high-quality results that allow systematic analytical classification according to existing guide-lines.

Acknowledgments

EAV was supported by grants PI10/2910 of the Instituto de Salud Carlos III (Spanish Ministry of Science and Innovation), and CSI004A10-2 (Consejería de Educación) and BIO39/VA27/10 (Consejería de Sanidad) of the Regional Government of Castilla y León. Alberto Acedo is supported by the European Social Fund and Consejería de Educación de la Junta de Castilla y León under the P.O. Castilla y León 2007-2013. ZGC is a member of the Genetic Counseling Program in the Comunidad Valenciana. SG-E is funded by a Miguel Servet contract (CP10/00617) from the Spanish Carlos III Health Institute. The research in HUVH-VHIO (Barcelona) was supported by a Miguel Servet Project grant CP10/00617. MH was supported by Fondo de Investigación Sanitaria (FIS) research grants PI 09/00859 and PI 12/00539, an initiative of the Insituto de Salud Carlos III (Spain), partially supported by European Regional Development FEDER Funds.

Bibliography

Acedo A, Sanz DJ, Duran M, Infante M, Perez-Cabornero L, Miner C, Velasco EA. 2012. Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res* 14:R87.

Biswas K, Das R, Eggington JM, Qiao H, North SL, Stauffer S, Burkett SS, Martin BK, Southon E, Sizemore SC, Pruss D, Bowles KR et al. 2012. Functional evaluation of BRCA2 variants mapping to the PALB2-binding and C-terminal DNA-binding domains using a mouse ES cell-based assay. *Hum Mol Genet* 21:3993-4006.

Bonnet C, Krieger S, Vezain M, Rousselin A, Tournier I, Martins A, Berthet P, Chevrier A, Dugast C, Layet V, Rossi A, Lidereau R 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.

de Garibay GR, Diaz A, Gavina B, Romero A, Garre P, Vega A, Blanco A, Tosar A, Diez O, Perez-Segura P, Díaz-Rubio E, Caldés T et al. 2013. Low prevalence of SLX4 loss-of-function mutations in non-BRCA1/2 breast and/or ovarian cancer families. *Eur J Hum Genet*. 21:883-886.

de Juan Jiménez I, García Casado Z, Palanca Suela S, Esteban Cardeñosa E, López Guerrero JA, Segura Huerta A, Chirivella González I, Sánchez Heras AB, Juan Fita MJ, Tena García I, Guillen Ponce C et al. 2013. . Novel and recurrent BRCA1/BRCA2 mutations in early onset and familial breast and ovarian cancer detected in the Program of Genetic Counseling in Cancer of Valencian Community (eastern Spain). Relationship of family phenotypes with mutation prevalence. *Fam Cancer*. PubMed PMID: 23479189

Dosil V, Tosar A, Canadas C, Perez-Segura P, Diaz-Rubio E, Caldes T, de la Hoya M. 2010. Alternative splicing and molecular characterization of splice site variants: BRCA1 c.591C>T as a case study. *Clin Chem* 56:53-61.

Guidugli L, Pankratz VS, Singh N, Thompson J, Erding CA, Engel C, Schmutzler R, Domchek S, Nathanson K, Radice P, Singer C, Tonin PN et al. 2013. A Classification Model for BRCA2 DNA Binding Domain Missense Variants Based on Homology-Directed Repair Activity. *Cancer Res* 73:265-275.

Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ; Breast Cancer Information Core (BIC) Steering Committee. 2004. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet*. 75:535-544.

Gutierrez-Enriquez S, Coderch V, Masas M, Balmana J, Diez 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.

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.

- Lindor NM, Guidugli L, Wang X, Vallee MP, Monteiro AN, Tavtigian S, Goldgar DE, Couch FJ. 2012. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Hum Mutat* 33:8-21
- Menendez M, Castellsague J, Mirete M, Pros E, Feliubadalo L, Osorio A, Calaf M, Tornero E, del Valle J, Fernandez-Rodriguez J, Quiles F, Salinas M et al. 2012. Assessing the RNA effect of 26 DNA variants in the BRCA1 and BRCA2 genes. *Breast Cancer Res Treat* 132:979-992.
- Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FB, Hoogerbrugge N, Spurdle AB, Tavtigian SV; IARC Unclassified Genetic Variants Working Group. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat*. 29:1282-1291.
- Sanz DJ, Acedo A, Infante M, Duran M, Perez-Cabornero L, Esteban-Cardenosa E, Lastra E, Pagani F, Miner C, Velasco EA. 2010. A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin Cancer Res* 16:1957-67
- Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM; IARC Unclassified Genetic Variants Working Group. 2008. Prediction and assessment of splicing alterations: implications for clinical testing. *Hum Mutat*. 29:1304-1313.
- Spurdle AB, Healey S, Devereau A, Hogervorst FB, Monteiro AN, Nathanson KL, Radice P, Stoppa-Lyonnet D, Tavtigian S, Wappenschmidt B, Couch FJ, Goldgar DE 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.
- Thomassen M, Pedersen IS, Vogel I, Hansen TV, Brasch-Andersen C, Brasen CL, Cruger D, Sunde L, Nielsen FC, Jensen UB, Bisgaard ML, Borg A et al. 2011. A BRCA2 mutation incorrectly mapped in the original BRCA2 reference sequence, is a common West Danish founder mutation disrupting mRNA splicing. *Breast Cancer Res Treat* 128:179-185.
- Walker LC, Whaley PJ, Couch FJ, Farrugia DJ, Healey S, Eccles DM, Lin F, Butler SA, Goff SA, Thompson BA, Lakhani SR, Da Silva LM 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:E1484-1505. PubMed PMID:23893897.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure Legends

