SPLICING FUNCTIONAL ANALYSIS OF DNA VARIANTS WITHIN THE BREAST CANCER TYPE 2 SUSCEPTIBILITY GENE (*BRCA2*), AND ITS EFFECT ON HEREDITARY *BREAST AND OVARIAN CANCER* (HBOC): A HYBRID MINIGENE APPROACH.

Valeria Velásquez Zapata

Eladio Andrés Velasco Sampedro
ADVISOR

María José Caloca Roldán
CO-ADVISOR

Máster en Investigación Biomédica
Instituto de Biología y Genética Molecular- IBGM
Universidad de Valladolid
2015

TABLE OF CONTENTS

1 INTRODUCTION

1.1 SPLICING DEFINITION

Splicing is a complex process that revolutionized the gene concept since its discovery in the 70s (Berget and Sharp, 1977; Gelinas and Roberts, 1977; Louise T Chow, Richard E. Gelinas, Thomas R., 1977). It soon became clear that the gene was not a simple functional unit or a heredity element, but rather a series of protein-coding sequences called exons, separated by long noncoding stretches called introns (Gerstein et al., 2007). Those introns must been excised from the gene transcript in order to obtain a mature mRNA. Some of them have an autocatalytic function (type I and II introns); and other are catalyzed by the spliceosome (spliceosomal introns, which are confined to the nuclear genome of all eukaryotes). The spliceosome is a complex ribonucleoprotein composed by five small RNAs (U1, U2, U4, U5, and U6 snRNAs) and more than 200 proteins (Rodríguez-Trelles et al., 2006). Therefore, splicing process is defined as the removal of introns and subsequent rejoining of exons.

The number of introns in a particular pre-mRNA can vary depending of the organism: in higher eukaryotes, it goes from zero to more than fifty (in vertebrates it ranges from 5.2 – Fugu- to 7.9 –Gallus-, with 7.8 introns per gene in humans) (Sakharkar et al., 2004). Furthermore, whereas exons are relatively short (10-400 nucleotides), intron sizes can reach more than 200,000 nucleotides (Krämer, 1996; Sakharkar et al., 2004). Additionally, eukaryote genes suffer an interesting phenomenon, known as alternative splicing (AS), which consists in the generation of multiple transcript variants, by differentially processing introns and exons. This process can result in an on-off switch for a particular gene; it can expand gene molecular diversity (by generating different mRNA and protein isoforms); and also may explain the apparent low number of genes in eukaryotes, especially the ones whose AS rates are high (such as mammals -in humans for example, it occurs in 95% of multiexon genes-) (Claverie, 2001).

AS patterns may vary from retention of introns or skipping of an entire exon, to more complex arrangements, as the production of multiple 5' and 3' splice sites (ss) within an intron, which would induce exon truncations/expansions (Irimia and Roy, 2014). These truncations generate prematurely terminated proteins upon translation, and can alter severely the protein function; so it may be expected this process would be highly regulated. Nonetheless, there remains a poorly understanding of the AS regulation and its functional implications (as it is not clear in which cases AS is functional or simply splicing "noise") (Ast, 2004; Pajares et al., 2007).

1.2 CONSTITUTIVE AND ALTERNATIVE SPLICING PROCESSES

CONSTITUTIVE SPLICING:

As it has been explained before, the excision of introns from pre-mRNAs by splicing process is an essential function in all eukaryotic organisms. In order to understand and classify this process, there have been described two types of spliceosomal introns. 1) The canonical introns (>99.5%) that start with the nucleotides GT and end with AG, and are catalyzed by the U2-dependent spliceosome. They are flanked by the 5' and 3' ss, which are called as

the splicing donor and acceptor. 2) The other ones are the non-canonical introns, which usually end with AT-AC, and are catalyzed by the U12-dependent spliceosome. They are found in vertebrates, insects, and plants, suggesting a common origin in those lineages with subsequent lost in many interior clades (Rodríguez-Trelles et al., 2006).

Depending on the spliceosome type, intron defining information in a pre-mRNA must comprise different conserved sequences (see Figure 1): 1) the 5′ and 3′ ss; 2) the branch point (BP), whose localization is unknown (some authors estimate it as far of 3′ ss as 200 nucleotides); and in higher eukaryotes, 3) a polypyrimidine tract or PPT, typically located 5-16 nucleotides upstream from the 3'ss (Sieliwanowicz, 1988; Will and Lührmann, 2011). Usually, large introns contain numerous spurious recognition sites and this information is insufficient for proper spliceosomal recognition (Lynch, 2006), so it becomes necessary additional cis-acting information elements such as exonic and intronic splicing enhancers (ESEs and ISEs) or silencers (ESSs and ISSs). They are typically short and diverse in sequence - four to ten nucleotides in length- and modulate constitutive and alternative splicing by binding regulatory proteins (which control assembly of spliceosomal complexes at an adjacent splice site) (Sieliwanowicz, 1988; Will and Lührmann, 2011).
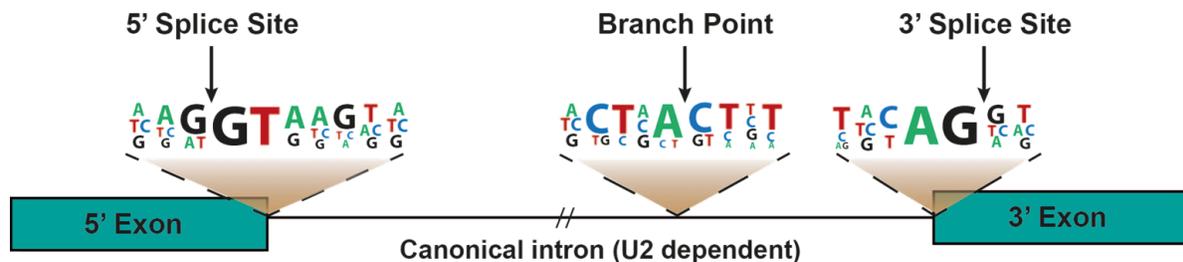
Figure 1. Consensus cis-acting sequences of intron recognition.

Pre-mRNA introns are removed by two trans-esterification reactions. First, within the intron, the 2′ OH group of the BP adenosine carries out a nucleophilic attack on the 5′ ss. Then, this site is cleaved and the 5' ss is ligated to the BP adenosine, forming a lariat structure. Second, the 5′ exon 3′-OH group attack the 3′ ss, inducing the ligation of the 5′ and 3′ exons, and consequently, the intron release (Moore et al., 1993). In order for these reactions to take place, the spliceosome must assemble and interact with the pre-mRNA through two main splicing assembling pathways: the U2-dependent spliceosome, which is composed from the U1, U2, U5, and U4/U6 snRNPs and numerous non-snRNP proteins; and the U12-dependent spliceosome, comprised by U11, U12, U5, and U4atac/U6atac snRNPs (Will and Lührmann, 2011).

ALTERNATIVE SPLICING:

Although splicing is a high-regulated process, there are deviations from a preferred gene-splicing pattern to other forms of mature mRNA (see Figure 2). The exon recognition is given by the interaction between cis-acting elements and trans-acting factors (splicing code, Figure 3), which collaborate in the promotion or inhibition of spliceosome assembly around the weak splice sites. It has been seen that cis-acting elements function additively. Additionally, the enhancer elements have dominant roles in constitutive splicing, while the silencers have relatively more control in the alternative splicing. Cis-acting elements include:

1) the promoters, ESEs and ISEs (which are bound by positive trans-acting factors, such as SR proteins); and 2) the inhibitors, ESSs and ISSs (which are bound by negative acting factors, such as heterogeneous nuclear ribonucleoproteins –hnRNPs-) (Wang et al., 2014).
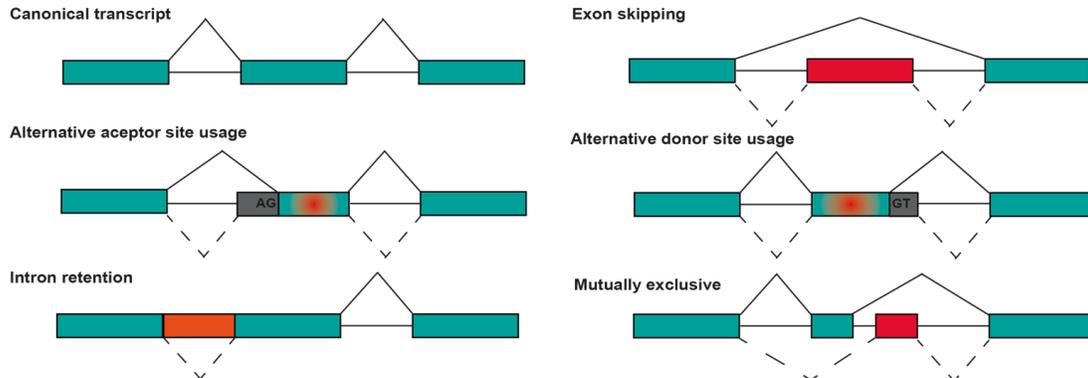
Figure 2. Types of splicing patterns. Continuous lines indicate splicing reactions of every pattern, while the discontinuous ones show the expected reaction of a canonical transcript.

Referring to trans-acting factors, SR proteins are the most important positive-acting ones. They bind ESEs through their RNA recognition motifs (RRMs) and mediate protein interactions through their RS domains, which change its phosphorylation state during alternative splicing regulation. On the other hand, hnRNPs are the most important negative trans-acting factors. Their mechanisms of action are not fully understood, but it has been seen that they bind to the ESS excluding SR proteins, repress the spliceosome binding to the splice sites, or create a exon sequestration through the creation of a pre-RNAm loop (Han et al., 2010; Wang et al., 2014).
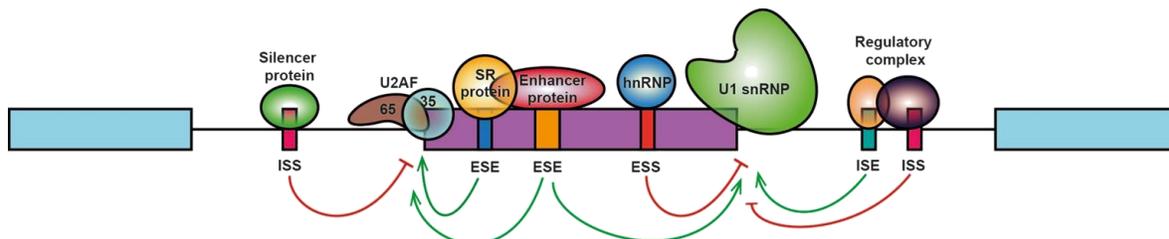
Figure 3. Cis-acting and trans-acting factors in splicing process. Green arrows indicate activation, and red lines indicate inhibition of splice site recognition.

1.3 SPLICING MUTATIONS AND DISEASE

ABERRANT SPLICING PATTERNS AND PATHOLOGY:

As noted above, the splicing is a dynamical and high-regulated process that seeks the preservation of precise exon-intron junction recognition by the spliceosome, and the generation of different and functional mRNA variants, increasing the molecular diversity of eukaryote genes. In order to achieve that, there should be a correct conservation of splice sites, and an equilibrium of cis-acting and trans-acting factors (particularly important when a gene has large introns and spurious recognition sites) (Lynch, 2006; Sieliwanowicz, 1988; Will and Lührmann, 2011). However, if splicing elements were mutated, they would induce

an aberrant splicing pattern, causing disease, increasing its susceptibility or modifying the severity of the deleterious phenotype (Padgett, 2012; Wang and Cooper, 2007). Remarkably, recent evidence shows how the splicing motifs as mutational targets may induce a range of 15% to more than 60% of human genetic disorders (Acedo et al., 2015; Lynch, 2006; Wang and Cooper, 2007).

This wide range of splicing mutations associated with genetic diseases suggests: in first place, our lack of knowledge about the splicing code; second, the deficiency in spliceogenic variant detection; and then, the necessity of increasing research in this topic (Wang and Cooper, 2007). There are important evidences that support these assertions: 1) 300 genes with splicing mutations are associated with 370 diseases (Wang et al., 2012); 2) at least 10% of the inherited diseases are related with mutations upon intron-exon junction (Padgett, 2012); and 3) It seems that mutations outside the canonical splice sites may account for a bigger percentage, since this type of variants have been routinely overlooked (Krawczak et al., 2007; Padgett, 2012).

Although it is prevailing the research conducted through the roles of splicing in disease, it already has given us insights about the splicing mechanisms and its regulation. For example, the discovery that most pathogenic 5′ ss mutations give rise to exon skipping rather than intron retention, supports the exon definition concept; or, that the effects of exon mutations on splicing contributed to the description of splicing code. Technology advances (which must be understood as new molecular biology techniques and high-throughput methods) will allow the implementation of these discoveries in clinic, and the knowledge application toward diagnosis and treatment. In order to successfully predict the effect of splicing mutations in disease, it is necessary to implement an holistic approximation, which integrates computational methods with experimental evidence (Wang and Cooper, 2007).

