UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Human Development and Health

Splicing regulation of *BRCA1* exon 11

by

Claudia Tammaro

Thesis for the degree of Doctor of Philosophy

October 2013

UNIVERSITY OF SOUTHAMPTON

Abstract

FACULTY OF MEDICINE

Human Development and Health

Doctor of Philosophy

Splicing regulation of *BRCA1* exon 11

by Claudia Tammaro

BRCA1 is one of the major genes associated with familial breast cancer and mutations in this gene are responsible for about 40-45% of hereditary breast cancer. These include DNA changes that cause *BRCA1* aberrant splicing by affecting regulatory sequences.

Alternative splicing is a post-transcriptional process allowing the generation of multiple protein isoforms from the same gene. *BRCA1* alternative splicing is altered in breast cancer therefore mutations in splicing regulatory elements along this gene sequence can contribute to the disease. Three *BRCA1* isoforms are present in different proportions in breast cancer due to alternative splicing of exon 11. Mutations in exon 11 that affect splicing regulatory elements can alter the normal isoform ratio and cause breast cancer.

An Unclassified variant (UV) c.693G>A was found in *BRCA1* in a breast cancer patient with a family history of breast and ovarian cancer. This unclassified variant is situated in exon 11 in *BRCA1* gene.

Using a mini-gene approach as a splicing assay and when possible, blood samples from the patients, I show that the unclassified variant c.693 G>A has a strong effect on the splicing isoform ratios of *BRCA1*.

Systematic site directed mutagenesis of the area surrounding the nucleotide position c.693 in exon 11 and mini-gene splicing assay suggests a splicing

regulatory element exists in this region that is disrupted with the c.693 G>A sequence variant.

Bioinformatic analysis and in vitro (Pull-Down) analysis have detected both enhancer and silencer factors as the regulatory proteins that bind this area regulating BRCA1 alternative splicing.

Correction strategies to revert aberrant splicing of exon 11 in the *BRCA1* gene in the presence of the synonymous variant c.693 G>A using bifunctional oligonucleotide were undertaken. However these failed to alter the inclusion of exon 11.

List of Contents.

List of Figure.

List of Table.

Declaration of Authorship.

I *Claudia Tammaro* declare that this and the work presented in it are my own and has been generated by me as the result of my own original research.

"Splicing regulation of *BRCA1* exon 11"

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- Part of this work has been published before submission, or parts of this work have been published as:

Raponi M., Douglas AG.,Tammaro C., Wilson DI, Baralle D. (2012) "Evolutionary Constraint Helps Unmask a Splicing Regulatory Region in BRCA1 Exon 11" PlosOne;7(5):e37255. Epub 2012 May 16.

Tammaro C., Raponi M, Wilson DI, Baralle D. “BRCA1 Exon 11 alternative splicing, multiple functions and the association with cancer”. 2012 Biochemical society Transactions 40, (768-772).

Signature: ……Date:

Acknowledgements.

Finally the end!!!
When I start the PhD I could never have imagined what challenge it would have been. I could never image the passion that you need to bring with you every day, how many times you need to repeats to yourself “I can do it and I will do it!!!”I could never image how much you need to be in love with science to finish the PhD.
However, it would not have been possible reach this point without all the people I had on my side, without their support and helpful criticism which involves laughs and sometimes tears.

I would like to thanks my supervisors: **Dr Diana Baralle**, **Dr Michela Raponi** and **Prof David Wilson** for their constant guide and their precious suggestions.
Thanks to **Diana**, to give me the opportunity to realise my dream, for all her help and the trust she had in me and in my project throughout my PhD.
Thanks to **David** for his constant guidance and precious advice, for so many conversations, coffees laughs and for sharing through and honesty, for his precious time, and thanks for being always on my side supporting me. Most of all, thanks for your friendship. I will always remember it.
And finally I would like to thanks **Michela**, who is not only my supervisor, she is also a great colleagues and a fantastic friend. Thanks for give me the opportunity to feel science, for teaching me, for sharing thoughts, for giving me your time, for helping and believe in me during my PhD. Without you I would not be here. It feels like an honour to share science with you. But above all, thanks for your friendship; you are part of my family. Thanks!!!!

Special thanks to my group (the splicing girls). Thanks to Lindsay, it was pleasures for me share the splicing knowledge with you. I really enjoy all the bifunctional oligonucleotide discussion; you are a great colleague and a good friend!!! Thanks to my friend Michela, we share thought, laugh and sometimes

tears. I will always remember all the moment together. And thanks to all people that in this year's decide to feel in love with splicing, we always had a great time in and outside the lab.
Thanks to everyone in the surrounding laboratory and the offices, for making my workplace a pleasant and challenging environment, sharing together science and life.