Figure 1

A) Capillary EP analysis (33 cycles RT-PCR assays) identifies one aberrant transcript in carriers. B) Capillary EP (45 cycles RT-PCR assays) identifies five additional aberrant transcripts. Size callings are compatible with exon 9 skipping (225 nt peak), and activation of five cryptic acceptor sites (CASs). C) A *BRCA2* sequence including exon9 (capital letters) and intronic boundaries is shown. Five CASs are displayed (bold letters). Human Splicing Finder (HSF) scores (% of wild-type acceptor site score) are indicated in brackets. D) Direct sequencing of RT-PCRs demonstrates exon9 skipping and CAS4 activation. E) 45 cycle PCR assays allowed semi-quantitative analysis of aberrant transcripts. The chart represents the relative amount of each aberrant transcript compared with the overall aberrant transcript signal, as determined in five carriers (mean and SD are shown)

Table 1. Splicing analysis in 35 *BRCA2* variants

Location	HGVS description ¹	dbSNP	MAF ²	In Silico Splicing Analysis ³		Analytical classification of In Vitro Splicing Assays Outputs ⁴			
				HSF	MaxEnt	Published Data	Routine Assay	Minigene	qPCR
Intron 1	c.-39-26G>A	-	.0004	no effect	no effect	-	Class-2	-	-
Intron 4	c.426-89T>C	rs3783265	.022	(c.426-81; 86.04%)	no effect	-	Class-2	-	-
Intron 6	c.516+14C>T	rs182828913	.0016	no effect	no effect	Class-2 ⁵	Class-2	-	-
Intron 7	c.631+25C>T	-	.0008	no effect	no effect	-	Class-2	-	-
Exon 8	c.680C>T (p.Ala227Val)	rs149565664	.0004	-0.57%	-3.04%	-	Class-2	-	-
Intron 8	c.681+69C>T ‡	-	.0004	no effect	no effect	-	Class-2	-	-
	c.682-2A>G ‡	-	.0008	-33.39%	-112.46%	-	Class-5	-	-
Exon 10	c.1114A>C (p.His372Asn)	rs9534262	.11	no effect	no effect	Class-2 ⁶	-	-	-
	c.1773_1776del (p.Ile591fs)	rs80359305	.0004	no effect	no effect	-	Class-5	-	-
Exon 11	c.3264dupT (p.Gln1089fs)	rs80359380	.0097	(c.3268; 103.25%)	(c.3283; 7.52%)	-	Class-5	-	-
	c.3396A>G (p.Lys1132=)	rs1801406	.25	no effect	no effect	-	-	-	-
	c.3807T>C (p.Val1269=)	rs543304	.14	no effect	no effect	-	-	-	-
	c.6825G>A (p.Glu2275=) ‡	-	.0004	no effect	no effect	-	Class-2	-	-
Exon 12	c.6928A>C (p.Thr2310Pro) ‡	-	.0004	(c.6935; 96.34%)	no effect	-	Class-2	-	-
Exon 14	c.7242A>G (p.Ser2414=)	rs1799955	.20	no effect	no effect	-	-	-	-
Intron 14	c.7435+8A>G ‡	-	.0004	(c.7436+7; 92.39%)	no effect	-	Class-2	-	-
Exon 15	c.7469T>C (p.Ile2490Thr)	rs11571707	.0024	no effect	no effect	-	Class-2	-	-
Intron 15	c.7617+1G>A ⁷	-	.0016	-29.43%	-83.47%	non classifiable (Class-4) ^{8,*}	Class-5	-	-
Exon 16	c.7626G>A (p.Thr2542=)	rs61754138	.0004	no effect -	no effect	-	Class-2	-	-
	c.7796A>G (p.Glu2599Gly) ‡	-	.0004	(c.7796; 88.03%)	no effect	-	Class-2	-	-
Intron 16	c.7806-14T>C	rs9534262	.40	no effect	-4.68%	-	non classifiable**	Class-4(Class-3)	Class-4(Class-3)
Intron 17	c.7976+5G>T ‡	-	.0004	-16.83%	-74.19%	-	Class-2	-	-
	c.7976+57G>C	-	.0004	no effect	no effect	-	Class-2	-	-
Exon 18	c.8010_8032del (p.Ala2671fs)	rs80359690	.0004	no effect	no effect	-	Class-5	-	-
	c.8149G>T (p.Ala2717Ser)	rs28897747	.0004	(c.8168; 55.22%)	no effect	-	Class-2	-	-
Exon 20	c.8503T>C (p.Ser2835Pro)	rs11571746	.0024	no effect	no effect	-	Class-2	-	-
	c.8524C>T (p.Ar2842Cys) ⁹	rs80359104	.0004	(c.8535; 88.19%)	no effect	-	Class-2	-	-
	c.8592C>T (p.Ala2864=)	-	.0004	(c.8591; 79.82%)	no effect	-	Class-2	-	-
Exon 22	c.8850G>T (p.Lys2950Asn)	rs28897754	.0008	no effect	no effect	-	Class-2	-	-
	c.8851G>A (p.Ala2951Thr)	rs11571769	.0004	no effect	no effect	-	Class-2	-	-
Intron 22	c.8954-5A>G	-	.0008	-0.08 % (c.8954-4; 103.97%)	-43.77 % (c.8954-4; 98.45%)	non classifiable (Class-4) ^{10,*}	non classifiable (Class-4)*	Class-5	-
Exon 23	c.9004G>A (p.Glu3002Lys) ^{9,11}	rs80359152	.0004	(c.9008; 87.90%)	(c.9008; 38.16%)	-	Class-2	-	-
	c.9026_9030del (p.Thr3009fs)	rs80359741	.0020	no effect	no effect	-	Class-5	-	-
	c.9038C>T (p.Thr3013Ile)	rs28897755	.0008	no effect	no effect	Class-2 ⁶	Class-2	-	-
Exon 27	c.9976A>T (p.Lys3326Ter)	rs11571833	.009	no effect	no effect	-	Class-5	-	-

¹ Description according to HGVS guide-lines (<http://www.hgvs.org/mutnomen>). Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the GeneBank reference sequence NM_000059.3.