Researchers have also classified different types of splicing mutations effects on disease. Cis-acting mutations comprise splice site variants, single-nucleotide substitutions, and micro insertions/deletions. These variants take place primarily within exons, which can cause a splicing defect, rather than changes in the predicted protein-coding sequence (since aberrant splicing and mRNA degradation mechanisms prevail over mutated protein expression). The uncertainty about the influence of these variants on disease is due to its ample range of sensitivities within a specific gene, which is directly related with the binding of the splicing elements (Wang and Cooper, 2007). Trans-acting mutations involve defects in the splicing machinery. Normally, a null mutation in the basal splicing machinery leads to lethal consequences (Faustino et al., 2003). Despite this observation, there are two diseases associated with the assembly or function of the snRNPs, in which two different groups of neurons are affected: spinal muscular atrophy and retinitis pigmentosa. Additionally, there are diseases linked with trans-acting mutations (which are related with the disruption of alternative splicing regulators), such as myotonic dystrophy or cancer (Faustino et al., 2003; Wang and Cooper, 2007). While a cis-acting mutation is expected to be restricted to a specific exon, a trans-acting one may have a global effect (because these regulate thousands of alternative splicing events in the transcriptome) (Acedo, 2013; Wang and Cooper, 2007). Furthermore, it is necessary to consider that a specific mutation effect may range from being the disease cause to affecting its susceptibility.

SPLICING MUTATIONS IN CANCER:

During cancer progression, a common feature involves the aberrant splicing of the genes implicated. The main affected genes throughout the transition from a normal cell function to the development of malignancy are those that regulate apoptosis, cell migration, cell growth, hormones responsiveness, cell–cell and cell–matrix interactions, and response to chemotherapy. Currently, the evidence has shown how cis-acting mutations on oncogenes, tumor suppressors and other cancer-relevant genes have a direct impact in cancer initiation and progression; however, the majority of the splicing changes must be due to trans-acting mutations, since there is no detection of changes within the genes involved. Besides these findings, there remain several important issues about the role of splicing in cancer. In first place, there is uncertainty about the extent of influence of splicing mutations on cancer initiation and progression. Second, if the previous premise is true, what are the most important genes involved in this process? Third, what is the effect of such mutations on the protein expression and properties? Those questions give rise to an ample range of diagnosis and therapeutic possibilities such as: the use of the alternative splicing signatures to predict a pathological outcome and evaluate a treatment efficiency; or the development of drug searching using cancer-specific splicing patterns as targeting mechanism (Faustino et al., 2003; Wang and Cooper, 2007). According to the literature, cancer-associated mutations in cis-elements (inherited or acquired) are being underestimated. Though there are many examples of this type of cancer-associated alterations, in only a few cases a cause-effect relationship has been proved (Srebrow and Kornblihtt, 2006). This is due to the traditional mutation screening approach, which focuses on the primary structure analysis of genomic DNA, and predicted protein effect, ignoring alterations on splicing.

First, as cancer-associated splicing mutations are studied, there are being gathered more genes related to them. One example is Adenomatous Polyposis Coli gene (*APC*) which contains a AG>AT mutation on the acceptor site of the exon 4, frustrating its recognition, and implicating it in hepatic and colorectal cancer (Kurahashi et al., 1995). Others cancer-associated genes also have this kind of genetic variants, such as the tumor suppressor gene *hSNF5*, the estrogen receptor gene (*ER*), the neurofibromatosis type 1 and 2 genes (*NF1* and 2), the tumor suppressor (*LKB1*), the oncogene *KIT*, the liver intestine cadherin (*CDH17*), and the Kruppel-like Znfinger transcription factor gene (*KLF6*) (Colapietro et al., 2003; De Klein et al., 1998; Srebrow and Kornblihtt, 2006; Taylor et al., 2000; Wang et al., 1997). Second, in order to determine the real frequency of splicing variants, it is required a comparative analysis of DNA, mRNA, and even protein sequences. To exemplify this point, a study of 80 unrelated patients with neurofibromatosis type 1 was conducted to show how the percentage of splicing mutations varies according to the nucleic acid used during the analysis: it increased from 37% to 50%, by shifting from DNA to mRNA analysis. Third, it is a fact that some exonic nonsense mutations are associated with exon skipping rather than the introduction of a premature stop codon (Pajares et al., 2007).

Regarding to cancer-associated trans-acting mutations, it has been found a correlation between the splicing regulatory factor expression (mostly SR proteins and hnRNPs) and the aberrant alternative splicing transitions, which are associated with malignant transformation (Faustino et al., 2003). One of them is related to changes of the relative abundance of specific SR proteins during cancer progression. These changes can be used as a pre-neoplasia marker, nonetheless, they are not predictive in the determination of tumor

incidence or invasiveness, which indicates that they are only one factor within a complex of cellular changes for neoplasia and malignancy (Faustino et al., 2003). An increase of SR protein expression is associated to mammary gland tumorigenesis in a well-characterized mouse model and various human cancers (Kaida et al., 2012; Nowak et al., 2008; Stickeler et al., 1999). In addition, hnRNP protein family is also implicated in cancer development controlling alternative splicing of the *IG20/MADD* and *Ron* genes (Kaida et al., 2012).

HEREDITARY BREAST AND OVARIAN CANCER (HBOC):

Breast cancer is the most common malignancy and the second most common worldwide cause of death by cancer in women. In 2012, 1.67 million new cases were diagnosed, reaching a 25% of all cancers in females. The statistics indicate that one in eight people of this gender will suffer from this disease. Additionally, if she has first-degree relatives who suffer of breast cancer, her risk of developing the disease is twice than general population. This hereditary component of the pathology, which is called hereditary breast and ovarian cancer (HBOC), ranges from 10% to 30% of the cases (25–40% when the patients are under 35 years old). However, only 5%–10% are identified with a strong inherited component, and 4%-5% are explained by mutations in high penetrant genes, which have an autosomal dominant inheritance (Stratton and Rahman, 2008).

The most important high penetrant genes involved in HBOC are the breast cancer type 1 and 2 susceptibility genes (*BRCA1* and *BRCA2*), which cause 16% of HBOC cases, and whose pathogenic variants multiply the risk of developing breast cancer between 10 to 20 times. The HBOC predisposition has an autosomal dominant inheritance throughout families who carry mutations in the *BRCA* genes; however, at cellular level, they act as recessive genes (Lux et al., 2006). These genes encode nuclear phosphoproteins, which have tumor suppressor activity, through the maintenance of genomic stability. Despite they act in a common genome protection pathway, they have different roles: BRCA1 is a protein involved in checkpoint activation and DNA repair, whereas BRCA2 is involved in the core mechanism of homologous recombination. Currently, the associations between these proteins remains undefined, but it is clear they must exist, since the germline mutations of the genes have an obvious similarity regarding their cancer susceptibility (Roy et al., 2011).

BRCA1 gene has 81.19 Kb of genomic sequence that contains 23 exons, 22 of them protein-coding. Its transcription generates an mRNA of 7094nt, whose translation produce a protein of 1863 aminoacids and 220 kDa. The protein contains several functional domains (Figure 4): 1) an amino-terminus RING domain, which is important for the interaction with other proteins; 2) a pair of nuclear localization signals (NLS) in the core region; 3) several phosphorylation sites; 4) a helical domain to interact with PALB2 and BRCA2; and 5) a BRCT domain at the C terminus. The protein interacts with other proteins to form the BRCA1-associated genome surveillance complex, which participates in regulating multiple cellular processes as: gene stability maintenance, cell cycle control, ubiquitination, transcription, chromatin remodeling and p53-independent apoptosis (Linger and Kruk, 2010). Mutations within this gene cause a risk of breast and ovarian cancer as high as 80% and 40%, respectively (Apostolou and Fostira, 2013).

On the other hand, BRCA2 is involved in the repair of double-strand DNA breaks (DSBs) through the homologous recombination (HR) pathway. It is a large gene, which comprises

27 exons (26 of them protein-coding) and 84.2 Kb. The BRCA2 resulting protein contains 3418 amino acids and 384 KDa. It has several functional domains (see Figure 4): 1) at the N-terminus it has a PALB2 interaction domain; then, within the gene, it contains 2) eight BRC repeats that bind the RAD51 recombinase; at the C-terminus a 3) DNA-binding domain (DBD), whose functions is the binding with single and double stranded DNA (Roy et al., 2011). It has been found that male carriers of *BRCA2* mutations have a risk of 20% of developing prostate cancer, a 6% of breast cancer, and a 3% of pancreatic cancer. Female carriers have a risk between 26% to 84% for breast cancer, and 20% for ovarian cancer (Apostolou and Fostira, 2013). A large number of pathogenic variants are classified as frameshift and nonsense, but most DNA changes (including ~900 different missense mutations from the BIC database) are variants of unknown clinical significance (VUSs) (Apostolou and Fostira, 2013).
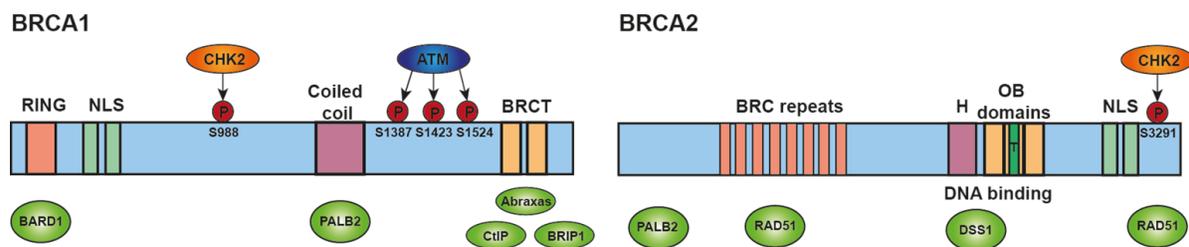
Figure 4. Functional domains of BRCA1/2 proteins.
Green circles indicate the interaction with other proteins

Research focusing on the *BRCA1* and *BRCA2* aberrant genetic variants have barely linked them to 16% of cases of HBOC, since there is a high number of inconclusive results due to variants of unknown significance. Then, how to explain the 80% remaining cases with a strong family HBOC record? There have been proposed several approaches to answer this question: first, researchers have identified different mutated genes that conferee a moderate risk (as *ATM, BRIP1, CHEK2, PALB2, or RAD51*), nonetheless, their frequency is very low to explain this high percentage (Acedo, 2013). These genes mostly codify proteins involved in DNA repair, but there is no explanation to the question of why damage of these genes specifically increase the susceptibility to HBOC. The most accepted theory suggests that menstrual cycle causes oxidative DNA damage driven by hormonal growth. This oxidative DNA damage induces replication stress and consequently, DSBs, which requires the BRCA1–BRCA2–HR pathway to be repaired (Roy et al., 2011).

Second, genome-wide association studies have identified 41 common single nucleotide polymorphisms (SNP) associated with the HBOC susceptibility, however the authors suggest that more than 1,000 additional loci are involved (Michailidou et al., 2013). This evidence suggest the importance of determining the real effect of the unclassified *BRCA1/2* genetic variants, which constitute the major limitation of *BRCA1* and *BRCA2* genetic testing. Most VUSs are mainly missense but also can be intronic or even silent variants. The interpretation of such variations constitutes a diagnosis problem, which jeopardizes the HBOC prevention and surveillance protocols. Currently, there are multiple approximations to evaluate the possible effect of a VUS variant. The *in silico* analysis tries to estimate the variant phylogenetic conservation and the possible protein modification it causes. Additionally, there are variant and disease segregation analysis; however, the best

way to improve the HBOC genetic testing is to clarify the functional effect of the VUS (Apostolou and Fostira, 2013).

SPLICING MUTATIONS WITHIN *BRCA* GENES:

At present, there are specialized DNA databases, which compile the updated *BRCA1/2* reported variants. The Breast Cancer Information Core -abbreviated as BIC- (http://research.nhgri.nih.gov/bic/) is one of the most important. At May of 2015, this database has reported 1781 *BRCA1* variants, and 2000 to *BRCA2*. Most of them are single-nucleotide substitutions, but also there are small insertions and deletions. BIC database classifies the mutations according to the protein translation predicted effect: as frameshift, nonsense, missense, or synonymous. Additionally it indicates the in frame Ins/Dels, the mutations within introns or UTR regions, and finally those that alter the splicing process.

Interestingly, about one-half of the total BIC reported variants remain as VUSs. Our laboratory formerly has reported one third of disease-causing variants as splicing-disrupting (belonging to 14 exons of *BRCA1/2*) (Acedo et al., 2012; Sanz et al., 2010). Furthermore, recent studies showed that splicing variants are enriched in cancer-related genes (Acedo et al., 2015; Sterne-weiler and Sanford, 2014). The characterization of an aberrant splicing variant is a complex task, which requires a functional validation, since a completely accurate bioinformatic prediction of its effect on the mRNA processing is not currently available (a number of studies have failed to confirm experimentally the initial *in silico* findings). Subsequently, there is an increasing research effort to describe the functional effect of such variants, in both *BRCA1* and *BRCA2*, and its relationship with HBOC. Just to give an example, the first proved splicing deleterious variant of *BRCA1* was an inherited mutation in exon 18, classified as nonsense. It actually disrupts an ESE sequence and provokes exon 18 skipping (Mazoyer et al., 1998). Since then, the splicing enhancer and silencer mutation descriptions have considerably grown (Srebrow and Kornblihtt, 2006). There are also examples of *BRCA2* that have been deeply described:

1) The c.145C>T mutation within exon 3 is already annotated at BIC as nonsense. Our laboratory demonstrated that it actually affects splicing, causing a high percentage of transcripts with exon 3 skipping. Bioinformatic analysis indicates that it eliminates an ESE sequence that binds to the SF2/ASF factor (Sanz et al., 2010).