I would like to thanks all my friends in Southampton to be so amazing and making my life fantastic!!!Thanks to be my family, for helping me during my time in Soton, for sharing with me dinner, (some of them in the middle of night☺), laughing, amazing chat and of course some tears. Thank you to those of you whom gave me the opportunity to see and organise concerts☺. Thank you to those of you whom have always been on my side, in the shade, give me suggestion and advice, to those of you that have shared with me all of the difficulties of these years, to those of you that decide to be my friend. In a simple way you are the BEST friends and you will always be part of my life. I always said that the true fellowship follows us wherever we are and I am sure that we will be always in touch.
I would like to thanks to my closest friends in Italy, for making the distance being so short. Thank you to all of them for still being my friends, and for understanding me and my life, to those of you that share with me my passion!

Thanks to my family and in particular my parents (Umberto and Immacolata), for understanding me and being always there. For so many talks that made distance be nothing. They have supported me during my PhD and without whom it would have not been possible, and for that reason I will dedicate my thesis to them.
To all my surrounding family: my brothers, my sisters and my brothers in law, my parents in law and my grandparents. Thank you to all of you for encouraging, supporting and be always on my side, during the challenging times of this PhD. Thanks for loving me and understand me during this years. A special thanks to my

nephew (Placido) and my niece (Monica), thanks for loving me even if I was not always present, and thanks for making my worst day a fantastic day.

Last but not at least I would like to express a sincere thanks my fiancé Raffaele Cavallaro.
Thanks Lello for encouraging and supporting me during this PhD. Thanks for be always on my side, for holding my hand, for making me smile and for guide me during all this time.
Thank you for your incredible patience, for teaching me to trust in my ability and give me calm, and optimism, that have absolutely made me who I am today. Without you I would never have been able to reach this point!!! Thanks for loving me!!!

Definitions and abbreviations

The standard abbreviations used in this thesis follow IUPAC rules. All the abbreviations are also defined in the text when they are introduced for the first time.

A	Adenosine
aa	Amino acid
APS	Ammonium persulfate
ATP	Adenosine triphosphate
AS	Alternative splicing
bp	Base pair
BSA	Bovine serum albumin
BP	Branch point
C	Cytosine
cDNA	Copy DNA
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleoside triphosphate (A, C, G and T)
DMEM:	Dulbecco`s Modified Eagle Medium;
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
E.Coli	*Escherichia Coli*
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
FBS:	fetal bovine serum;
G	Guanosine
HRP	Horse radish peroxidise
hnRNP	Heterogeneous ribonucleoprotein
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer

kb	Kilo base
kDa	Kilo Dalton
LB	Luria Bertani broth
µ	Micro
M	Molar
MCF7	Breast cancer cell line
Min	Minute
mRNA	Messenger RNA
N	Nucleotide (A or C or G or T)
NMD	Nonsense-mediated decay
nt	Nucleotides
o/n	Overnight
OD_{600}	Light absorbance at 600 nm wavelength
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PPT	Polypyrimidine tract
Pu	Purine
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS gel electrophoresis
Syn	synonymous
snRNA	Small Nuclear Ribonucleoprotein
snRNP	Small nuclear ribonucleoprotein particles
SOC	Super optimal culture solution
SR	Arginine-serine rich protein
SRSF 1	serine arginine splicing factor 1 (ASF/SF2)
SRSF 5	serine arginine splicing factor 5 (SRp40)
SRSF 6	serine arginine splicing factor 6 (SRp55)

SRSF 7	serine arginine splicing factor 7 (9G8)
SRSF 9	serine arginine splicing factor 9 (SRp30c)
SFRS 10	serine arginine splicing factor 10 (Tra2α/β)
ss	Splice site
SR	Arginine-serine rich protein
T	Thymine
TA	Annealing temperature
TBE	Tris-borate-EDTA (buffer)
TBST	Tris-buffered-saline-tween
TEMED	Tetramethylethylenediamine
U	Uracil
wt:	wild type;
λ:	wavelength

Oligonucleotides:

2’-O-Me-2’-O-Methyl

2’-O-MePs-2’-O-Methyl Phosphorothioate

Chapter 1

Introduction.

1 Introduction.

The central dogma of molecular biology according to Crick and Watson states that genetic information is sequential. Information is transferred from DNA to RNA to protein, but cannot be transferred from protein to nucleic acid (Crick, 1970). The transfer of information includes transcription (DNA to RNA) and translation (RNA to protein). From transcription to translation there is a maturation process of the RNA which includes poly-adenylation and RNA capping.