² MAF in index cases from HBOC families (N=2558 chromosomes). ³ De novo splice site predictions are shown in brackets. ⁴ Both 2008 [Spurdle et al, 2008] and 2013 [Walker et al, 2013] guide-lines have been considered. If discrepant, 2013 Class is shown in brackets. Published data was classified by authors of the present manuscript. ⁵ [Bonnet et al., 2008]. ⁶ [Sanz et al., 2010]. ⁷ Segregation analyses supported that this is indeed a Class-5 variant [Thomassen et al., 2011]. ⁸ [Gutierrez-Enriquez et al., 2009; Thomassen et al., 2011]. ⁹ Inactivating in protein functional assay [Guidugli et al., 2013]. ¹⁰ [Menendez et al., 2012]. ¹¹ Inactivating in protein functional assays [Biswas et al., 2012]. * UV allele contribution to wild-type expression was not tested.

** Poor quality semi-quantitative data does not allow classification. ‡ Novel variants submitted to the Breast Informative Core (BIC) database (<http://research.nhgri.nih.gov/bic/>).

A) Capillary EP analysis (33 cycles RT-PCR assays) identifies one aberrant transcript in carriers. B) Capillary EP (45 cycles RT-PCR assays) identifies five additional aberrant transcripts. Size callings are compatible with exon 9 skipping (225 nt peak), and activation of five cryptic acceptor sites (CASs). C) A BRCA2 sequence including exon9 (capital letters) and intronic boundaries is shown. Five CASs are displayed (bold letters). Human Splicing Finder (HSF) scores (% of wild-type acceptor site score) are indicated in brackets. D) Direct sequencing of RT-PCRs demonstrates exon9 skipping and CAS4 activation. E) 45 cycle PCR assays allowed semi-quantitative analysis of aberrant transcripts. The chart represents the relative amount of each aberrant transcript compared with the overall aberrant transcript signal, as determined in five carriers (mean and SD are shown)

254x190mm (96 x 96 DPI)

Patients and Samples.

All variants under study were first identified through clinical testing in index cases of high risk hereditary breast and ovarian cancer families undergoing genetic counseling according to Spanish National Health System recommendations (N=1279). According to the frequency observed in the index case cohort, we refer throughout the text to rare ($f < 0.1$) and common ($f > 0.1$) *BRCA2* variants. We have performed *in vitro* RNA analysis in index cases (or relatives) carrying the genetic variants of interest. In addition, rs9534262 (c.7806-14T>C) has been genotyped in 376 healthy control individuals, a consecutive series of 173 women diagnosed of breast cancer, and 438 index cases from BRCAX families (high risk families negatives for *BRCA1* and *BRCA2* germ-line mutations after comprehensive genetic testing). The BRCAX cohort has been extensively described elsewhere [de Garibay et al., 2013; Dosil et al., 2010]. All participants signed an informed consent. The study was approved by the Ethics Committee of the participating institutions.

BRCA2 sequence variants description.

All *BRCA2* genetic variants are described according to HGVS guide-lines (<http://www.hgvs.org/mutnomen>). Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the GeneBank reference sequence NM_000059.3

In silico splice predictions

Consensus splice sites predictions were performed with HSF and MaxEnt algorithms. Splicing enhancer predictions were performed with ESE finder, RESCUE-ESE, PESE and EIE algorithms. Similarly, splicing silencer predictions were analyzed with ESS, Fas-ESS, PESS and IIE algorithms. All analyses were performed on-line with Human Splicing Finder version 2.4.1 (default settings).

Blood-based RT-PCR *In Vitro* Splicing Assays (Routine Assay).

RNAs were extracted from fresh blood using a MagnaPure Compact workstation and Magna Pure RNA kits according to the manufacturer’s instructions (Roche Diagnostics). RT was performed with 300ng of RNA and the PrimeScript RT kit (Takara Biotechnology) following manufacturer’s protocol. Depending on the interrogated variant, cDNAs were amplified with different combinations of forward and reverse primers (Supp. Table 1). Amplicons were designed taking into consideration both *in silico* predictions and “naturally occurring” alternative splicing events. The later might be relevant for the correct analysis of *BRCA2* splicing variants, as has been previously shown for *BRCA1* splicing variants [Dosil et al., 2010]. Fluorescent PCRs were performed with FastStart Taq (Roche Diagnostics) in a final volume of 25 µL that contained 1 µL of cDNA. Samples were denatured at 95° for 10 min, followed by 45 cycles consisting of 95° for 30 sec, 58 ° for 30 sec, and 72° for 30 sec, and final extension step at 72° for 7 min. For semi-quantitative analysis, the number of PCR cycles was reduced to 33. RT-PCR products were analyzed by capillary electrophoresis on an ABI3130 automated sequencer (Life Technology). Depending on the expected sizes, LIZ-500 or LIZ-1200 (Life Technology) was used as internal size-standard. Size-calling and peak areas (semi-quantitative assays) were analyzed with the Genemapper software v.3.7 (Life Technology). Bi-allelic expression was analyzed by direct sequencing of RT-PCR reactions performed at informative (heterozygous) loci: either the interrogated variant itself (exonic variants) or linked exonic SNPs (in the case of intronic variants). Cycle sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing kit (Life Technology).