2) The c.470_474del5 mutation in exon 5 is registered in the BIC database as frameshift. This mutation affects splicing by excluding the majority of exon 5, due to the elimination of a ESE sequence (Sanz et al., 2010).

3) The mutation c.8331G>A in exon 18 initially had an unknown clinical effect. Splicing functional assays have shown that it alters exon 18 donor site causing exon skipping. The variant finally causes the production of a truncated protein (Sanz et al., 2010).

4) It has been proved that missense mutations with an unknown clinical effect can alter splicing. The c.517G>T in exon 7 of *BRCA2*, does it. In this particular example, a significant percentage of transcripts have exon 7 skipping, due to the disruption of the acceptor sequence (Gaildrat et al., 2012).

5) Normally, intronic splice site mutations are classified as clinically significant, since they affect splicing. However, there are other deleterious intronic variants, as the c.6937+594T>G within the intron 12 of *BRCA2*. It activates a cryptic splicing site, inducing the a pseudo-exon introduction in the mature mRNA (Anczuków et al., 2012). There are other examples as the intron 2 PPT mutation c.68-7T>A, which promotes partial exon 3 skipping (Sanz et al., 2010).

Altogether, these results demonstrate how crucial is to implement holistic approaches to interpret the molecular basis of a genetic disease, and to define the nucleotide substitutions within coding regions that act at the translational or the splicing level. In order to achieve this, it is necessary to combine the use of bioinformatics tools with RNA functional assays, which can be developed though the implementation of RT-PCR from patient's RNA samples, or by the use of hybrid minigenes. The first one is the most straightforward and direct method to assess the clinical relevance of DNA changes, but collecting patient samples is a difficult task, so it is possible to overcome the issue by using the hybrid minigene technology.

This work focuses on the study of splicing variants within the exon 17 of *BRCA2*. Remarkably, it possesses a non-canonical GC donor site, which represents less than 1% of these motifs, compared with the canonical GT site. In the non-canonical donor site exons, it is expected that the other nucleotides in the signal adhere more closely to the consensus sequence (A/C)AG|GC(A/G)AGT, apparently, compensating for the T to C substitution that is unfavorable for splicing. This rare class of donor splice signals also has been implicated in alternative splicing, being highly important the signal strength for splice site selection (Churbanov et al., 2008; Thanaraj and Clark, 2001). Based on this premise, splicing variants within this exon are more susceptible of being deleterious, which becomes it interesting for the study of its regulation.

1.4 HYPOTHESIS

A large proportion of DNA variants of any type alters pre-mRNA splicing in many human inherited diseases, including HBOC.

1.5 OBJECTIVES

GENERAL

We aim to assess exon 17 splicing variants involvement in the genetic susceptibility to HBOC, by a bioinformatic analysis and mRNA functional assays through a hybrid minigene strategy.

SPECIFIC

1) To select the candidate variants of the functional analysis based on a bioinformatic prediction, using online databases as BIC and algorithms as the presented on the Human Splicing Finder Database (HSF).
2) To construct and validate a wild type minigene that contains the 14-20 exons of *BRCA2* and can be used as a vector of the functional analysis.
3) To generate each of the selected *BRCA2* variants within the minigene.

4) To perform a functional analysis of the variants (driven in eukaryote cells), in order to evaluate its effect on the mRNA splicing.
5) To characterize the new-generated splicing patterns.
6) To contribute to the understanding of the HBOC genetic predisposition spectrum.

2 MATERIALS AND METHODS

2.1 BIOINFORMATIC ANALYSIS OF SPLICING VARIANTS

All the *BRCA2* exon 17 reported variants were searched at two *BRCA* databases: the BIC and the Universal mutation –UMD (http://www.umd.be/) (Caputo et al., 2012). Then, the mutant and normal sequences were analyzed with several bioinformatic tools to identify potential splicing mutations. Disruption/creation of splice sites was evaluated with NNSPLICE version 0.9 (http://www.fruitfly.org/seq_tools/splice.html) (Reese et al., 1997). Additionally, the human splicing finder -HSF version 2.4.1 (http://www.umd.be/HSF/) (Desmet et al., 2009) was used to predict branch point disruptions, enhancer and/or silencer modifications. Specifically, HSF computational tools predict disruption of SR proteins binding motifs such as SRp40, SC35, SF2/ASF, SRp55, hnRNP A1 (repressor), Tra2-b, and 9G8 protein binding motifs. The criteria to select the splicing variants that would be object of a functional analysis were: 1) acceptor or donor sites elimination or significant score reduction; 2) generation of new splice sites, mainly exonic ones; 3) disruption of the branch point; 4) mutations localized within positive microdeletions; 5) disruption of enhancer motifs or creation of silencers, focusing on hnRNPA1.

After the selection of the *BRCA2* DNA variants, the site directed mutagenesis primers were designed, by using Oligo 7 software (Molecular Biology Insights, Inc. DBA Oligo, Inc. Vondelpark, CO, USA).

2.2 CONSTRUCTION OF THE *BRCA2* EXON 14-20 MINIGENE

The *BRCA2* exon 17 mutation variants were introduced in the pSAD-minigene with exons 14-20 of *BRCA2*. It produced a canonical transcript of the expected exon composition and size (1809 nucleotides). This new splicing plasmid, pSAD (Patent P201231427-CSIC, Priority Patent Application filed), has a pSPL3 backbone with the following features: 1) intron size reduction of 1.2 kb; 2) introduction of Beta-Galactosidase (LacZ) as a second selection marker; 3) a new polylinker with 27 restriction sites (pSPL3, 10 sites); 4) a vector exon V2 acceptor site strengthening (NNSPLICE score 0.99 vs. 0.64) by triple mutagenesis; and 5) elimination of old pSPL3 restriction sites (HindIII, XbaI, and SalI) that interfere with those of the new multiple cloning site (MCS).

MGBR2EX14-20 (see Figure 5) was constructed in four steps, which comprise overlapping extension PCR or classical restriction digestion/ligation cloning. The construction process had three intermediate constructs: MGBR2EX17-18, MGBR2EX16-18, and MGBR2EX16-20. All the inserts were amplified with Phusion High Fidelity polymerase (Life Technologies, Carlsbad, CA, USA) and primers indicated on Table 1. Exons 17-18 were subcloned into the pSAD vector by an overlapping extension PCR. Then, exon 16 was added by the same technique. Exons 19-20 were inserted by cutting with XhoI and BamHI restriction enzymes, both the vector and the PCR product that contains the exons 19-20, and by ligating the

fragments. Finally, a similar process was conducted to introduce the exons 14-15, using the EagI and SacI restriction enzymes. To conclude, after each step of the minigene construction, a functional analysis was conducted to evaluate whether they generate a correct splicing pattern.

Table 1. Primers used on the MGBR2EX14-20 minigene.

Primer	5'-3' Sequence	Size
MGBR2_ex14-15-SacIFW	CACACAGAGCTCTGAAGATGTGAAGGTGAGAGAA	2405
MGBR2_ex14-15-EagIRV	CACACACGGCCGGTTCAGGGCTATCAGTTATTCA	
MGBR2_ex17-18_Ins-FW	GCTCTAGAACTAGTGGATCCCCCGGTCAGTATGATACTTTGATACATGT	1556
MGBR2_ex17-18_Ins-RW	ATAAGCTTGATATCGAATTCCTGCAGTGGAATAGGGATCTGATCAA	
MGBR2_ex16_Ins-FW	GCTCTAGAACTAGTGGATCCCCCGGCCCTTTTGTTTCCCATCTAAGT	805
MGBR2_ex16_Ins-RW	GACATGTATCAAAGTATCATACTGAATAAATGCCTAAGAAAAATGT	
MGBR2_ex19-20_XhoI-FW	CACACACTCGAGATAGCATTAAGAACTTGTAGCA	1188
MGBR2_ex19-20-KpnI-RV	CACACAGGTACCATTACAAATGGCTTAGACCTGA	
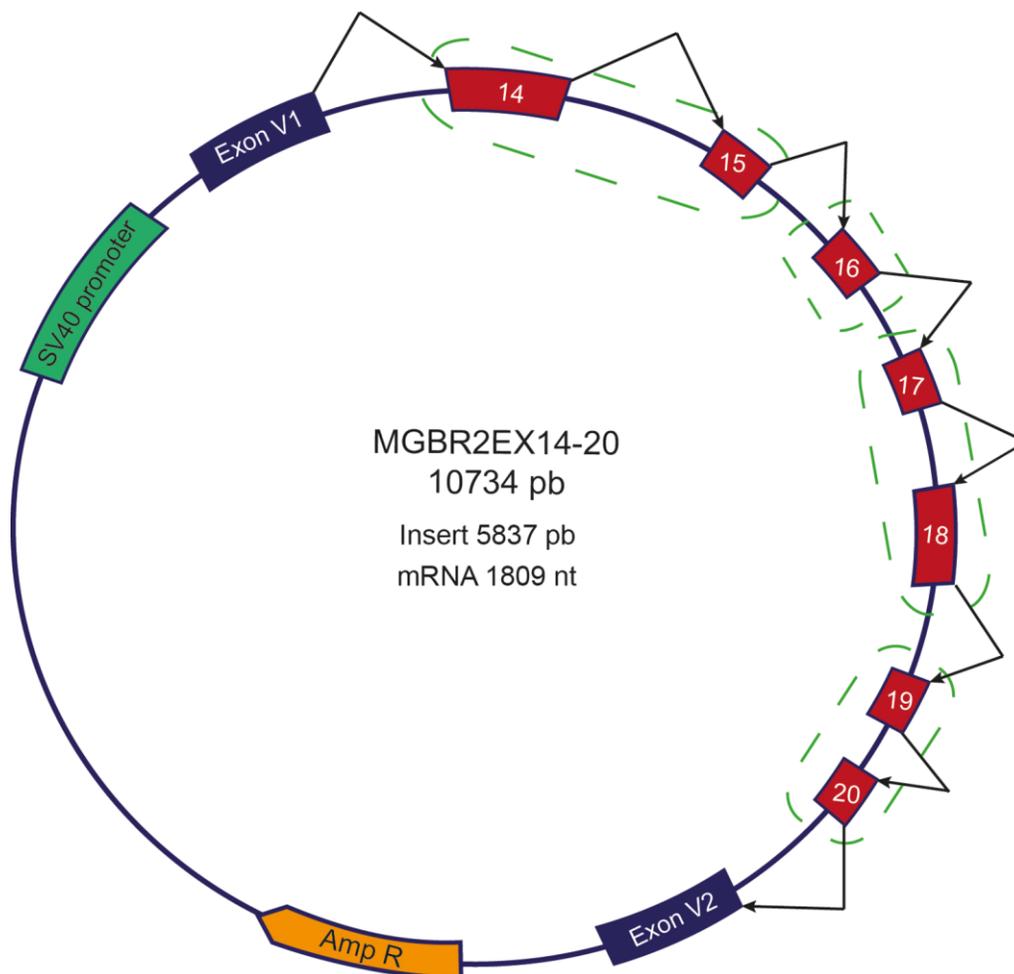
Figure 5. MGBR2EX14-20 structure inside the pSAD vector. The four cloning steps are shown as dashed green boxes. Arrows indicate expected splicing reactions in eukaryotic cells.

2.3 SITE-DIRECTED MUTAGENESIS

Mutagenesis was carried out with the Quick-change Lightning Site-Directed Mutagenesis kit (Agilent, Santa Clara, CA, USA). Wild type (wt) minigene MGBR2ex14–20 was used as template to generate 14 bioinformatically selected DNA variants, and confirmed by sequencing. Each mutant clone was cultured at 37 ºC for 24 hours, in LB liquid medium (Broth-LB Miller, Merk) supplemented with ampicillin at a final concentration of 100 ug/ml. Plasmid DNA minigenes were purified with the AccuPrep Plasmid Extraction kit (Bioneer, Alameda, CA, USA), and were stored for the subsequent analysis steps.

DNA was quantified with a Quawell 3000 Nanodrop spectrophotometer (ThemoScientific, Waltham, MA, USA), and then sequenced with the BigDye Terminator Cycle kit Sequencing v3.1 (Applied Biosystems, Carlsbad, CA, USA) following the manufacturer's instructions, adding 25ng/Kb of DNA, and depending on the variant position, primers that allow the mutagenesis identification: RT18-RV (5'TCCCTTTCCATTATCTTTTT3') or RT16-FW (5'TATGGACTGGAAAAGGAATAC3'). The capillary electrophoretic run was conducted in a sequencer automatic ABI Prism 3130 (Applied Biosystems) with a 36 cm- capillary array, using polymer POP-7 (Applied Biosystems) and running buffer with EDTA (Applied Biosystems). The analysis of chromatograms was performed with SnapGene 2.7.2 software (GSL Biotech LLC, Chicago, IL, USA).