In eukaryotes most transcripts undergo an additional maturation step (splicing) where non-coding sequences (introns) are removed from the pre-mRNA and exons (coding sequence) are joined together. One gene can produce different mRNA and different protein products thanks to an alternative splicing (AS) process in which different exons can be included or excluded to produce diverse mature RNA primary transcripts.
Mutations in the DNA sequence can affect processes involved in the creation of the correct proteins, leading to the creation of non-functional proteins or aberrant proteins that could cause disease. These aberrant DNA sequence changes include modifications both in the coding region of the gene and in the non-coding region. Nonsense and missense mutations in the coding regions are commonly assumed to cause their effect by altering protein coding. However these can also affect genes by inducing aberrant splicing. A nonsense mutation is a point mutation that produces a truncated protein product by creating a premature stop codon; whereas missense mutations create a single nucleotide change which changes the amino acid.

Conversely, silent sequence variants are thought not to result in a change of the amino acid sequence of a protein and are usually considered to be neutral. These assumptions could be correct in some cases, but when we consider the RNA level these assumptions could be misleading. This is because sequence variants can also

affect RNA sequences important for the splicing process and therefore potentially affect the final mature RNA transcript and the final protein product.

1.1 Pre-mRNA splicing reaction.

In eukaryotes, the coding sequence of the genes, the exons, are interrupted by non-coding sequences, termed introns. After transcription, introns are removed from the precursor-messenger RNA (pre-mRNA) by a process called splicing and the exons are joined together to form the mature mRNA. The splicing reaction is an essential step in the post transcriptional regulation of gene expression in eukaryotes. This process takes place in a complex machine called spliceosome, a complex molecular machine containing U1, U2, U5 and U4/U6 small ribonucleoproteins particles (snRNPs), in association with almost 170 different proteins factors (Maniatis and Tasic, 2002).

The first system used for studying mRNA splicing and the structure of the pre-mRNA molecule was the late stage of adenovirus infection in mammalian cells. The presence of introns was first described in the mRNA segment of adenovirus, coding for the Hexon polypeptide, the major virion structural protein (Berget et al., 1977). In this system, different RNA molecules called “mosaics” were identified, which contained sequences from non-contiguous sites in the viral genome. Studying the process before infection identified that the long RNA contained the sequence of the late RNA in addition to the sequence that has been called “the intervening sequences” (intron). After the discovery in adenovirus of the presence of long sequences, the introns were found in other viral and eukaryotic gene such as haemoglobin and immunoglobulin (Breathnach et al., 1978). After that the characterization of the RNA splicing process in viruses, the presence of introns was also reported in eukaryotic genes.

The maturation of RNA, which follows the pre-mRNA transcription by RNA polymerase II (Pol II), is required not only for the splicing process but also for other steps such as capping, polyadenilation and RNA editing. Some studies have delineated functional relationships between these maturation processes and each of them (with the exception of the editing process) can occur co-transcriptionally.
Many findings point out that during mRNA bio-genesis the different processing machineries responsible for capping, splicing, poly-adenylation, modification and transport of mRNAs can interact with the elongating RNA Polymerase II, suggesting that all these processes are physically and functionally intertwined. Consequently the RNA transcription and processing seem to take place in "gene expression factories" and where distinct activities are functionally combined in order to maximize efficiency and to extend possibility of regulation. In fact during the RNA process a complex network of functional interactions are formed between the elongation RNA polymerase II and the different processing machines responsible for capping, splicing, polyadenylation, modification, and transport of mRNAs (Bentley, 2002; Kornblihtt et al., 2004; Maniatis and Reed, 2002; Proudfoot et al., 2002).

In fact, the process of splicing occurs in all species, and the chemistry of splicing is highly conserved from yeast to human, although yeast has few short introns and humans multiple long introns. The high prevalence of introns in human contributes to the proteomic diversity in a process called alternative splicing, where the exon can be spliced in different ways and create different mRNA products that contribute to the formation of distinct proteins.
A change in alternative splicing can be related to human disease, unfortunately most changes caused by alternative splicing are hard to detect, due to the complexity of the splicing process (Kelemen et al., 2013).

1.1.1 Chemical aspects of splicing.

The process of splicing is a complex process requiring different proteins and several small RNAs forming the spliceosome complex. The primary signals that determine where the process of splicing occurs, are included within the transcript itself. These are the donor site (5' splice site), the acceptor site (3' splice site), the branch point that contains the branch point nucleotide Adenosine residue, and the polypyrimidine tract (PPT) characterized by 20-40 nucleotides (nt) of pyrimidine located between the branch point nucleotide and the 3' splice site (Figure 1.1).
The splicing reaction takes place through two consecutive *trans* esterification steps (Konarska et al., 1985; Moore and Sharp, 1993; Ruskin et al., 1984).
In the first reaction of splicing the 5' exon is cleaved through nucleophilic attack by the 2 hydroxyl of the specific branch-point adenosine (located within the 3' intron) to generate a 5' exon fragment and a lariat intermediate that contains intron and 3' exon sequences and the branched adenosine. In the second reaction, cleavage at the 3' exon/intron boundary occurs through nucleophilic attack of the 3' hydroxyl of the 5' exon at the 3' splice site, which ligates the exons and releases the intron in the form of a lariat (Figure 1.2) (Umen and Guthrie, 1995).