Minigene assays

Construction of Minigenes:

The construction of the minigene containing *BRCA2* exons 23 and 24 and flanking intronic sequences (MGBR2_Ex23-24) inserted into the splicing vector pSPL3b was previously described [Acedo et al, 2012]. The minigene containing exons 16 to 20 of the *BRCA2*

(MGBR2_Ex16-20) gene was constructed in three sequential cloning steps in a new splicing vector derived from pSPL3 called pSAD (Patent **P201231427** CSIC, *Priority Patent Application filed*; Acedo et al, manuscript in preparation): exons 17+18 in a single fragment (1,556 bp; intron 17= 485 bp), exon 16 (805 bp) and exons 19+20 in a single fragment (1165 bp; intron 19= 398 bp). These sequences were amplified with the Phusion Hot Start II High Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA) following the manufacturer's instructions, and using the following primers:

for exons 17+18:

mgbr2_ex17fw, 5'GCTCTAGAACTAGTGGATCCCCCGGTCAGTATGATACTTTGATACATGT3'

mgbr2_ex18rev, 5'ATAAGCTTGATATCGAATTCCTGCAGTGGAATAGGGATCTGATCAA3',

for exon 16:

mgbr2_ex16fw, 5'GCTCTAGAACTAGTGGATCCCCCGGCCCTTTTGTTCCTCCATCTAAGT3'

mgbr2_ex16rev, 5'GACATGTATCAAAGTATCATACTGAATAAATGCCTAAGAAAAATGT3'

For exons 19+20:

mgbr2_ex19fw, 5'CACACACTCGAGATAGCATTAAAGAACTTGTAGCA3'

mgbr2_ex20rev, 5'CACACAGGTACCATTACAAATGGCTTAGACCTGA3'

For cloning purposes a 5' tail (underlined) was added to each primer. The final structure of the insert of the minigene from the 5' to the 3' end was: ivs15 (333 bp) - exon 16 (188 bp)- ivs16 (234 bp)- // - ivs16 (181 bp) - exon 17 (171 bp) – ivs17 (485 bp) – exon 18 (355 bp) – ivs18 (314 bp)- // - ivs18 (235 bp) – exon 19 (156 bp) – ivs19 (398 bp) – exon 20 (145 bp)– ivs20 (207 bp). Introns 16 (4,595 bp) and 18 (6,853 bp) were too large so they were shortened to facilitate cloning. All the new constructs were sequenced with the Big Dye 3.1 Sequencing Kit (Life Technologies) to confirm the fidelity of cloned sequences.

Site directed mutagenesis:

Mutagenesis was carried out according to the PCR mutagenesis protocol (<http://www.methodbook.net/pcr/pcrmut.html>) with Pfu Turbo DNA polymerase (Agilent,

Santa Clara, CA, USA). Wild type (wt) minigenes MGBR2_Ex23-24 and MGBR2_Ex16-20 were used as templates to generate DNA variants.

For c.7806-14T>C of intron 16 primers used were

5' TTTATGATAATATTCTACCTTTATTTGTTTCAGGGCTC 3'

5' GAGCCCTGAACAAATAAAGGTAGAATATTATCATAAA 3'

For c.8954-5A>G of intron 22 primers used were

5' TCTTTCTCATCTTTCTCCAGACAGTTATACTGAGTATTTG 3'

5' CAAATACTCAGTATAACTGTCTGGAGAAAGATGAGAAAGA 3'

Transfection of HeLa cells:

Approximately 10⁵ HeLa cells (human cervical carcinoma) were grown to 90% to 95% confluency in 0.5 mL of growth medium (DMEM, 10% fetal bovine serum, 1% glucose and 1% penicillin/streptomycin) in four-well plates (Nunc, Roskilde, Denmark). Cells were transfected with 1 µg of plasmid minigene and 2 µL of Lipofectamine 2000 (Life Technologies) following the manufacturer's protocol. RNA was purified 24-48 hours after transfection with Nucleospin-RNA-II (Macherey-Nagel, Düren, Germany) that includes on-column rDNase treatment and quantified in a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific)

RT-PCR of minigenes:

Retrotranscription was carried out with 200 ng of RNA and the transcriptor first strand cDNA synthesis kit (Roche Diagnostics). Fluorescent RT-PCRs were performed in triplicate in a final volume of 20 µL that contained 2 µL of cDNA and the flanking primers of constitutive exons of pSPL3, SD6-PSPL3_RTFW (FAM-5' TCACCTGGACAACCTCAAAG 3') and SA2-PSPL3_RTREV (5' TGAGGAGTGAATTGGTCGAA 3'). Samples were denatured at 94°C for two minutes, followed by 25-30 cycles consisting of 94°C for 20 seconds, 58°C for 20 seconds, and 72°C for 30-90 seconds, and a final extension step at 72°C for two minutes. The PCR products were analyzed by capillary electrophoresis. Expected sizes of the RT-

PCR products were 488 bp and 1199 bp for minigenes MGBR2_Ex23-24 and MGBR2_Ex16-20, respectively

Relative quantification of *BRCA2* $\Delta 17,18$ expression level.

The relative expression level of *BRCA2* $\Delta 17,18$ was measured by real-time PCR with a dedicated FAM-labeled TaqMan gene expression assay. The assay was designed and synthesized using the Assay-by-Design service (Life Technology) that includes proprietary specificity and amplification efficiency QC testing. As endogenous control, we used the VIC-labeled Custom TaqMan gene expression assay Hs01037420_m (Life Technology) that targets *BRCA2* exon 24 (since no *BRCA2* $\Delta 24$ splicing isoform has been reported, *BRCA2* exon 24 probably reflects the overall *BRCA2* expression level). Both assays were performed in a single tube reaction including 1 μ L of cDNA, 1X TaqMan Universal PCR MasterMix (Life Technology) and 1X TaqMan assays. Experiments were performed in a 7500 Fast Real-Time PCR system (Life Technology) with default thermal cycling conditions (Life Technology). For quantification purposes, we measured ΔC_T , defined as $C_T \text{ } BRCA2\Delta 17,18 - C_T \text{ } BRCA2e24$. Therefore, ΔC_T reflects the expression level of *BRCA2* $\Delta 17,18$ relative to the overall expression level from the same locus. ΔC_T was analyzed in triplicate in all tested samples.