2.4 FUNCTIONAL ANALYSIS OF SPLICING VARIANTS

Transfection of HeLa cells: 10^5 HeLa cells (human cervical carcinoma) were grown to 90% confluence with 0.5 ml of medium (DMEM, 10% fetal bovine serum, 1% glucose, and 1% penicillin/streptomycin) in four-well plates (Nunc, Roskilde, Denmark). Cells were transiently transfected with 1µg of each minigene and 2µl of Lipofectamine 2000 (Life Technologies). After 48 hours, nonsense-mediated decay (NMD) was inhibited in the cell culture by adding cycloheximide (Sigma–Aldrich, St. Louis, MO, USA) at a final concentration of 300µg/ml. After four hours, RNA was purified with the GeneMATRIX Universal RNA Purification Kit (EURx, Gdańsk, Poland) with on-column DNAase I treatment.

RT-PCR was carried out with 100 ng of RNA and the RevertAid H Minus First Strand cDNA Synthesis Kit (Life Technologies), using gene specific primer RTPSPL3-RV (5'TGAGGAGTGAATTGGTCGAA3'). Samples were incubated 25°C for 5 min, 42°C for 1 hour, and a final extension step at 70°C for 5 min. Then, cDNA was amplified with RTPSPL3-RV and RT16-FW using Platinum Taq DNA polymerase (Life Technologies). The final transcript had an expected size of 1015 nucleotides.

Previous purification of the cDNA with the AccuPrep PCR Purification kit (Bioneer), sequence reactions were performed using the kit BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) following the manufacturer's instructions. Semiquantitative fluorescent RT-PCRs were done in triplicate with a RT16-FW FAM-labeled primer. Seven nanograms of DNA from the FAM-labeled RT-PCR products was mixed with 20µl of Hi-Di Formamide and 1µl of LIZ-1200 Size Standard (Life Technologies). Samples were run on an ABI3130 sequencer and analyzed with the Peak Scanner software (Applied Biosystems). Only peaks with heights ≥ 50 RFU (Relative Fluorescence Units) were taken into account. Mean peak areas of each transcript and standard deviations were calculated.

3 RESULTS

3.1 BIOINFORMATIC ANALYSIS OF SPLICING VARIANTS

Sixty-six DNA variants within exon 17 of *BRCA2* were identified from the database search. Using the selection criteria, fourteen of them were finally chosen for functional analysis, and site directed mutagenesis primers were designed (Figure 6 and Table 2). According to the BIC and UMD annotations, five mutations were classified as Intervening sequences (intronic), five as missense mutations, two as frame shift mutations, and two as synonymous.
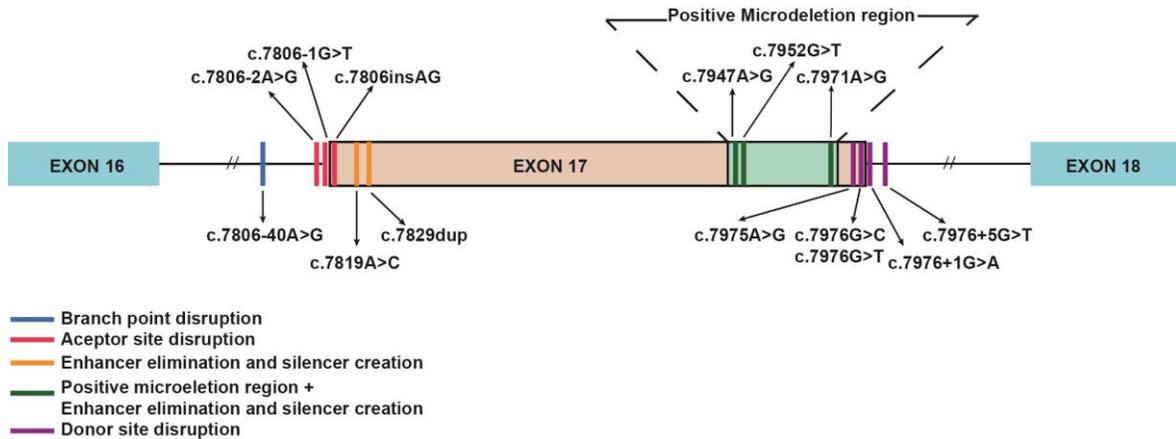
Figure 6. Exon 17 selected variants localization.
The different colors indicate the main criteria used to the selection.

When these results were contrasted with the bioinformatic analysis, eight variants were found to induce changes in the splice sites scores, affecting the ss consensus sequences, and then, they were selected for functional analysis. Three of them were associated with the acceptor site (c.7806-2A>G, c.7806-1G>T and c.7806insAG), and five with the donor site (c.7975A>G, c.7976G>C, c.7976G>T, c.7976+1G>A and c.7976+5G>T). Bioinformatics also predicted one disruption on a branch point site (c.7806-40A>G), and multiple enhancer disruptions and/or silencer creations (c.7819A>C and c.7829dup), which were also associated and prioritized if they belonged to a previously identified positive microdeletion region with putative active ESEs (Acedo, 2013) (c.7947A>G, c.7952G>T and c.7971A>G).

3.2 MINIGENE CONSTRUCTION

Since it is very difficult to obtain patient samples to test the presence and effect of *BRCA2* splicing variants, we opted to study the selected variants with a hybrid minigene approach. MGBR2EX14-20 was constructed in several cloning steps, as previously described (Acedo, 2013). Before the functional analysis, it was necessary to guarantee the fidelity and reproducibility of the minigene, which had to be able to create an adequate genomic context for variant analysis of exon 17 of *BRCA2*, object of this study. Therefore, each intermediate construct was functionally assayed. The amplification products were run on an electrophoresis gel and its size was checked (see Figure 7).

Table 2. Bioinformatic analysis of the variants selected to functional assay.
[+] and [-] symbols indicate creation or disruption, respectively, of splicing regulatory sequences.

DNA variant	Source	Variant type	Bioinformatics[2]	Site-directed Mutagenesis primers (5'- 3')
c.7806-40A>G	BIC	Intervening Sequence	**HSF: [-]branch site 45 nt upstream (67.63→38)**; [+]SRp55; [+]PESEs (26.79); [-] 2Sironi ESS (63.33; 61.92); [+]Fas-ESS; [-] hnRNP A1(66.19)	FW:GTTGAATTCAGTATCATCCTGTGTGGTTTTTATGATAATAT RV:ATATTATCATAAAAACCACACAGGATGATACTGAATTCAAC
c.7806-2A>G	BIC	Intervening Sequence	**[-] Aceptor site (NNS: 0.95 →<0.4)**; [-]9G8 (59.87); [+] 3 Sironi ESS (75.30;65.03;70.71); [-] hnRNP A1(82.86)	FW:TCTACTTTTATTTGTTCGGGGCTCTGTGTGACACT RV:AGTGTCACACAGAGCCCCGAACAAATAAAAGTAGA
c.7806-1G>T	BIC	Intervening Sequence	**[-] Aceptor site (NNS: 0.95 →<0.4)**; [-]SC35; [+] 2 ESS Zhang; [-] hnRNPA1 (82,86)	FW:CTACTTTTATTTGTTCATGGCTCTGTGTGACACTC RV:GAGTGTCACACAGAGCCATGAACAAATAAAAGTAG
c.7806insAG	BIC	Frame shift	**[-] Aceptor site (MaxEnt: 5.95 → -14.25)**; [+]hnRNPA1 (82,86)	FW:TCTACTTTTATTTGTTCAGAGGGCTCTGTGTGACACTCC RV:GGAGTGTCACACAGAGCCCTCTGAACAAATAAAAGTAGA
c.7819A>C	BIC	Missense	**[-]SF2/ASF (72.31), [-]SC35 (77.63), [-] 2 SRp40 (78.98;85.27)**; [+] 2 Sironi ESS (62.90; 63.22)	FW:TGTTCAGGGCTCTGTGTGACCCTCCAGGTGTGGATCCAAAG RV:CTTTGGATCCACACCTGGAGGGTCACACAGAGCCCTGAACA
c.7829dup	UMD	Frame shift	**[-]SC35 (83.34), [-] 2 SRp55 (78.60); [-] 2 Rescue-ESE; [-] 9G8 (64.90); [+] 2 Fas-ESS**	FW:TCTGTGTGACACTCCAGGTGTTGGATCCAAAGCTTATTTCTA RV:TAGAAATAAGCTTTGGATCCAACACCTGGAGTGTCACACAGA
c.7947A>G[1]	BIC	Synonymous	**[-]SF2/ASF (77.23), [-] SRp40 (79.40)**; [-] Rescue-ESE); [-] Sironi ESS (64.11)	FW:GCTAATAGATGCCTAAGCCCGGAAAGGGTGCTTCTTCAACT RV:AGTTGAAGAAGCACCCTTTCCGGGCTTAGGCATCTATTAGC
c.7952G>T[1]	UMD	Missense	**[-] SRp40 (79.40)**; [-] Rescue-ESE); [-] 2 Sironi ESS (64.93;61.98); [+] PESEs (28); [-] hnRNP A1 (83.81)	FW:TAGATGCCTAAGCCCAGAAATGGTGCTTCTTCAACTAAAAT RV:ATTTTAGTTGAAGAAGCACCATTTCTGGGCTTAGGCATCTA
c.7971A>G[1]	UMD	Synonymous	**[-] 2 Tra2-β (81.02;85.42); [+]hnRNPA1 (70,48)**	FW:AGGGTGCTTCTTCAACTAAAGTACAGGCAAGTTTAAAGCAT RV:ATGCTTTAAACTTGCCTGTACTTTAGTTGAAGAAGCACCCT
c.7975A>G	BIC	Missense	**[-] GC donor site (MaxEnt: 3.61 → -4.34)**; [-] SRp40 (91.86), [+] SRp55 (75.65); [-] hnRNPA1(79.29)	FW:CTTCTTCAACTAAAATACGGGCAAGTTTAAAGCATTA RV:TAATGCTTTAAACTTGCCCGTATTTTAGTTGAAGAAG
c.7976G>C	BIC	Missense	**[-] GC donor site (MaxEnt: 3.61 → -4.45)**; [+]SF2/ASF; [+] Rescue-ESSs; [-] 2hnRNPA1 (74.76, 66.43)	FW:TTCTTCAACTAAAATACACGCAAGTTTAAAGCATTAC RV:GTAATGCTTTAAACTTGCGTGTATTTTAGTTGAAGAA
c.7976G>A	BIC	Missense	**[-] GC donor site (MaxEnt: 3.61 → -5.13)**; [-] Sironi ESS (69.76);[-] 2 hnRNPA1(74.76;66.43)	FW:GCTTCTTCAACTAAAATACAAGCAAGTTTAAAGCATTACAT RV:ATGTAATGCTTTAAACTTGCTTGTATTTTAGTTGAAGAAGC
c.7976+1G>A	BIC	Intervening Sequence	**[-] GC donor site (HSF: 73.16 →43.33)**; [-] SRp40 (91.86); [-] Sironi ESS (69.76); [-] hnRNP A1(74.76)	FW: TCTTCAACTAAAATACAGACAAGTTTAAAGCATTACA 3 RV: TGTAATGCTTTAAACTTGTCTGTATTTTAGTTGAAGA
c.7976+5G>T	BIC	Intervening Sequence	**[-] GC donor site (MaxEnt: 3.1 →1.01)**; [-] 2 Fas-ESS [+] PESEs (32.82)	FW: CAACTAAAATACAGGCAATTTTAAAGCATTACATTAC RV: GTAATGTAATGCTTTAAAATTGCCTGTATTTTAGTTG

1 These variants belong to a positive microdeletion region with putative ESE sequences.
2 HSF: Human Splicing Finder. NNS: NNSplice.
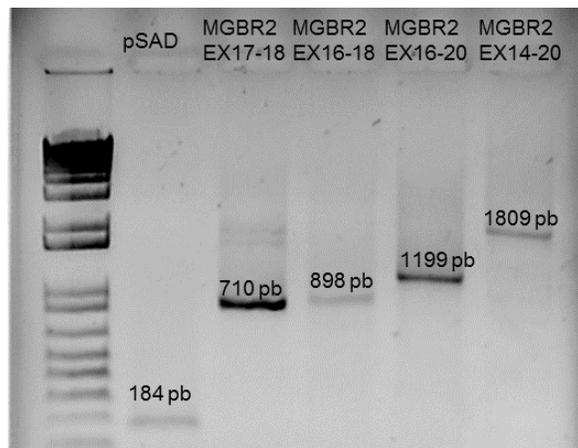
Figure 7. Agarose gel electrophoresis (1%) of RT-PCR results of the wt MGBR2EX14-20 minigene and its intermediate constructs, with the predicted fragment size.

The Initial MGBR2 17-18 minigene showed an intron retention, but splicing reactions were stabilized by inserting the exon 16. Once the final minigene was obtained and it produced a correct splicing pattern, no additional material is needed (such as mRNA from patients), which avoids sample artifacts. Furthermore, the minigene usage leads to the analysis of an isolated allele (which easily allows the quantification of the different mRNA isoforms generated by a DNA variant).