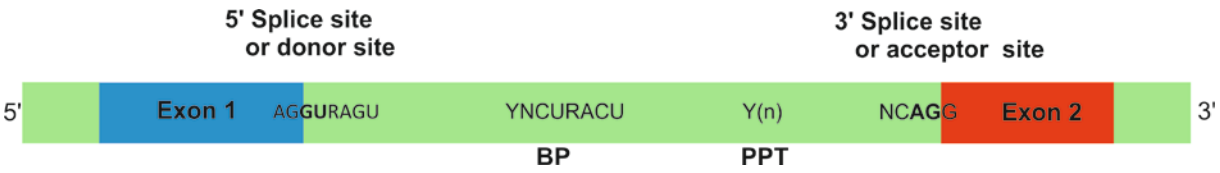


Figure 1.1 Schematic representation of exon/intron boundaries.

The basic elements of a transcript. The blue and the red boxes represent exons. The green box represents the introns. The consensus sequence for the donor site (5'splice site), acceptor site (3' splice site), Adenosine branch point (BP), and polypyrimidine tract (PPT) are given. R=Purine, Y=Pyrimidine, N=any nucleotide.

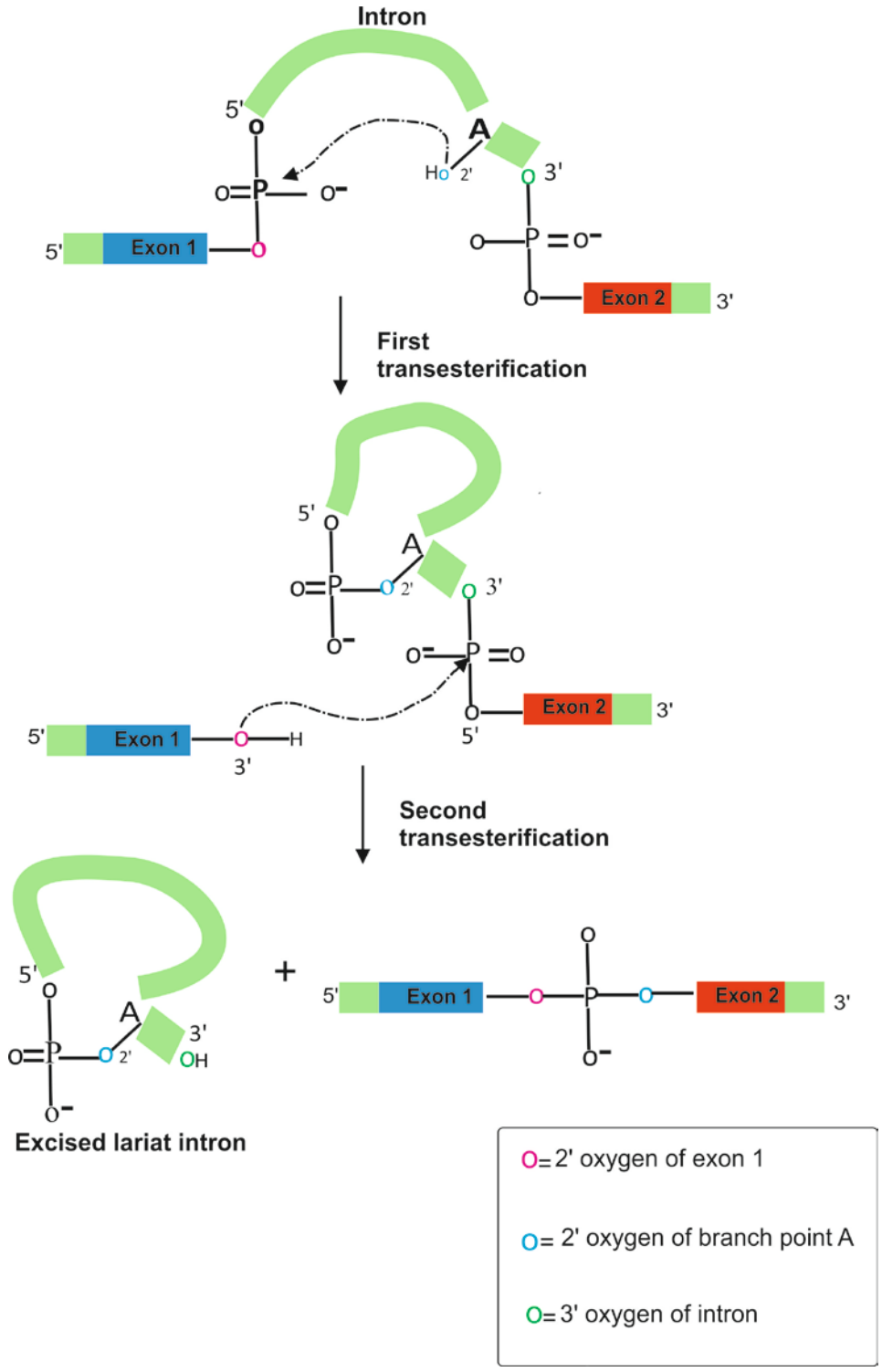


Figure 1.2: Schematic representation of the chemical steps in the splicing process.