Haplotype analysis of recurrent variants

Haplotype analysis was performed with up to 9 short tandem repeats (STR) covering 14.4Mb of chromosome 13 surrounding the *BRCA2* gene. Primer sequences, order of markers and physical distances among D13S260, D13S1699, D13S1698, D13S1697, D13S1701, D13S1715, D13S1493, D13S267, and D13S1694 were obtained from the Ensembl database (<http://www.ensembl.org>). Polymerase chain reaction (PCR) products of STR markers were generated using fluorescently labelled primers. Size-calling was performed by capillary electrophoresis. Additionally, 3 *BRCA2* SNPs (rs1799943, rs144848, and rs9534262) were genotyped by HRMA.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Statistical Analyses.

BoxPlots, Kaplan-Meyer survival analysis, chi-square, and ANOVA test were performed with SPSS v13.0 software.

For Peer Review

Supp. Figure S1. A family carrying the BRCA2 variant c.682-2A>G is displayed. Segregation analysis allowed us to determine the phase between the variant allele and two SNPs located in BRCA2 exon 10 (rs9534262) and exon14 (1799955). Direct sequencing of RT-PCR assays (33-cycle PCR) performed with RNA isolated from individual ID_1 showed that mutant allele contribution to overall BRCA2 expression (if any) is minimal.

Supp. Figure S2. A, B) capillary EP analysis (33 cycles RT-PCR) revealed an aberrant transcript 182nt shorter than wt in c.7617+1G>A carriers, suggesting *BRCA2* exon 15 skipping. Direct sequencing confirmed exon 15 skipping. C) An informative locus in exon 14 (rs1799955) allowed us to demonstrate that the variant allele does not contribute to full-length expression. Note that RT-PCR products generated with an exon15 reverse primer (right) are full-length specific. Segregation analysis allowed us to determine the phase between the variant allele and rs1799955 (data not shown).

Supp. Figure S3. Eight Spanish families carrying the *BRCA2* genetic variant c.7617+1G>A. Numbered Individuals were included in the haplotype analysis (See Supp. Table S2).

A

B

Supp. Figure S4. Capillary EP analysis of the RT-PCR with primers at exon 21 and 24 in a c.8954-5A>G carrier. Direct sequencing demonstrates an insertion corresponding to the last 4nt of intron 22.

Supp. Figure S5. Splicing outcomes of variants of exon 23 of BRCA2 in minigene MGBR2_Ex23-24. RT-PCR products labelled with FAM (blue peaks) of wild type and mutant minigenes were run in an ABI3130 sequencer with Genescan ROX 500 (red peaks) as size standard. The different RNA isoforms are indicated. A) Wild type minigene; B) c.8954-3C>G (intron 22); C) c.9117+1G > A c; D) c.8954-5A>G; E) Sequencing of isoforms produced from c.8954-5A>G.

Supp. Figure S6. Four Spanish families carrying the *BRCA2* genetic variant c.8954-5A>G. Numbered Individuals were included in the haplotype analysis (See Supp. Table S3).

Supp. Figure S7. a) RT-PCR with primers located in exons 1 and 4 shows a similar splicing pattern in controls and c.-39-26G>A carriers, with the presence of full-length and $\Delta 3$ *BRCA2* transcripts. b) Capillary EP analysis of semi-quantitative RT-PCR assays (33 cycle PCR) shows no obvious differences in the amount of $\Delta 3$ *BRCA2* transcripts in c.-39-26G>A carriers. c) Bi-allelic expression was confirmed by direct sequencing of an RT-PCR assay containing an informative locus (rs144848).

Supp. Figure S8. Capillary EP analysis of semiquantitative RT-PCR with primers located at exons 2 and 8 shows similar patterns of splicing in controls, c.426-89T>C, c.516+14C>T and c.631+25C>T carriers. Full-length, Δ3 and Δ6q7 *BRCA2* transcripts were detected in all samples at similar levels. Bi-allelic expression was confirmed by direct sequencing of RT-PCR assays containing informative loci (rs1799944, rs766173, rs1799943).

Supp. Figure S9. Capillary EP analysis of semi-quantitative RT-PCR assays (33 cycle PCR) performed with primers located at exon 4 and 10 shows similar patterns of splicing in wt controls and c.680C>T carriers. Full-length and $\Delta 6q7$ *BRCA2* transcripts were detected in all samples at similar levels. Direct sequencing of RT-PCR assays confirmed bi-allelic expression in c.680C>T carriers.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supp. Figure S10. Capillary EP analysis of RT-PCR assays performed with primers located in exons 7 and 10 shows the presence of a single transcript both in wt controls and one c.681+69C>T carrier. Bi-allelic expression was confirmed by direct sequencing (reverse strand) of RT-PCR assays containing an informative locus in exon 14 (rs1799955).

Supp. Figure S11. RT-PCR with primers located in exons 11 and 12 (E11.29-S and E12-FAM-AS) shows the presence of a single transcript both in wt controls and one c.6825G>A carrier. Direct sequencing confirmed bi-allelic expression in variant carriers.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supp. Figure S12. RT-PCR with primers located in exons 12 and 15 (E12-S and E15-FAM-AS) shows the presence of a single transcript in wt controls and one *c.7435+8A>G* carrier. Bi-allelic expression was confirmed by direct sequencing of RT-PCR assays containing an informative locus at exon (rs1801406).

Supp. Figure S13. a) RT-PCR with primers located at exon 14 and 16 (E14.2-S and E16.1-AS) shows the presence of a single transcript in control and c.7469T>C carrier. **b)** Sequencing of this product confirms the expression of both alleles.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supp. Figure S14 a) RT-PCR with primers located at exon 15 and 17 (E15.2-S and E17.1-AS) shows the presence of a single transcript in controls, c.7796A>G carrier and c.7626G>A carrier. **b)** Sequencing of these products confirms the expression of both alleles in all cases.