3.3 FUNCTIONAL ANALYSIS OF DNA SELECTED VARIANTS

The selected variants were tested with a functional assay of mutant and wt minigenes. As a prior analysis, the Figure 8 contains the cDNA amplification results on an agarose gel, of each variant compared with a wild type sequence. This figure lets the identification of the main transcript produced by each variant, but its resolution is not enough to determine the details of every splicing pattern. In broad terms, the donor-associated variants had an altered splicing outcome, congruent with exon 17 skipping.
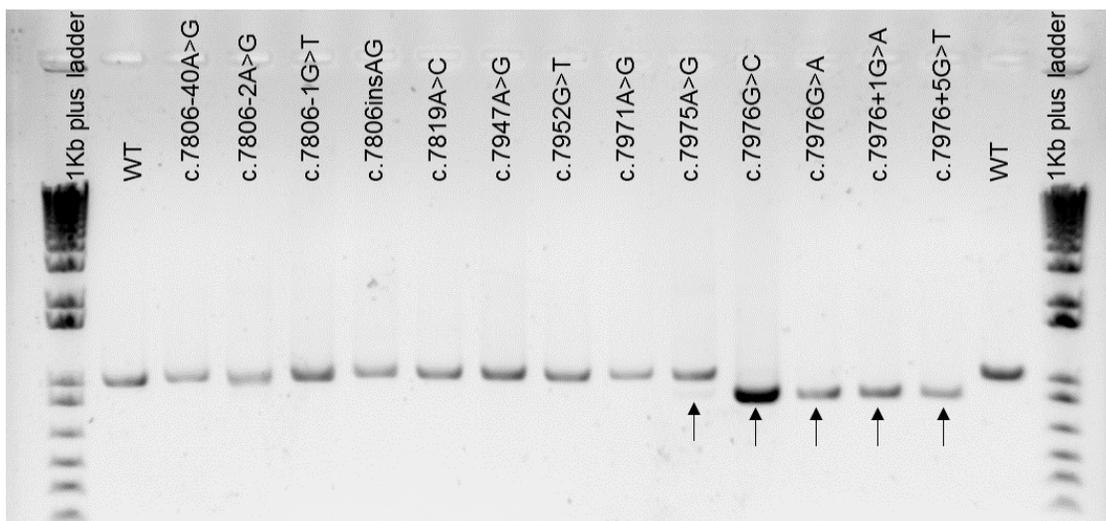
Figure 8. Agarose gel electrophoresis (1%) of RT-PCR results of the wild type MGBR2EX14-20, and the selected DNA variants. Arrows indicate abnormal patterns/transcripts.

All transcripts were characterized by capillary electrophoresis of the FAM-labeled RT-PCR products and by DNA sequencing, results are summarized in Table 3. Besides the splicing outcome, the RNA and protein effect according to the HGVS nomenclature is also reported. Eight out of thirteen mutations (61.5%) had a consequence in the splicing process of exon 17 of *BRCA2*. On next sections, each variant result is analyzed in detail.

Table 3. Functional assay results of selected variants, with the RNA and protein effect.*

DNA variant	Splicing outcome	RNA effect	Protein Effect
c.7806-40A>G	Canonical transcript	-	-
c.7806-2A>G	Alternative site usage 19 nt downstream (NNS: 0.17) (54.04%) Ex17 del20	r.7806_7825del	p. Ala2603Cysfs*8
	Alternative site usage 69 nt downstream (NNS: 0.54) (38.91%) Ex17 del69	r.7806_7874del	p.Ala2603_Arg2625del
	Exon 17 skipping (7.05%)	r.7806_7976del	p.Ala2603_Arg2659del
c.7806-1G>T	Alternative site usage 19 nt downstream (NNS: 0.17) (96.85%) Ex17 del20	r.7806_7825del	p. Ala2603Cysfs*8
	Alternative site usage 69 nt downstream (NNS: 0.54) (3.15%) Ex17 del69	r.7806_7874del	p.Ala2603_Arg2625del
c.7806insAG	Alternative site usage 69 nt downstream (NNS: 0.54) (2.82%) Ex17 del69	r.7806_7874del	p.Ala2603_Arg2625del
	Exon 17 skipping (1.42%)	r.7806_7976del	p.Ala2603_Arg2659del
	Exon17-insAG (95.76%)	r.7805_7806insAG	p. Ala2603Glyfs*46
c.7819A>C	Canonical transcript	-	p. Thr2607Pro
c.7947A>G	Canonical transcript	-	-
c.7952G>T	Canonical transcript	-	p. Arg2651Met
c.7971A>G	Canonical transcript	-	-
c.7975A>G	Exon 17 skipping (8.31%) Canonical transcript (91.69%)	r.7806_7976del	p.Ala2603_Arg2659del p. Arg2659Gly
c.7976G>C	Exon 17 skipping (96.47%) Alternative site usage 150 nt upstream (NNS: 0.94) (3.53%) Ex17 del150	r.7806_7874del r.7826_7976del	p.Ala2603_Arg2659del p.Val2610_Arg2659del
c.7976G>A	Exon 17 skipping (97.51%) Canonical transcript (2.49%)	r.7806_7976del	p.Ala2603_Arg2659del p. Arg2659Lys
c.7976+1G>A	Exon 17 skipping (100%)	r.7806_7976del	p.Ala2603_Arg2659del
c.7976+5G>T	Exon 17 skipping (100%)	r.7806_7976del	p.Ala2603_Arg2659del

* The variant c.7829dup could not be tested, since the site-directed mutagenesis was unsuccessful, despite the great effort on the standardization (annealing temperature, primer concentration and DMSO were tested, trying to obtain positive colonies). When we analyzed the mixed oligos duplexes formation, the variant c.7829dup had a very stable duplex at high temperature, which could lead to the experiment failure.

CANONICAL-TRANSCRIPT ASSOCIATED VARIANTS

Five variants showed no alteration of the canonical splicing pattern: c.7806-40A>G, c.7819A>C, c.7947A>G, c.7952G>T and c.7971A>G. Bioinformatic analysis indicated that c.7806-40A>G caused the disruption of a branch site; moreover, the other variants, which belong to a positive microdeletion region as is shown in Figure 9 (A), have been predicted to disrupt multiple enhancer sequences (such as SF2/ASF, SC35, SRp40 and SRp55) and to create silencer-binding sites. In addition, Figure 9 (B) shows the variant's capillary electrophoresis, compared to a wild type transcript; and Figure 9 (C) shows the representation of the canonical transcript splicing pattern. Two variants induced predicted missense mutations: c.7819A>C (p. Thr2607Pro) and c.7952G>T (p. Arg2651Met).
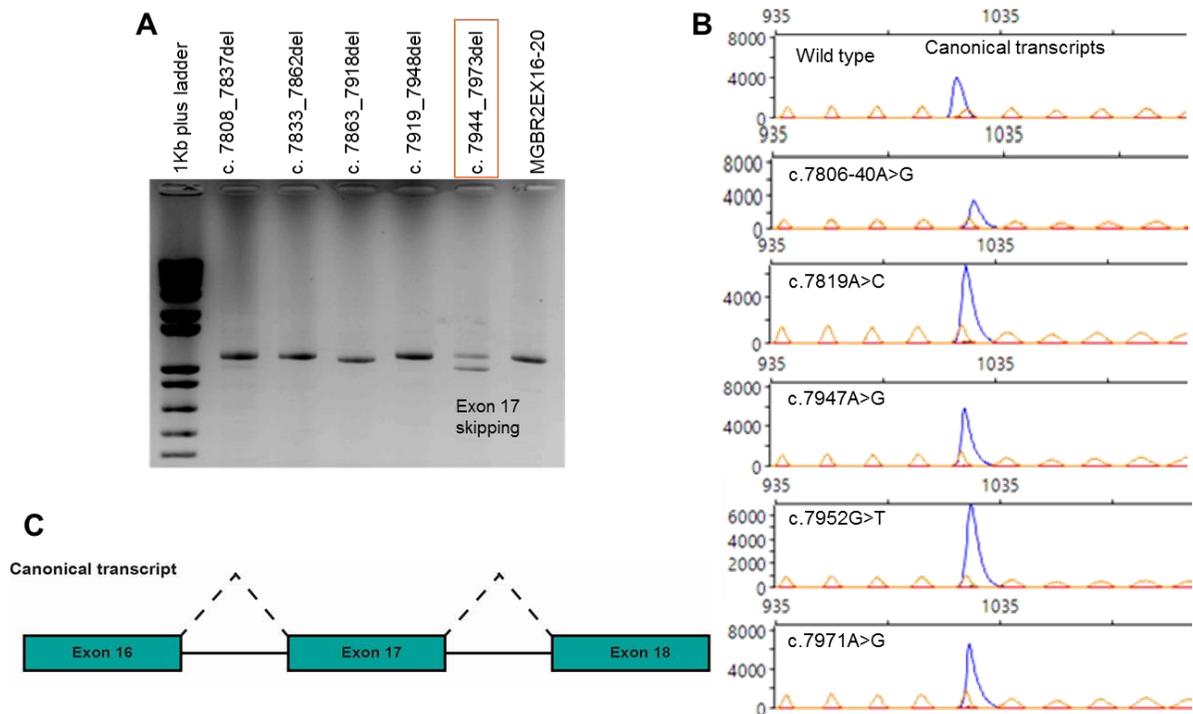
Figure 9. (A) Agarose gel electrophoresis (1%) of functional assays of *BRCA2* exon 17 microdeletions (adapted from **Acedo, A.** (2013) doctoral thesis). **(B)** Capillary electrophoresis of FAM-RT-PCR products (blue peaks) of the wild type MGBR2EX14-20, and the selected DNA variants. The Liz-1200 DNA size standard is shown as orange peaks. **(C)** Representation of the canonical splicing pattern.

ACCEPTOR SITE ASSOCIATED VARIANTS

The variants c.7806-2A>G, c.7806-1G>T and c.7806insAG, which affect the consensus 3' splice site sequence, induced aberrant splicing patterns. The bioinformatic analysis evidenced that the two first mutations caused the disruption of the natural acceptor site, but also the elimination of different enhancers and the silencer creation (c.7806-2A>G and c.7806-1G>T disrupt a 9G8 and a SC35 site, respectively). On the other hand, c.7806insAG was predicted to disrupt an acceptor site by the MaxEnt algorithm, and to create an hnRNPA1 binding site. The databases classification of this variant differs from our results, since c.7806insAG is classified only as a frame shift mutation, when it actually represents an intronic duplication of the AG acceptor site (c.7806-2_7806-1dup).

Figure 10 (A and B) shows the functional results, and 10 (C) schematizes the aberrant splicing patterns produced by these variants. c.7806-2A>G and c.7806-1G>T variants induced usage of two cryptic alternative acceptor sites, localized 19 and 69 nt downstream (exon 17-del19 and exon 17-del69), the first one with a higher proportion in both cases (54.04% and 96.85%). Moreover, c.7806-2A>G also induced exon 17 skipping (7.05%). Finally, the c.7806insAG mutation generated an apparently canonical transcript with a frame shift mutation (AG insertion; 95.76% of isoforms), as well as exon 17-del69 (2.82%) and exon 17 skipping (1.42%). The protein prediction effects of these variants include frame shift changes with premature termination codons (PTC), p. Ala2603Glyfs*46 (exon 17-insAG) and p.Ala2603Cysfs*8 (exon 17-del19), and two in-frame deletions, p.Ala2603_Arg2625del (exon 17-del69) and p.Ala2603_Arg2659del (exon 17 skipping).
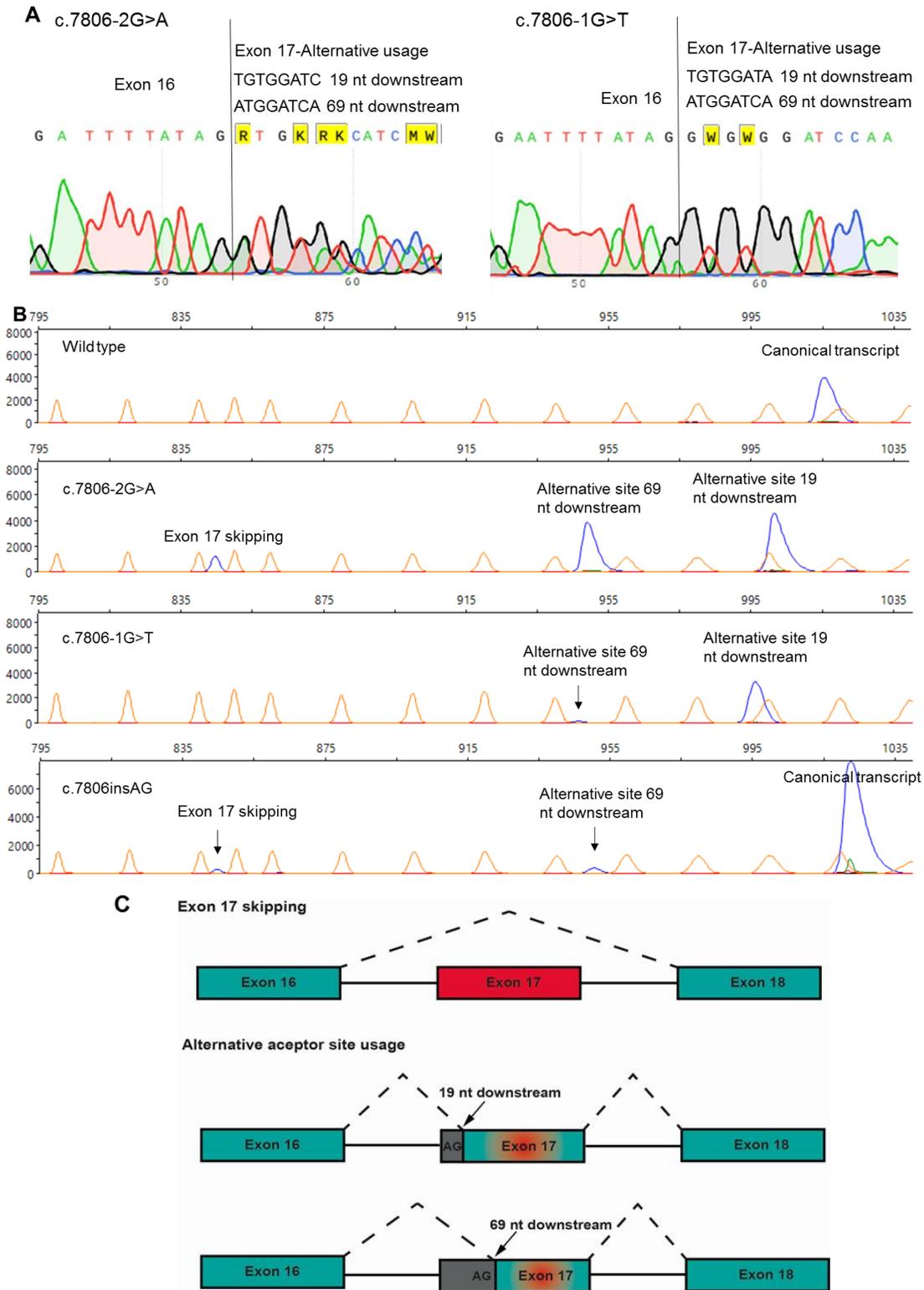
Figure 10. (A) Sequence and **(B)** capillary electrophoresis of RT-PCR products of c.7806-2A>G, c.7806-1G>T and c.7806insAG variants, compared to a wild type splicing outcome. The Liz-1200 DNA size standard is shown as orange peaks. **(C)** Representation of the aberrant-associated splicing patterns.