In the first reaction, the ester bond between the 5' phosphorous of the intron and the 3' oxygen (red) of exon 1 is exchanged for an ester bond with the 2' oxygen (blue) of the branch-site A residue. In the second reaction, the ester bond between the 5' phosphorous of exon 2 and the 3' oxygen (green) of the intron is exchanged for an ester bond with the 3' oxygen of exon 1, releasing the intron as a lariat structure and joining the two exons.

1.1.2 The spliceosome.

The pre-mRNA splicing reaction requires the formation of a highly dynamic macromolecular complex called the spliceosome composed of several small nuclear ribonucleoproteins (snRNPs) and a large number of non-snRNPs splicing factors (Jurica and Moore, 2003; Rappsilber et al., 2002; Wahl et al., 2009). The spliceosome complex acts through different types of interactions such as RNA-RNA, RNA-protein and protein-protein interactions. Each snRNPs particle is formed from a uridine-rich small nuclear RNA (U snRNA) molecule (or two in case of U4/U6) in complex with a common set of Sm proteins or Sm-like proteins (Will and Lührmann, 2001) (Figure 1.3). The snRNPs are also associated with other different proteins to form the spliceosome (Bessonov et al., 2008).
The structural core of snRNPs is formed by eight proteins, called Sm proteins, B', B, D1, D2, D3, E, F and G. These classes of common proteins play an essential role in the biogenesis of the snRNPs. The Sm proteins form three distinct heterogenic complexes prior to their interaction with the highly conserved Sm site (PuAU4-6GPu flanked by two stem-loop structures) of the U1, U2, U4 and U5 snRNAs (Raker et al., 1996).

In particular U1 snRNP consists of ten different proteins, seven common Sm proteins, and another three (U1-70K, U1-A and U1-C) U1-specific proteins (Stark et al., 2001). 17S U2snRNP represents the active form of U2 that is active in complex assembly (Behrens et al., 1993). Apart from the seven Sm proteins, numerous U2 specific proteins have been identified. These include the stably associated U2-A' and U2-B' polypeptides, and the heteromeric protein complexes SF3a and SF3b (Kramer, 1996; Kramer et al., 1999).

Five different proteins have been described associated with U4/U6 snRNA (Teigelkamp et al., 1998) and 8 different Sm proteins are associated with U5 snRNA (Turner et al., 2004; Will et al., 2002) (Figure 1.3).