Supp. Figure S15. Semiquantitative RT-PCR with primers located at exon 16 and 19 shows similar splicing patterns in controls, c.7976+57G>C, c.7976+5G>T and c.8149G>T carriers. Full-length and $\Delta 17,18$ *BRCA2* transcripts were detected in all samples at similar levels. Bi-allelic expression was confirmed by direct sequencing of RT-PCRs containing the *BRCA2* variant (c.8149G>T), or in the case of intronic variants, an informative exonic locus (rs144848).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supp. Figure S16. RT-PCR with primers located at exon 19 and 22 (E19-S and E22-FAM-AS) shows the presence of a single transcript in c.8503T>C, c.8524C>T, and c.8503T>C carriers. Direct sequencing of RT-PCR products confirmed the expression of both alleles.

Supp. Figure S17. RT-PCR with primers located at exon 19 and 23 (E19-S and E23-FAM-AS) shows the presence of a single transcript in wt control, c.8850G>T and c.8851G>T carriers. Direct sequencing confirmed bi-allelic expression.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supp. Figure S18. RT-PCR with primers located at exon 21 and 24 (E21-S and E24-FAM-AS) shows the presence of a single transcript in control, c.9004G>A and c.9038C>T carriers. Direct sequencing confirmed bi-allelic expression.

Supp. Figure S19 Splicing analysis in carriers of three different *BRCA2* PTC mutations did not identify splicing alterations. However, all predicted PTC transcripts were detected. With the single exception of p.Tyr3009fs, PTC transcripts were detected at similar levels than wt transcripts.

Supp. Figure S20 Splicing analysis in carriers of two different *BRCA2* PTC mutations did not identify splicing alterations. PTC transcripts were detected at similar levels than wt transcripts (sequencing data).

Supp. Figure S21. qPCR analysis of *BRCA2* $\Delta 17,18$ splicing isoform using *BRCA2* exon 24 expression (a proxy for overall *BRCA2* expression level) as an endogenous control. $\Delta C_T = (C_T \text{ } BRCA2 \Delta 17,18 - C_T \text{ } BRCA2 \text{ exon24})$. 7 T/T, 20 T/C, and 11 C/T samples were analyzed in triplicated. Statistically significant differences between C/C and T/T genotypes were observed (*BRCA2* $\Delta 17,18$ is upregulated in C/C carriers). Quantification in T/C carriers is compatible with a dosage effect. Overall, the data indicates that rs9534262 (c.7806-14T>C) modulates *BRCA2* alternative splicing.

Supp. Figure S22. Minigene analysis of *BRCA2* $\Delta 18$ and $\Delta 17,18$ expression level using *BRCA2* full-length as an internal reference. Relative signal=(peak area splicing isoform/peak area full-length). *BRCA2* $\Delta 17,18$ and *BRCA2* $\Delta 18$ are upregulated in rs9534262 C minigene.

Supp. Figure S23. A) Association between rs9534262 genotype and breast cancer risk. Allelic frequencies and two genetic models are tested. Healthy controls, sporadic breast cancer, and BRCAX (index cases from BRCAX families) cohorts are described in on-line method. B) Kaplan Meier analysis of the age at diagnosis (a dominant model is tested in the sporadic breast cancer cohort). The median age of diagnosis was 62 years for c.7806-14C carriers (CC+CT) and 54 years for non-carriers (TT), but differences were not statistically significant. C) Similar results were obtained when median age of diagnosis was compared between CC and TT genotypes (64 vs. 54 years).

Supp. Table S1. Analytical classification of *in vitro* splicing assays outputs.

Class	2008 original IARC guide-lines ¹	2013 up-dated guide-lines ²
5	The variant allele produces major transcript(s) carrying a PTC, or in-frame deletions disrupting known functional domain(s)	The variant allele produces only transcript(s) carrying a PTC, or in-frame deletions disrupting known functional domain(s), determined by semi-quantitative or quantitative methods.
4	The variant allele produces minor transcript (s) carrying a PTC, or in-frame deletions disrupting known functional domain(s), at higher level than similar transcripts corresponding to physiological alternative splicing, or absent in wt controls.	The variant allele produces aberrant transcript(s) carrying a PTC, or in-frame deletions disrupting known functional domain(s), these transcripts are not found in controls assayed in tandem, but further assessment to determine if the UV allele does or does not contribute to production of wildtype transcript.
3	The variant allele produces major or minor transcript(s) carrying an in-frame deletion not disrupting known functional domain(s), at a higher level than similar transcripts corresponding to physiological alternative splicing.	All other transcript patterns e.g. the variant allele produces transcript(s) carrying an in-frame deletion not disrupting known functional domain(s), variant allele upregulates transcripts corresponding to physiological alternative splicing, at least some of which do not carry a PTC or are not an in-frame deletion disrupting known functional domains.
2	The variant allele produces the same transcript(s) as the wildtype allele.	The variant allele produces transcript patterns comparable to that of wildtype controls assessed in tandem.

¹[Spurdle et al., 2008], ²[Walker et al., 2013]

In 2008, the IARC proposed a classification system aimed to achieve robustness in the analytical interpretation of *in vitro* splicing assays. The scheme has been up-dated recently. Somehow inevitably; these guide-lines introduce vague terms such as major, minor, or comparable (highlighted in bold italics by authors of the present report). Note that the 2008 original guide-lines describe splicing assay outputs, while the update is more concerned with methodological issues. By doing so, the *meaning* of the classification system has been slightly modified. For instance, *analytical* Class-4 does not reflect any more UV alleles expressing both wild-type and PTC transcripts, but a mere lack of analytical information (the UV contribution to wild-type expression has not been tested). Note that analytical guide-lines do not make a distinction between PTCs located 5' or 3' to the last 50 nucleotides of the penultimate exon, despite the fact that this might have clinical relevance.