DONOR SITE ASSOCIATED VARIANTS

Variants c.7975A>G, c.7976G>C, c.7976G>A, c.7976+1G>A, and c.7976+5G>T, which can be classified as part of the consensus 5' splice site sequence, induced aberrant splicing patterns. The functional results and splicing pattern schemas are shown in Figure 11. The variant bioinformatic analysis, made through the NNSplice database, did not show the site disruption, since the exon 17 has a non-canonical donor sequence. However, other algorithms as MaxEnt from Human Splicing Finder detected changes in the atypical GC exon 17-donor site. Those changes were estimated to vary the scores as follows: -220%, -223%, -242%, -36%, and -67% respectively; therefore, a mutation within these sites should prevent recognition by the spliceosome.

According to databases, the c.7975A>G mutation eliminates a SRp40 binding site. The relative isoforms quantification detected by capillary electrophoresis revealed the existence of two transcripts, a canonical one (which induces a predicted missense p.Arg2659Gly) with a percentage of 91.69%, and exon 17 skipping with 8.31%. The BIC database had annotated this variant as missense, nonetheless, the functional assay results showed how the exon skipping causes a BRCA2 in-frame deletion of 57 amino acids (p.Ala2603_Arg2625del) that correspond to the DNA binding region of BRCA2 (helical domain: amino acids 2479-2667).

The following variants, c.7976G>C and c.7976G>A (annotated at BIC as missense), are substitutions of the last nucleotide of exon 17, and cause exon skipping in a major proportion. c.7976G>C causes exon 17 skipping in 96.47% of detected isoforms, and an alternative donor site recognition localized 150 nt upstream of the natural site (with a percentage of 3.53%). Similarly, c.7976G>A variant generates exon 17 skipping in high percentage (97.51%), but differs of the last one since it produces a canonical transcript in 2.49% of isoforms. Furthermore, the protein effect of such changes are in-frame deletions of 57 amino acids (which corresponds to exon 17 skipping, see above), and p.Val2610_Arg2659del, a deletion of 50 amino acids caused when an alternative donor 150 nt upstream is recognized.

The c.7976+1G>A mutation also induces exon 17 skipping, but it does it in 100% of transcripts. Bioinformatics also predicted a disruption of a SRp40 binding site. The BIC database classified it as an intervening sequence; however, it affects an essential nucleotide of the donor site causing a deletion of 57 amino acids as is shown in Table 3 (p.Ala2603_Arg2625del).

Finally, the c.7976+5G>T variant, induces exon 17 skipping in all the detected isoforms of the splicing outcome. This variant is considered and Unclassified Variant by the UMD database but here we demonstrated that c.7976+5G>T triggers the RNA isoform r.7806_7976del, and therefore, the predicted p.Ala2603_Arg2625del protein deletion, which belongs to the conserved DNA-binding domain, which could be probably associated with the disease. Interestingly, the splicing effect of this variant was replicated in a patient by lymphocyte RT-PCR (Infante and Durán, 2015. Cancer Genetics Laboratory, IBGM).
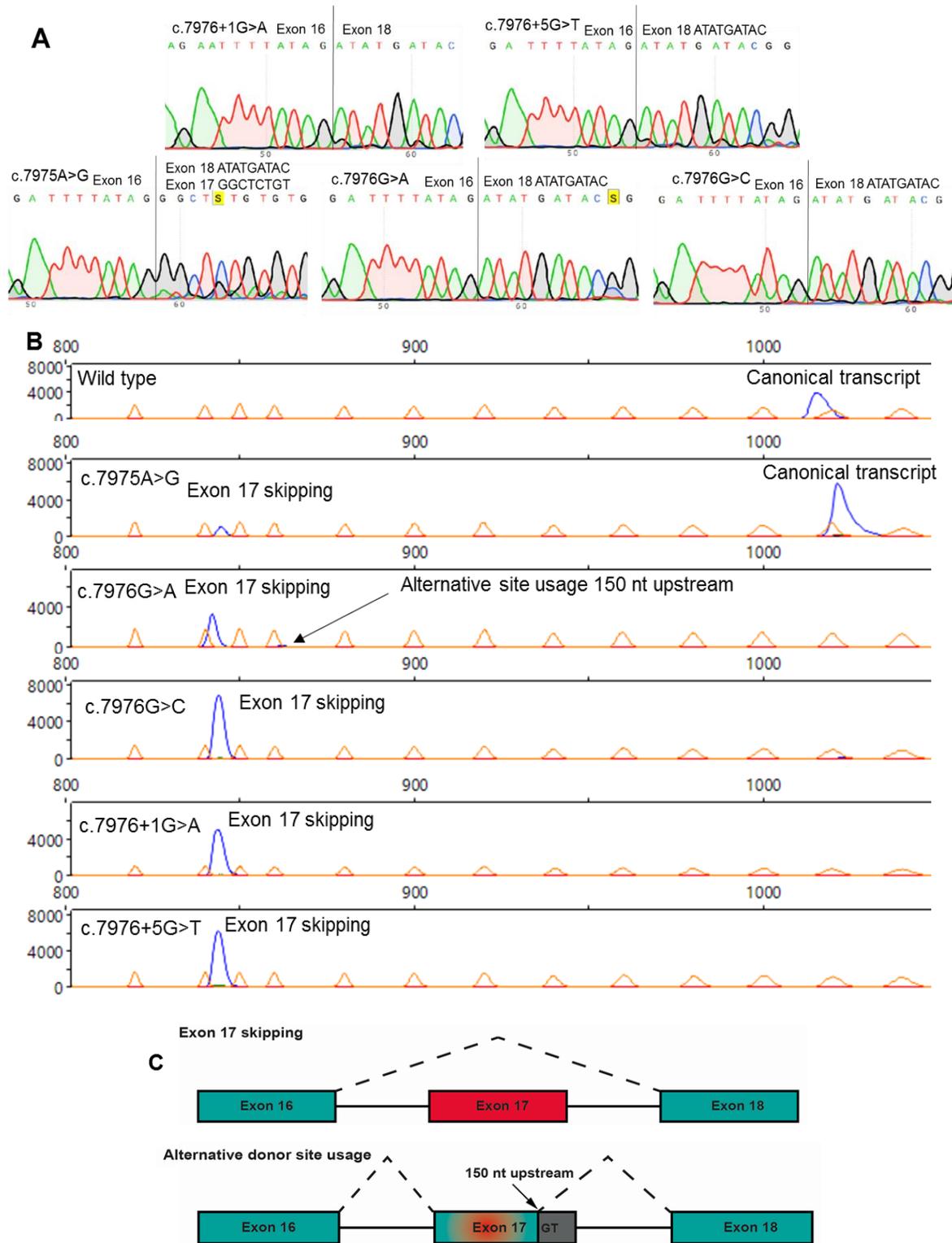
Figure 11. (A) Sequence and **(B)** capillary electrophoresis of variants c.7975A>G, c.7976G>C, c.7976G>A, c.7976+1G>A, and c.7976+5G>T, compared to a wild type splicing outcome. The Liz-1200 DNA size standard is shown as orange peaks. **(C)** Representation of the aberrant-associated splicing patterns.

20

4   DISCUSSION

The splicing is an important eukaryote cellular process, which plays a crucial role in the generation of biological complexity and genetic diversity. Pre-mRNA processing is an essential step in gene expression, regulated by a complex network of interactions; therefore, a disruption of any regulatory factor can have a determinative role in genetic diseases, as a direct cause, a modifier of disease severity, or a determinant of disease susceptibility (Faustino et al., 2003; Kaida et al., 2012; Wang and Cooper, 2007). Splicing aberrant patterns have also been associated with cancer initiation and progression, and are acquiring a growing relevance, being important in genetic testing, especially in hereditary cancer. In this way, the clinical classification of DNA variants regarding its effect on disease predisposition (which leads clinical management of patients and asymptomatic carriers) represents a challenge for geneticists; nonetheless, the identification of genomic variants that trigger aberrant splicing is not as simple.

4.1   HYBRID MINIGENE APPROACH VALIDATION

Although the study of RNA samples from patients is the ideal way to assess the splicing effect of DNA variants, this method has limitations that restrict a broad implementation. First, it is frequently difficult to obtain, since the patient sample might not be available and, second, the source of RNA could mask the actual splicing alterations because of the differential tissue-specific alternative splicing events (leukocytes are used rather than HBOC target tissue) (Baralle et al., 2009). As a reliable and straightforward substitute methodology, *ex vivo* assays of DNA variants have emerged. Splicing reporter minigenes are valuable tools to corroborate bioinformatic data, which had been reported successful for the identification of *BRCA1/2* splice-site variants with impact on splicing (Acedo et al., 2012; Acedo et al., 2015; Anczuków et al., 2012; Cooper, 2005; Gaildrat et al., 2010; Gaildrat et al., 2012; Sanz et al., 2010). We attempted to use a splicing reporter strategy by implementing a hybrid minigene to study the candidate DNA variants on exon 17 of *BRCA2*. Functional minigene analysis represents a powerful tool to determine the biological and the pathological significance of sequence variants detected in genetic screenings of disease-predisposing genes (Gaildrat et al., 2010). The designed vector, pSAD, allows the cloning of any exon of any human disease gene, so that any real or theoretical mutation can be generated by site-directed mutagenesis, and tested under the splicing perspective in eukaryotic cells (where unique vector exon primers specifically amplify the minigene-synthesized RNA). Moreover, the methodological strategy is straightforward, consisting of the following steps: Cloning into pSAD- Transfection- RNA isolation- RT-PCR – Sequencing and quantification of induced transcripts by fluorescent capillary electrophoresis.

Likewise, the high resolution of the fluorescent RT-PCR technique in splicing reporter assays should be highlighted as it allowed accurate detection of isoforms that otherwise could be masked. Its high resolution is capable of detecting minor transcripts; even less than 2% of total mRNA isoforms, which could not be visualized in agarose gels (see Figures 8, 10 and 11). Furthermore, the minigene approach is a single-allele assay that leads to a specific quantification of the different RNA isoforms without the normal allele interference (as in lymphocyte RT-PCR). Nevertheless, the success of this approach depends on the consideration of its main disadvantage: it may differ from the patient RNA analysis, since the genomic context (it has fewer inserted exons than the actual gene), and the control of the expression varies (it uses a different promoter). This could lead to changes in the expression level and transcription rate, influencing splicing process). In order to overcome the lack of

genomic environment, it is important to assure the correct exon recognition, by adding enough natural flanking exons to the analyzed one.

As an example, a previous study showed the effect of genomic context on the splicing outcome generated by two different *NF1* minigenes (Neurofibromatosis type I). The first one contained only the exon 37 of the *NF1* gene, and the pathogenic variant 6792C>G was analyzed, which was found to principally induce exon 37 skipping. Conversely, a minigene with exons 31 to 38, with the same variant introduction, replicated almost exactly the splicing pattern of a patient sample (canonical transcript, exon 37 skipping and exons 36+37 skipping) (Buratti et al., 2006). Moreover, our laboratory has already tested the influence of the genomic context on the functional analysis of variants through hybrid minigenes: in (Acedo, 2013), the c.8488-1G>A mutation (corresponding to exon 19 of *BRCA2*), was analyzed by a patient lymphocyte RT-PCR, and by functional splicing assays in MGBR2EX19-20 and MGBR2EX19-27 minigenes. The minigene results showed the presence of the two aberrant isoforms detected in the RT-PCR of lymphocytes; however, MGBR2EX19-20 also revealed a third transcript in major proportion (deletion of exon 20), differing from the patient pattern. When MGBR2EX19-27 was used, it practically restored the lymphocyte-splicing pattern supporting the hypothesis of the flanking exons requirement for proper recognition.