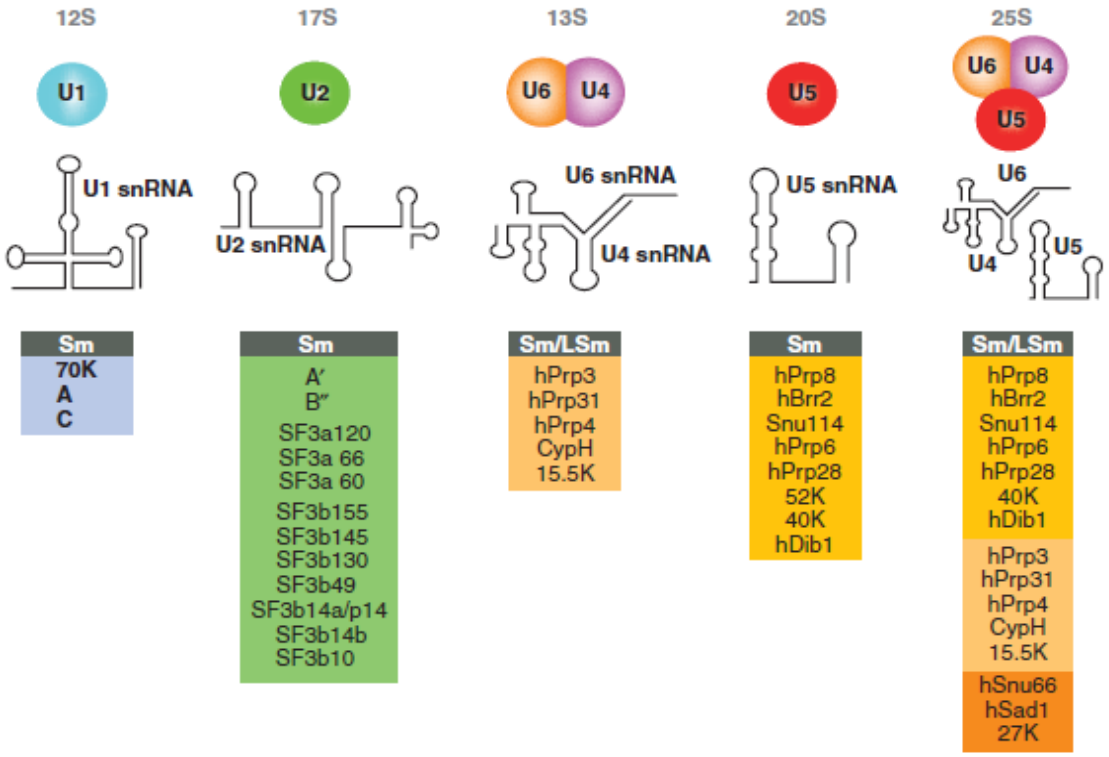


Figure 1.3: Proteins compositions and RNA secondary structure of snRNPs.

The figure represents the protein composition and snRNA secondary structures of the major human spliceosome snRNPs. All seven Sm proteins (B/B', D3, D2, D1,E, F, G1) or Lm proteins Lsm (2-8) are indicated by Sm or Lm on the top of the box (figure adapted from (Ou et al., 1995).

The spliceosome formation consists of a precise and ordered assembly/disassembly of the different snRNP particles onto the pre-mRNA. The spliceosome needs to recognize specific RNA sequences, known as the 5' splice site (MAG/GTAAGTA), the 3'splice site (NYAG/R), the adenosine branch point and the polypyrimidine tract (M indicates C or A, N indicates any nucleotide; Y pyrimidine; R purine; / intron-exon boundary) (Bessonov et al., 2010; Reed, 1989; Zhang, 1998; Zhuang et al., 1989).

In mammals four distinct spliceosomal complexes, which form in the temporal order E, A, B and C, have been detected (Jenkins et al., 2013; Jurica and Moore, 2003; Zhang et al., 2013). The E, A and B complexes contain unspliced pre-mRNA whereas the C complex has the products of catalytic step I of the splicing reaction (exon 1 and lariat-exon 2). Although the earliest assembly step (E complex) is energy independent, subsequent spliceosomal steps require NTP hydrolysis (Staley and Guthrie, 1998) (Figure 1.4). The E (Early) complex (commitment complex) assembly starts with the recognition of the 5'splice site driven by the U1 snRNP in an ATP-independent manner (Figure 1.4). This interaction is mediated by base pairing of the U1 snRNA with the 5'splice site, as well as by protein-protein and protein-pre-mRNA interaction involving U1-70K and U1-C proteins (Will et al., 1996). In addition to the U1-5'splice site interaction the Early complex phases also involve the binding of the splicing factor 1/branch point protein (SF1/BBP) to the branch point, which is usually located 20-40 nucleotides (nt) upstream of the 3'splice site (Berglund et al., 1997). The SF1/BBP protein binds the Adenosine Branch site and interacts with U2 Auxiliary factor (U2AF) through its C-terminal RNA recognition motif (RRM) (Behrens and Luhrmann, 1991).

U2 snRNP auxiliary factor (U2AF) is constituted of two subunits of 65-kDa (U2AF65) and 35-kDa (U2AF35). The U2AF65 subunit specifically recognizes the splicing factor SF1/BBP (Valcarcel et al., 1996a), but is also able to bind the branch site through its N-terminal arginine- and serine-rich (RS) domain. In the

proline-rich region of U2AF65, between the RS motif and RNA-binding motifs, is located the binding domain for U2AF35 (Newman, 1997; Sander et al., 2006).

The small subunit U2AF35 is involved in the recognition of the essential AG dinucleotide of the 3'splice site during the earliest stage of spliceosome assembly (Wu et al., 1999). U2AF35 is made up of a central homology motif (UHM motif) flanked by two zinc fingers in the N-terminal region while the C-terminus contains an arginine-serine-rich domain (RS domain) and a glycine tract. U2AF35 binds both U2AF65 and the pre-mRNA through its RRM domain. Both U2AF subunits (U2AF35 and U2AF65) bind to intronic sequences only during the early steps of spliceosome assembly and subsequently are replaced by the U5 snRNP during the formation of the B complex (Figure 1.4).