Supp. Table S2. Primers used for Conventional RT-PCR Splicing Assays

Name	Sequence
<i>Forward primers</i>	
E1	AGCTTACTCCGGCCAAAAA
E2	GGATCCAAAGAGAGGCCAAC
E4	GCAGATGATGTTTCCTGTCC
E7	GGTCGTCAGACACCAAAACA
E9	TGTGACAGACAGTGAAAAGAGAA
E10.6	GCCTCTGAAAAGTGGACTGGA
E10.7	AAAAACACAGAAGGAATCGTCA
E10	CTAATTAAGTGTTCAGCCAG
E11.29	GCAGGCTTCACCTAAAAACG
E11	CCAAGTCATGCCACACATTC
E14	GAGACACTTGATTACTACAG
E15	GAATGCCAGAGATATACAG
E16.2	ATGGCGTTTCTAAACATTGC
E19	CATCGCTTTTCAGTGATGGA
E21	AGTTCGTGCTTTGCAAGATG
E22	AATCATTTTGTCTAGTAAGGTC
E25	AGGGCCACTTTCAAGAG
<i>Reverse primers (Fam-labeled)</i>	
E3.1	ATTGGTACAGCGGCAGAGTC
E4	GGACAGGAAACATCATCTGCT
E8	ATCATGAGGAAAATACAGTTTCAGAT
E10	TTCCAATGTGGTCTTTGC
E10.3	TCTTTCTTTCTTTTGTCTCTGTGTC
E11	CAGAATTGTCCCAAAAGA
E11.6	CGCAGGCTTGCTCAGTTTCTTT
E11.7	TGTACTCTTCTGCAATATGT
E12	GCCATCTGGAGTGCTTTTTG
E13	AGTTGTGCCAAAGGGTAC
E15	TTCGCATATCCTGTATATCTCTGG
E16.1	ACCAAAATAATCTTCAGTGTGAAAC
E17.1	CCATAGCTGCCAGTTTCCAT
E19	GGGTCAGGAAAGAATCCAAG
E22	TCATTCAACATTTGCCTGTGA
E23	TTTGTAGATTTTGAAGTTGCAAGA
E24	CCGTGGCTGGTAAATCTGA
E27	ATGCAAGTTCCTCGTCAGCT

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Supp.Table S3. In silico analysis of splicing regulatory elements.

		Splicing Enhancer Elements				Splicing Silencer Elements				
Location	HGVS description ¹	ESE Finder	RESCUE-ESE	PESE	EIE	Silencer motifs	ESS	Fas-ESS	PESS	IIE
Exon 2	c.-39-26G>A	-	-	-	2 new sites	-	-	2 sites broken	-	1 new site
Exon 5	c.426-89T>C	2 new sites	1 new / 1 broken	-	4 sites broken	-	-	-	1 site broken	2 sites broken
Exon 6	c.516+14C>T	1 new site	site broken	1 site broken	3 sites broken	1 site broken	-	-	-	3 new sites
Exon 7	c.631+25C>T	-	3 sites broken	-	4 sites broken	1 site broken	-	-	1 new site	2 new sites
Exon 8	c.680C>T (p.Ala227Val)	2 sites broken	-	1 site broken	-	1 new site	-	-	-	2 new sites
	c.681+69C>T	2 sites broken	-	-	-	-	-	2 new / 1 broken	-	2 new sites
Exon 9	c.682-2A>G	-	3 sites broken	1 new site	3 sites broken	1 new site	-	-	2 sites broken	-
Exon 10	c.1773_1776 delTTAT (p.Ile591fs)	-	-	-	3 new / 4 broken	-	-	-	4 sites broken	4 sites broken
Exon 11	c.3264dupT (p.Gln1089fs)	1 site broken	2 new sites	new site	2 site broken	1 site broken	-	-	-	-
	c.6825G>A (p.Glu2275=)	1 site broken	1 new site	5 new sites	1 new site	1 site broken	-	-	-	-
Exon 12	c.6928A>C (p.Thr2310Pro)	1 new/ 2 broken	1 site broken	1 new site	1 new site	1 new site	-	-	-	-
Exon 14	c.7435+8A>G	1 new site	-	-	1 new/1 broken	1 new site	-	-	-	2 new sites
Exon 15	c.7469T>C (p.Ile2490Thr)	-	-	-	1 new site	1 site broken	-	-	-	-
	c.7617+1G>A	2 sites broken	2 new sites	-	4 new sites	1 new site	-	1 site broken	1 new site	1 site broken
Exon 16	c.7626G>A (p.Thr2542=)	-	site broken	-	-	-	-	-	2 new sites	1 site broken
	c.7796A>G (p.Glu2599Gly)	-	-	2 sites broken	1 site broken	1 new site	-	-	1 site broken	-
	c.7976+5G>T	-	-	-	2 sites broken	2 sites broken	-	1 site broken	1 site broken	2 new sites
	c.7976+57G>C	1 new site	site broken	-	-	2 new / 1 broken	-	2 sites broken	-	-
Exon 18	c.8010_8032del (p.Ala2671fs)	2 sites broken	6 sites broken	1 site broken	1new/14 broken	1 new/4 broken	-	1 new site	1 site broken	2 sites broken
	c.8149G>T (p.Ala2717Ser)	new site	1 site broken	1 site broken	3 sites broken	2 sites broken	-	-	-	1 new site
Exon 20	c.8524C>T (p.Arg2842Cys)	-	-	-	-	1 site broken	-	-	-	4 new sites
	c.8503T>C (p.Ser2835Pro)	2 new sites	site broken	1 new site	1 new /2 broken	-	-	-	-	-
	c.8592C>T (p.Ala2864=)	-	-	1 new site	3 sites broken	-	-	-	-	2 new sites
Exon 22	c.8850G>T (p.Lys2950Asn)	1 new site	site broken	-	4 new / 1 broken	1 site broken	-	-	1 new site	-
	c.8851G>A (p.Ala2951Thr)	1 new site	2 new sites	2 new sites	6 new sites	1 site broken	-	-	-	-
Exon 23	c.8954-5A>G	6 new sites	2 sites broken	-	2 new sites	2 sites broken	-	-	-	-
	c.9004G>A (p.Asp3002Lys)	2 sites broken	3 sites broken	4 sites broken	-	1 new/ 2 broken	-	-	-	-
	c.9026_9030del (p.Tyr3009fs)	1 new site	-	-	2 sites broken	1 new/1 broken	-	-	-	2 new/3 broken
	c.9038C>T (p.Thr3013Ile)	1 site broken	4 sites broken	1 site broken	2 sites broken	1 site broken	-	-	-	-
Exon 27	c.9976A>T (p.Lys3326Ter)	-	-	-	1 site broken	-	-	-	2 new sites	3 new sites

¹ Description according to HGVS guide-lines (<http://www.hgvs.org/mutnomen>). Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the GeneBank reference sequence NM_000059.3

Supp. Table S4. Haplotype analysis performed in 18 c.7617+1G>A carriers identified in 8 Spanish and 3 Danish families.