Consequently, larger minigene constructions with more exons should be carried out in order to mimic the natural genomic background. In the case of the MGBR214-20, we compared the minigene and lymphocyte RT-PCR results (using the c.7976+5G>T variant, which was evaluated in the Cancer Genetics Laboratory, IBGM), and showed the reliability of the minigene, which offers the sufficient genomic context for exon 17 study. Furthermore, another *BRCA2* exon 17 variant (c.7976G>A) was studied in patient samples (Hofmann et al., 2003), being completely congruent with our results. To these findings must be added the previous exposed evidence, and the fact that the minigene contains proper flanking exons to exon 17, giving rise to the assertion that the genomic context of the MGBR2EX14-20 to this exon provides the required architecture for its successful recognition.

4.2   VARIANT ANALYSIS: BIOINFORMATIC PREDICTION AND FUNCTIONAL ASSAYS

In order to test the splicing effect of significant variants and increase the probability of identifying *BRCA2* variants associated with HBOC, there was made a bioinformatic analysis of the sixty-six database reported variants. Bioinformatic predictions of DNA variants are useful, but only complementary, tools to identify possible splicing mutations because they must necessarily be confirmed by a functional assay. The algorithms used in those computational platforms are not enough precise by their own, because they systematically recognize all the degenerate motifs that can act as putative regulators. Their sensitivity can be increased by filtering the output data with other parameters such the motif evolutionary conservation, its strength, and the proximity of the splice sites (Fairbrother et al., 2004; Pettigrew et al., 2005).The bioinformatic analysis of exon 17 of *BRCA2* showed several features about it and its associated variants that were contrasted with the functional assay.

First, the NNSPLICE software did not recognize the donor site of exon 17, which is attributed to the non-canonical nature of the 3' consensus sequence. Nevertheless, other algorithms as HFS and MaxEnt could recognize it, giving the criteria to select these variants. The weak donor site of exon 17 of *BRCA2* does not appear to be an obstacle to function as such, because the exon recognition process is combinatorial, in which a series of factors (site robustness, presence of splicing enhancers or enhancers, genomic context, secondary structure,

transcription kinetics, etc.) would facilitate accurate detection of donor and acceptor sites. Nevertheless, the imprecision of the computational analysis could lead to mask the variant effect determination; then, it is important to take this into account at the time of the variant selection for functional analysis, having a special carefulness on the ones that are located within the consensus sequence. Indeed, when these variants were functionally tested, they caused aberrant splicing patterns.

Second, the current database annotation of *BRCA1/2* variants has a classification, which is mostly based on the predicted protein translation, with several categories as frame shift, nonsense, missense, intervening sequence and synonymous. However, the functional analysis showed that except for the variants classified as synonymous, at least one mutation of each type has been found to affect the splicing process and therefore, the protein translation, which lead us to recommend that these variants be reclassified as *spliceogenic*. Reclassification of VUSs as deleterious under the splicing perspective will increase the number of HBOC families who may benefit from tailored prevention protocols. Consequently, aberrant splicing should be considered as a primary mechanism of pathogenicity, important in the study of VUSs, and complementary to the prediction of the protein effects (Sanz et al., 2010).

In addition to bioinformatic analysis, we used a special criterion to select DNA variants for functional assays: if they were located within a positive microdeletion region previously identified, they would be prioritized. None of the variants that belonged to this region affected splicing, generating canonical transcripts. One way to explain this is the fact that cis-acting factors work additively: probably at the time of testing the variants individually, there was not a sufficient lack of regulation binding sites to alter exon recognition (i.e. the enhancer disruption and silencer creation, as the bioinformatic analysis predicted), as is evidenced in the complete microdeletion functional analysis, or we can even consider that selected variants do not affect active regulatory elements. Though these variants did not affect splicing, two of them cause missense changes with unknown clinical significance on the BRCA2 protein sequence: c.7819A>C causes a Thr to Pro amino acid change on the protein position 2607, and c.7952G>T causes an Arg to Met change on the 2651 position (whose possible functional effects on BRCA2 are being discussed on next section).

Another variant that did not generate changes in the splicing pattern was c.7806-40A>G. Despite bioinformatics predicted that c.7806-40A>G disrupted a putative branch point, the functional assay showed a canonical transcript. The human branch point sequence is degenerate and its localization is highly variable, so its identification can be confusing and inaccurate; certainly, the human consensus BP has never been extensively explored *in vitro* (Gao et al., 2008). Therefore, there are two main explanations to the functional analysis results: first, the bioinformatic prediction was wrong; and second, after the BP sequence disruption, another sequence was used to generate the lariat structure.

On the other hand, the positive functional results were obtained with the variants c.7806-2A>G, c.7806-1G>T, c.7806insAG, c.7975A>G, c.7976G>C, c.7976G>A, c.7976+1G>A and c.7976+5G>T. They generated four aberrant isoforms (Figures 10 and 11) that were caused by different splicing events including exon 17 skipping and use of alternative cryptic acceptor or donor splice sites. Moreover, most variants (Table 2) produced two or more distinct RNA isoforms. Our results are consistent with the literature, which indicates that the most common events of alternative splicing are exon skipping and alternative splice site selection, whereas intron retention is the less frequent phenomenon in physiological alternative splicing since this event is usually associated with changes in the reading frame and the introduction of

premature stop codons (Ast, 2004; Kim et al., 2008). In addition, these variants were located within the acceptor and donor consensus sequences, a fact that must be contextualized on the evolutionary conservation of these sites. Actually, the first, last and penultimate nucleotides of exons as well as the intronic positions +1 to +5 and -3 to -1 are also highly conserved and should be considered as potential splicing-affecting targets (Whiley et al., 2011). Therefore, this evidence suggests that the disruption of the splice sites is the causative mechanism, although some of them were not predicted to affect them (as it was explained before, on the non-canonical donor site).

4.3   PUTATIVE PATHOGENICITY AND BRCA2 PREDICTED EFFECTS

Taking all the results together, 12.1% of all the variants (8/66) of *BRCA2* exon 17 are associated with splicing defects, being the second most frequent deleterious mechanism, after the truncating mutations, which represent the 21.2% (14/66). The splicing mutations generate different types of protein isoforms: two nonsense, three missense, and three protein deletions, combined according to the protein effects of the splicing outcome of each DNA variant. Regarding to the putative pathogenicity of aberrant splicing-associated variants, there are two basic criteria (Brandão et al., 2011): first, total or almost complete absence of the canonical transcript in the splicing outcome; and second, the presence of predicted aberrant effects on protein translation, such as introduction of frame shifts and premature stop codons or loss of essential domains of BRCA2.

Seven out of eight variants fulfil these criteria (c.7806-2A>G, c.7806-1G>T, c.7806insAG, c.7976G>C, c.7976G>A, c.7976+1G>A and c.7976+5G>T), causing both the lack of canonical transcripts and protein modifications. It is important to clarify that c.7806insAG induces a frame shift, leading to the introduction of a premature stop codon, thus supporting a double deleterious mechanism: splicing disruption and protein truncation. These variants cause changes circumscribed to a highly conserved amino acids region of BRCA2 protein (see Figure 12), the helical domain that has been described to be responsible of the interaction between the BRCA2 protein with DSS1 (deleted in split hand/split foot), and the FANCD2 binding domain. It has been suggested that the DSS1 proteins are required for the stability of BRCA2 in mammalian cells (Li et al., 2006) and also has been implicated with normal cell growth in eukaryotes (Hofmann et al., 2003; Shahid et al., 2014), indicating an important role of exon 17 in encoding a relevant domain for maintaining the proper function of the BRCA2 protein. Furthermore, the FANCD2 binding domain protein has a role in DNA repair, which is important to guarantee the BRCA2 protein functionality and then, the integrity of the genome during cell replication (De Oca et al., 2005; Van Der Groep et al., 2008).
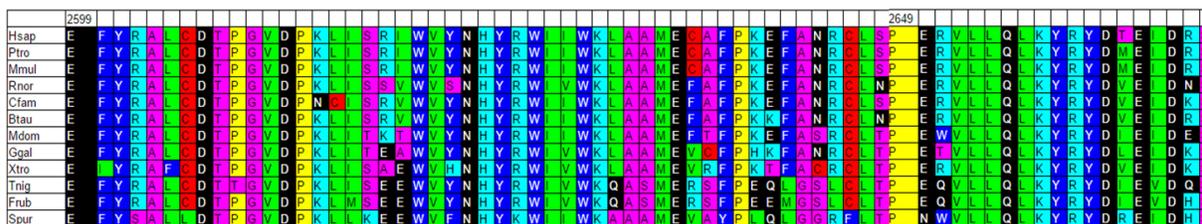
Figure 12. IARC BRCA2 alignment of the exon 17-corresponding amino acids (p.2603-2659). It contains the BRCA2 sequences from the following organisms: *Homo sapiens, Pan troglodytes, Macaca mulatta, Rattus norvegicus, Canis lupus familiaris, Bos taurus, Monodelphis domestica, Gallus gallus, Xenopus tropicalis, Tetraodon nigroviridis, Takifugu rubripes* and *Strongylocentrotus purpuratus.* Adapted from http://agvgd.iarc.fr/references.php

Additionally, two variants that generated canonical splicing patterns (c.7819A>C and c.7952G>T) induce missense mutations on the BRCA2 protein. The changes, p. Thr2607Pro p. Arg2651Met respectively, were identified within the IARC alignment (Figure 12) and then, tested in the PolyPhen-2 server (Adzhubei et al., 2010), in order to measure their possible effect on the protein functionality. The results indicate that p.Thr2607 is completely conserved within the BRCA2 alignment, and it was predicted as probably damaging on PolyPhem-2, with the maximum score of 1.0. On the other hand, p. Arg2651, is one of the most variable amino acid sites in the region (it has four different residues), but it was also classified as probably damaging, with a score of 0.983. This analysis shows how conserved the protein region is, and supports the idea of the deleteriousness of mutations within this domain.

Another variant altered splicing, and c.7975A>G, maintaining a major proportion of canonical transcripts. It had weak impact on splicing since aberrant pattern accounted for only 8.31% of total detected isoforms. Then, the remaining question is how to elucidate the implication in carcinogenesis of mutations with incomplete splicing effects? Variants with strong effects, as it was explained, are probably deleterious since they inactivate critical protein functions. Nevertheless, the variants whose splicing effect is partial, such as c.7975A>G, have unknown effects on protein function, and it pathological role is more uncertain so that they would remain as unclassified variants.

This question should be addressed by implementing an holistic study, including functional assays and epidemiological and statistical analyses (Easton et al., 2007). There must be taken into account the contribution of partial splicing mutations to a synergistic model of low-penetrance alleles in HBOC susceptibility. In fact, there has been estimated that more than 1,000 loci are involved in breast cancer susceptibility (Michailidou et al., 2013). All these parameters should be integrated in the definition of this cancer risk, and must include susceptibility and protector alleles, as well as environmental and life-style factors (Sanz et al., 2010). All this, in order to provide an assessment of the individual risk, which seems to be one of the most laborious undertakings.

5   CONCLUSIONS

- Functional analysis of *BRCA2* DNA variants was made through a hybrid minigene. It was constructed in the pSAD vector with exons 14 to 20, which allows the maintenance of the genomic context to exon 17. Therefore, it faithfully reproduces the physiological splicing reactions, avoiding the difficulty of the patient samples acquisition. This simple and reproducible approach analyzes the splicing impact of DNA variants by the introduction of them by site-directed mutagenesis, and uses the quantification of all splicing isoforms by the use of high-resolution capillary electrophoresis.

- Bioinformatic analysis of DNA variants comprises useful predicting tools that help to select candidates for functional analysis. This type of study increases the probability of finding relevant variants with an effect on the splicing process regulation. Nevertheless, the accuracy of these computational tools varies according to the type of the regulatory element disruption: the prediction of mutations associated with canonical splice sites and the creation of alternative ones are more reliable than the predictions of enhancers and silencers.

- Eight out of thirteen analyzed DNA mutations (61.5%) altered the pre-mRNA processing of the *BRCA2* exon 17. Seven of them can be likely classified as HBOC pathogenic variants (c.7806-2A>G, c.7806-G>T, c.7806insAG, c.7976G>C, c.7976G>A, c.7976+1G>A and c.7976+5G>T); the remaining one (c.7975A>G) has an incomplete effect on deleterious splicing, and the actual determination of its carcinogenic power requires further studies.

- The results indicate that an important fraction of DNA variants are associated with splicing aberrations, and should be reclassified on databases; therefore, splicing disruption must be considered as a primary cause of gene inactivation, important in the study of unclassified DNA variants, so that splicing functional assays provide an added value for the genetic counseling of hereditary diseases. Thus, the reclassification of *BRCA2* VUSs variants that affect splicing will significantly improve the genetic counseling procedure, increasing the number of HBOC families that can benefit from prevention protocols.