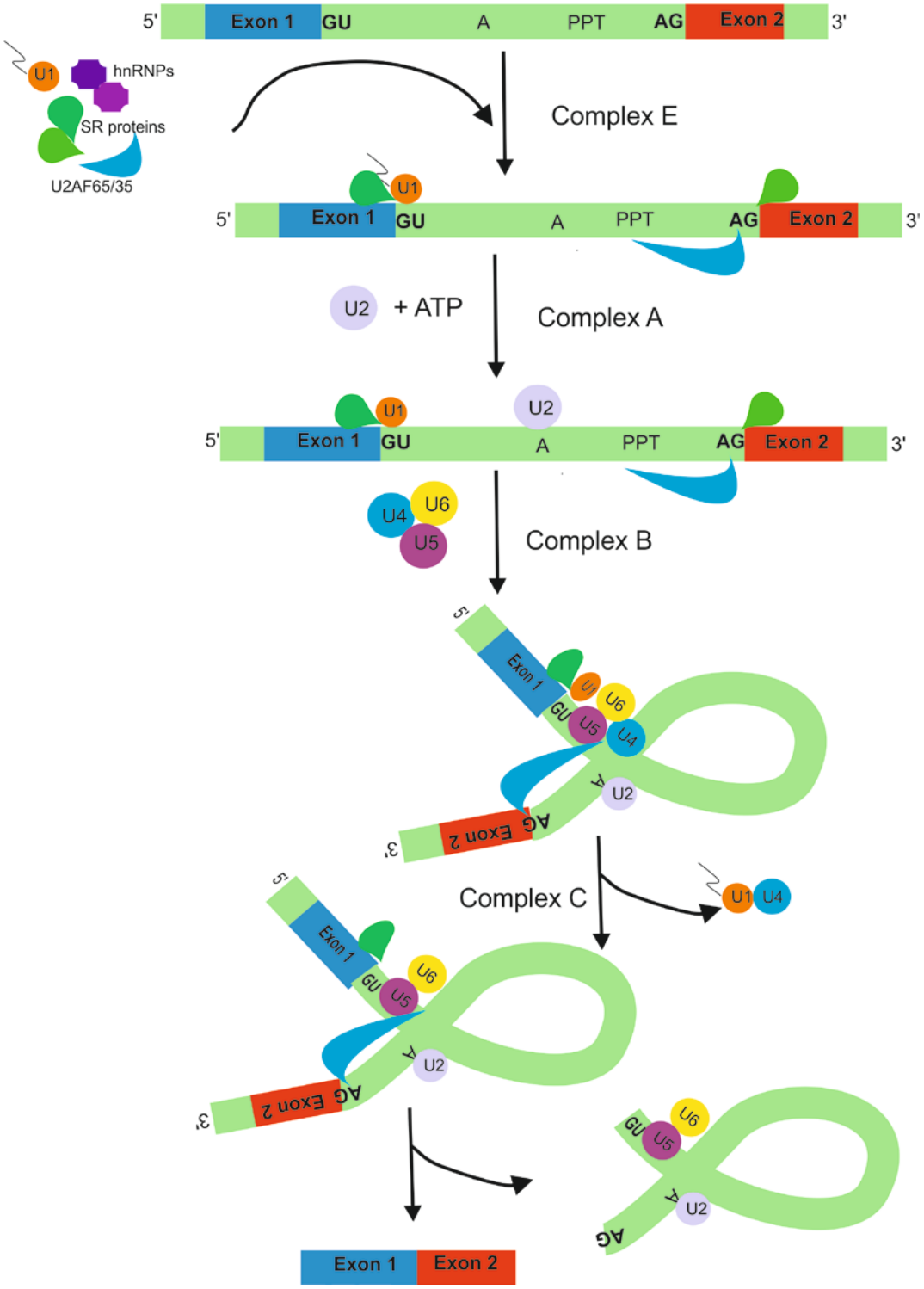


Figure 1.4: Spliceosome assembly.

The spliceosome assembles on the pre-mRNA in a stepwise manner. The E complex contains U1 snRNP bound to the 5' splice site, SF1 bound to the branch

point, and U2AF65 and U2AF35 bound to the pyrimidine tract and 3' splice site AG, respectively. In the A complex, SF1 is replaced by U2 snRNP at the branch point. The U4/U6/U5 tri-snRNP then enters to form the B complex. Finally, a rearrangement occurs to form the catalytically active C complex, in which U2 and U6 interact, and U6 replaces U1 at the 5' splice site. After that the intron is removed and exons joined together to form the mature RNA.

The formation of the A complex is characterized by the ATP-dependent recruitment of the U2 snRNP to the branch point sequence through replacement of the BBP/SF1 factor (Figure 1.4). This U2–branch site binding is mediated by U2 snRNA base pairing with the BP which is further stabilized through SF3a and SF3b subunits (Gozani et al., 1996) and also by the arginine-serine-rich domain of the U2AF65 protein (Valcarcel et al., 1996b).
The transition from the A complex to the B complex is characterized by the recruitment of the U4/U6-U5 Tris-snRNP to the un-spliced pre-mRNA (Turunen et al., 2013), making the spliceosome ready for the first trans esterification reaction (Hall and Padgett, 1996) (Figure 1.4). However it has been reported that the tri-snRNPs are also able to interact with the 5'splice site and the upstream 5' exon at earlier step of spliceosome assembly (Maroney et al., 2000).