Origen	FAMILY ID	ID	-0,45 Mb	-0,28 Mb	-0,18 Mb	-64 kB	-40 kB	-23 Kb	0	+6 Kb	+0,17Mb	+0,28MB	+1Mb	+1,29 Mb	+14Mb
Spain	HCSC_90	1	159/169	145/149	166/178	218/218	G/G	A/A	G/A	T/C	281/285	230/240	227/231	213/219	143/157
		2	159/161	145/149	158/178	218/218	G/G	A/C	G/A	T/T	281/293	230/240	227/231	219/219	143/159
		3	159/161	145/149	158/178	218/218	G/G	A/C	G/A	T/T	281/293	230/240	227/231	219/219	143/159
	HCSC_580	4	159/163	149/149	158/178	218/218	A/G	A/C	G/A	T/T	281/285	234/240	227/231	219/219	151/155
		5	159/163	149/149	158/178	218/218	A/G	A/C	G/A	T/T	281/285	234/240	227/231	219/219	149/151
		6	159/165	145/149	162/178	218/218	G/G	A/A	G/A	T/C	269/281	230/240	231/235	207/219	149/155
	HCSC_830	7	159/161	149/149	158/178	218/218	A/G	A/C	G/A	T/C	281/293	240/244	231/235	207/219	152/155
	HPH-55	10	159/159	145/149	172/178	218/218	A/G	A/A	G/A	T/C	281/289	240/244	223/231	207/207	145/153
	lov-99-09	11	159/161	149/149	162/178	218/218	G/G	A/C	G/A	T/T	281/289	240/244	223/231	219/219	151/157
		12	159/161	149/149	162/178	218/218	G/G	A/C	G/A	T/T	281/289	240/244	223/231	219/219	147/157
	200	13	159/161	149/149	164/178	218/222	A/G	A/C	G/A	T/C	285/293	230/240	231/231	219/219	147/157
		14	159/161	145/149	166/178	211/218	G/G	A/C	G/A	T/T	281/285	230/240	227/231	207/219	147/149
	1329	15	159/169	145/149	162/178	214/218	A/G	A/C	G/A	T/C	281/293	234/240	223/235	207/213	157/161
	696	16	159/165	145/149	174/178	218/222	G/G	A/A	G/A	T/C	281/289	240/244	223/231	207/219	151/157
		17	159/151	145/149	176/178	218/218	G/G	A/C	G/A	T/C	281/281	240/244	223/231	207/219	145/157
Denmark	D1	1	161/167	145/149	160/174	218/218	A/G	A/A	G/A	T/C	289/293	244/244	227/227	207/207	151/157
	D2	2	167/167	145/149	160/166	218/218	G/G	A/A	G/A	T/T	289/289	244/244	227/235	207/209	141/149
	D3	3	165/167	145/149	160/162	218/218	G/G	A/A	G/A	T/T	289/297	244/244	223/235	207/221	153/163

All eight families carrying this disease-causing mutation were resident in the Barcelona or Madrid metropolitan areas. We learnt through personal interviews that 7 families (we lack information on the remaining one) originated in the very same West-Spanish region (Extremadura). The length of the shared haplotype ranged from 0.51Mb to 1.74Mb. Interestingly, *BRCA2* c.7617+1G>A has been reported previously as a Danish founder mutation highly prevalent in the Vejle area [Thomassen et al., 2011]. The authors used SNPs arrays to define a shared haplotype of 0.2Mb to 3.3Mb in 19 families carrying this mutation. For that reason, we have included in our analysis three Danish families carrying the *BRCA2* c.7617+1G>A data. Our data (STR-based) is compatible with, but do not provide convincing evidence supporting a common origin for the Spanish and Danish mutations.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Supp. Table S5. Haplotype analysis in 10 c.8954-5A>G carriers identified in 4 unrelated Spanish families

		-0,45 Mb	-1.1 Kb	0	+1 Mb	+1.29 Mb	+14 Mb
Family ID	ID	D13S260	rs9534262	c.8954-5	D13S1493	D13S267	D13153
HCSC_896	1	168/168	T/C	A/G	227/235	207/209	151/157
Get483	5	168/168	T/C	A/G	227/243	207/209	157/161
	6	166/168	T/T	A/G	227/243	207/207	151/157
	7	162/168	T/C	A/G	227/235	207/209	145/157
	9	160/164	T/C	A/G	227/235	207/219	141/149
IVO_305	10	166/168	T/T	A/G	223/227	207/213	157/159
IVO_209	13	160/166	T/C	A/G	223/227	207/207	147/155
	15	160/160	T/T	A/G	223/227	207/213	151/153
	17	160/166	T/C	A/G	223/227	207/219	149/153
	19	160/166	T/C	A/G	223/227	207/219	153/153

All four families carrying this disease-causing mutation were residents in the Madrid or Valencia metropolitan areas. Personal interviews did not reveal any obvious common origin for these families. However, haplotype analysis identified a 1.28 Mb shared haplotype in all carriers, suggesting a founder effect (Supplemental Table 4). Families IVO_209 and IVO_305 have been reported previously [de Juan Jimenez et al., 2013]. BRCA2 c.8954-5A>G has been identified as well in three French families (<http://www.umd.be/BRCA2/>).