6   REFERENCES

Acedo, A. (2013). Alteraciones de splicing del gen BRCA2 en cáncer de mama y ovario. Diseño del nuevo vector de splicing pSAD para la construcción de minigenes híbridos.

Acedo, A., Sanz, D. J., Duran, M., Infante, M., Perez-Cabornero, L., Miner, C. and Velasco, E. a (2012). Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res.* **14**, R87.

Acedo, A., Hernández-Moro, C., Curiel-García, Á., Díez-Gómez, B. and Velasco, E. a. (2015). Functional Classification of   BRCA2 DNA Variants by Splicing Assays in a Large Minigene with 9 Exons. *Hum. Mutat.* **36**, 210–221.

Adzhubei, I. a, Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. and Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249.

Anczuków, O., Buisson, M., Léońe, M., Coutanson, C., Lasset, C., Calender, A., Sinilnikova, O. M. and Mazoyer, S. (2012). BRCA2 deep intronic mutation causing activation of a cryptic exon: Opening toward a new preventive therapeutic strategy. *Clin. Cancer Res.* **18**, 4903–4909.

Apostolou, P. and Fostira, F. (2013). Hereditary breast cancer: the era of new susceptibility genes. *Biomed Res Int* **2013**, 747318.

Ast, G. (2004). How did alternative splicing evolve? *Nat. Rev. Genet.* **5**, 773–782.

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

Berget, S. M. and Sharp, P. a (1977). A spliced sequence at the 5'-terminus of adenovirus late mRNA. *Brookhaven Symp. Biol.* **74**, 332–344.

Brandão, R. D., Van Roozendaal, K., Tserpelis, D., García, E. G. and Blok, M. J. (2011). Characterisation of unclassified variants in the BRCA1/2 genes with a putative effect on splicing. *Breast Cancer Res. Treat.* **129**, 971–982.

Buratti, E., Baralle, M. and Baralle, F. E. (2006). Defective splicing, disease and therapy: Searching for master checkpoints in exon definition. *Nucleic Acids Res.* **34**, 3494–3510.

Caputo, S., Benboudjema, L., Sinilnikova, O., Rouleau, E., Béroud, C., Lidereau, R., Sevenet, N., Bonnet, F., Longy, M., Hardouin, A., et al. (2012). Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-BRCA1/BRCA2 databases. *Nucleic Acids Res.* **40**, 992–1002.

Churbanov, A., Winters-Hilt, S., Koonin, E. V and Rogozin, I. B. (2008). Accumulation of GC donor splice signals in mammals. *Biol. Direct* **3**, 30.

Claverie, J. M. (2001). Gene number. What if there are only 30,000 human genes? *Science* **291**, 1255–1257.

Colapietro, P., Gervasini, C., Natacci, F., Rossi, L., Riva, P. and Larizza, L. (2003). NF1 exon 7 skipping and sequence alterations in exonic splice enhancers (ESEs) in a neurofibromatosis 1 patient. *Hum. Genet.* **113**, 551–554.

Cooper, T. a. (2005). Use of minigene systems to dissect alternative splicing elements. *Methods* **37**, 331–340.

De Klein, A., Riegman, P. H. J., Bijlsma, E. K., Heldoorn, A., Muijtjens, M., Den Bakker, M. a., Avezaat, C. J. J. and Zwarthoff, E. C. (1998). A G>A transition creates a branch point sequence and activation of a cryptic exon, resulting in the hereditary disorder neurofibromatosis 2. *Hum. Mol. Genet.* **7**, 393–398.

De Oca, R. M., Andreassen, P. R., Margossian, S. P., Gregory, R. C., Taniguchi, T., Wang, X., Houghtaling, S., Grompe, M. and D'Andrea, A. D. (2005). Regulated interaction of the Fanconi anemia protein, FANCD2, with chromatin. *Blood* **105**, 1003–1009.

Desmet, F. O., Hamroun, D., Lalande, M., Collod-Bëroud, G., Claustres, M. and Béroud, C. (2009). Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, 1–14.

Easton, D. F., Deffenbaugh, A. M., Pruss, D., Frye, C., Wenstrup, R. J., Allen-Brady, K., Tavtigian, S. V, Monteiro, A. N. a, Iversen, E. S., Couch, F. J., et al. (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.

Fairbrother, W. G., Holste, D., Burge, C. B. and Sharp, P. a. (2004). Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol.* **2**,.

Faustino, N. A., Cooper, T. a and Andre, N. (2003). Pre-mRNA splicing and human disease. *Genes Dev.* **17**, 419–437.

Gaildrat, P., Killian, A., Martins, A., Tournier, I., Frébourg, T. and Tosi, M. (2010). Use of Splicing Reporter Minigene Assay to Evaluate the Effect on Splicing of Unclassified Genetic Variants. *Methods Mol. Biol.* **653**, 249–257.

Gaildrat, P., Krieger, S., Di Giacomo, D., Abdat, J., Revillion, F., Caputo, S., Vaur, D., Jamard, E., Bohers, E., Ledemeney, D., et al. (2012). Multiple sequence variants of BRCA2 exon 7 alter splicing regulation. *J. Med. Genet.* **49**, 609–617.

Gao, K., Masuda, A., Matsuura, T. and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.* **36**, 2257–2267.

Gelinas, R. E. and Roberts, R. J. (1977). One predominant 5'-undecanucleotide in adenovirus 2 late messenger RNAs. *Cell* **11**, 533–544.

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., Emanuelsson, O., Zhang, Z. D., Weissman, S. and Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Res.* **17**, 669–681.

Han, S. P., Tang, Y. H. and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. *Biochem. J.* **430**, 379–392.

Hofmann, W., Horn, D., Hüttner, C., Classen, E. and Scherneck, S. (2003). The BRCA2 variant 8204G>A is a splicing mutation and results in an in frame deletion of the gene. *J. Med. Genet.* **40**, e23.

Infante, M. and Durán, M. (2015). Cancer Genetics Laboratory, IBGM. Personal communication.

Irimia, M. and Roy, S. W. (2014). Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb. Perspect. Biol.* **6**,.

Kaida, D., Schneider-Poetsch, T. and Yoshida, M. (2012). Splicing in oncogenesis and tumor suppression. *Cancer Sci.* **103**, 1611–1616.

Kim, E., Goren, A. and Ast, and G. (2008). Insights into the connection between cancer and alternative splicing. *Trends Genet.* **24**, 7–10.

Krämer, a (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* **65**, 367–409.

Krawczak, M., Thomas, N. S. T., Hundrieser, B., Mort, M., Wittig, M., Hampe, J. and Cooper, D. N. (2007). Single Base-Pair Substitutions in Exon–Intron Junctions of Human Genes: Nature, Distribution, and Consequences for mRNA Splicing. *Hum. Mutat.* **28**, 150–158.

Kurahashi, H., Takami, K. and Oue, T. (1995). Biallelic Inactivation of the APC Gene in Hepatoblastoma Binlie ! ic Inactivation of the APC Gene in Hepatoblastoma1. **5**, 5007–5011.

Li, J., Zou, C., Bai, Y., Wazer, D. E., Band, V. and Gao, Q. (2006). DSS1 is required for the stability of BRCA2. *Oncogene* **25**, 1186–1194.

Linger, R. J. and Kruk, P. a. (2010). BRCA1 16 years later: Risk-associated BRCA1 mutations and their functional implications. *FEBS J.* **277**, 3086–3096.

Louise T Chow, Richard E. Gelinas, Thomas R., B. and R. J. R. (1977). An amazing sequence arragement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8.

Lux, M. P., Fasching, P. a. and Beckmann, M. W. (2006). Hereditary breast and ovarian cancer: Review and future perspectives. *J. Mol. Med.* **84**, 16–28.

Lynch, M. (2006). The origins of eukaryotic gene structure. *Mol. Biol. Evol.* **23**, 450–468.

Mazoyer, S., Puget, N., L, P.-V., Lynch, H., Serova-Sinilnikova, O. and Lenoir, G. (1998). A BRCA1 Nonsense Mutation Causes Exon Skipping. *Am. J. Hum. Genet.* **62**, 713–715.

Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L., Schmidt, M. K., Chang-Claude, J., Bojesen, S. E., Bolla, M. K., et al. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–61, 361e1–2.

Moore, M. J., Query, C. C. and Sharp, P. a (1993). *Splicing of precursors to messenger RNAs by the spliceosome.*

Nowak, D. G., Woolard, J., Amin, E. M., Konopatskaya, O., Saleem, M. a, Churchill, A. J., Ladomery, M. R., Harper, S. J. and Bates, D. O. (2008). Expression of pro- and anti-angiogenic isoforms of VEGF is differentially regulated by splicing and growth factors. *J. Cell Sci.* **121**, 3487–3495.

Padgett, R. a. (2012). New connections between splicing and human disease. *Trends Genet.* **28**, 147–154.

Pajares, M. J., Ezponda, T., Catena, R., Calvo, A., Pio, R. and Montuenga, L. M. (2007). Alternative splicing: an emerging topic in molecular and clinical oncology. *Lancet Oncol.* **8**, 349–357.

Pettigrew, C., Wayte, N., Lovelock, P. K., Tavtigian, S. V, Chenevix-Trench, G., Spurdle, A. B. and Brown, M. a (2005). Evolutionary conservation analysis increases the colocalization of predicted exonic splicing enhancers in the BRCA1 gene with missense sequence changes and in-frame deletions, but not polymorphisms. *Breast Cancer Res.* **7**, R929–R939.

Reese, M. G., Eeckman, F. H., Kulp, D. and Haussler, D. (1997). Improved splice site detection in Genie. *J. Comput. Biol.* **4**, 311–323.

Rodríguez-Trelles, F., Tarrío, R. and Ayala, F. J. (2006). Origins and evolution of spliceosomal introns. *Annu. Rev. Genet.* **40**, 47–76.

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

Sakharkar, M. K., Chow, V. T. K. and Kangueane, P. (2004). Distributions of exons and introns in the human genome. *In Silico Biol.* **4**, 387–393.

Sanz, D. J., Acedo, A., Infante, M., Durán, M., Pérez-Cabornero, L., Esteban-Cardeñosa, E., Lastra, E., Pagani, F., Miner, C. and Velasco, E. a. (2010). A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin. Cancer Res.* **16**, 1957–1967.

Shahid, T., Soroka, J., Kong, E. H., Malivert, L., McIlwraith, M. J., Pape, T., West, S. C. and Zhang, X. (2014). Structure and mechanism of action of the BRCA2 breast cancer tumor suppressor. *Nat. Struct. Mol. Biol.*

Sieliwanowicz, B. (1988). Splicing of nuclear messenger RNA precursors. *Postepy Biochem.* **34**, 351–360.

Srebrow, A. and Kornblihtt, A. R. (2006). The connection between splicing and cancer. *J. Cell Sci.* **119**, 2635–2641.

Sterne-weiler, T. and Sanford, J. R. (2014). Exon identity crisis : disease-causing mutations that disrupt the splicing code. 1–8.

Stickeler, E., Kittrell, F., Medina, D. and Berget, S. M. (1999). Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. *Oncogene* **18**, 3574–3582.

Stratton, M. R. and Rahman, N. (2008). The emerging landscape of breast cancer susceptibility. *Nat. Genet.* **40**, 17–22.

Taylor, M. D., Gokgoz, N., Andrulis, I. L., Mainprize, T. G., Drake, J. M. and Rutka, J. T. (2000). Familial posterior fossa brain tumors of infancy secondary to germline mutation of the hSNF5 gene. *Am. J. Hum. Genet.* **66**, 1403–1406.

Thanaraj, T. a and Clark, F. (2001). Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Res.* **29**, 2581–2593.

Van Der Groep, P., Hoelzel, M., Buerger, H., Joenje, H., De Winter, J. P. and Van Diest, P. J. (2008). Loss of expression of FANCD2 protein in sporadic and hereditary breast cancer. *Breast Cancer Res. Treat.* **107**, 41–47.

Wang, G.-S. and Cooper, T. A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* **8**, 749–761.

Wang, M., Dotzlaw, H., Fuqua, S. a and Murphy, L. C. (1997). A point mutation in the human estrogen receptor gene is associated with the expression of an abnormal estrogen receptor mRNA containing a 69 novel nucleotide insertion. *Breast Cancer Res. Treat.* **44**, 145–151.

Wang, J., Zhang, J., Li, K., Zhao, W. and Cui, Q. (2012). SpliceDisease database: Linking RNA splicing and disease. *Nucleic Acids Res.* **40**, 1055–1059.

Wang, Y., Liu, J., Huang, B., Xu, Y., Li, J., Huang, L., Lin, J., Zhang, J., Min, Q., Yang, W., et al. (2014). Mechanism of alternative splicing and its regulation (Review). *Biomed. Reports* 152–158.

Whiley, P. J., Guidugli, L., Walker, L. C., Healey, S., Thompson, B. a., Lakhani, S. R., Da Silva, L. M., Investigators, K., Tavtigian, S. V., Goldgar, D. E., 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.

Will, C. L. and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, 1–2.