Although the B complex contains all of the snRNPs components required for splicing, it lacks a catalytic center. In order to activate the spliceosome, it is important to have the RNA-RNA rearrangements that involve the displacement of U1 by U6 snRNP via base pairing at the 5'splice site through its highly conserved ACAGAG motif, the disruption of the U4/U6 base pairing interaction and the formation of an intricate network of interactions among the U6, U2 snRNAs. At this stage the U1 and U4 snRNP are released from the spliceosome (Figure 1.4).
The U2 and U6 snRNAs are extensively base-paired with one another and with the branch site sequence and 5'splice site respectively, that provide a structural basis for making the branch site and 5'splice site ready for the first catalytic step (Boehringer et al., 2004; Reed, 2000). All these rearrangements contribute to transfer the premature B complex to a catalytically active B complex (B* complex) (Turner et al., 2004; Zhang et al., 2013).
The formation of the B* complex promotes the first catalytic step of splicing in order to generate the free 5' exon and the lariat-3' exon intermediates. This step is followed by the formation of the C complex, in which the second catalytic step of splicing reaction takes place (Figure 1.4). U5 snRNP plays an important role in the second catalytic step because it holds the 5'and the 3' splice site exons in

close proximity (Roca and Krainer, 2009). After the second catalytic step the spliceosome releases the mRNA and the U2 U5 and U6 snRNPs to be recycled for additional rounds of splicing (Wahl et al., 2009).

In addition to the major spliceosome there is a minor spliceosome that is used for splicing in a small fraction of mammalian cells. The minor spliceosome is present in different complex eukaryotic species such as plant, fungi and animals as well as single eukaryotic species but are absent in many species, including such common model organisms as *Caenorhabditis elegans* and *Saccharomyces cerevisiae* (Gunzl et al., 2002). The minor spliceosome uses alternative snRNPs and alternative splicing regulatory sequences. The major spliceosome U1, U2 and U4/U6 snRNPs are substituted with U11, U12 and U4atac/U6atac (Hwang and Cohen, 1996).

1.1.3 Cis-acting regulatory element.

The process of splicing is directed by the presence of specific consensus sequences at the exon/intron junction known as 5’ splice site and 3’ splice site (Burge and Sharp, 1999).
The recognition of these sequences could be a linear process; however the nucleotides surrounding the 3’ splice site and 5’ splice site are highly degenerate. There is evidence that in mammals only the dinucleotide core of the 5’ splice site (GU), the dinucleotide core of the 3’ splice site (AG), the adenosine branch point (BP), and the poly-pyrimidine (pY) tract between the BP and the 3’ splice site are universally conserved (Aebi et al., 1987).

1.1.4 The 5’ splice site.

The 5’ splice site motif in eukaryotes contains nine partially conserved nucleotides, MAG/GURAGU (M indicates A or C, R indicates purines and the slash the exon-intron boundary) at the exon-intron junction.
The GU dinucleotide core of the 5’ splice site consensus sequence is universally conserved and mutations in one of these two nucleotides completely abolish the splicing process (Langford et al., 1984).
The 5’splice site has been shown to be recognized by the U1 snRNA during early assembly of the spliceosome machinery (Horowitz and Krainer, 1994; Siliciano and Guthrie, 1988; Zhuang and Weiner, 1986). Although base pairing between U1 snRNA and the 5’ splice site is important, there is evidence that the selection of the 5’ splice site through the U1-C subunit is also possible when there is not high complementarity between the U1 snRNAs and the 5’ splice site (Zhu et al., 2001). Studies in literature demonstrated that the interaction between the U1 snRNP and the 5’ splice site is not highly complementary; in fact sometimes binding to a nonspecific 5’ splice site could be possible (Hicks et al., 2010). The binding between U1 snRNP and the 5’ splice site has been demonstrated *in vitro* using the RNase H protection assay (Gunzl et al., 2002).

During the assembly of the spliceosome, other factors are also important for the recognition of the 5’splice site, such as the U6snRNP. The interaction of U6 SNRP with the 5’ splice site is enhanced by the presence of U1 snRNP which in turn is displaced by U6 that binds the 5’ splice site and stimulates the first trans esterification reaction (Buratti et al., 2007; Grover et al., 1999) (Figure 1.5).
Other factors involved on the recognition of the 5’splice site and the correct choice of the 5’ splice site, if alternative 5’ splice site are present, these factors are the SR and the hnRNP proteins.
Hick et al. demonstrated that the SR proteins are able to bind in proximity of the 5’ splice site and are able to stimulate the use of the 5’ splice site located downstream of their binding sites rather than the one upstream (if present) that

will be inhibited (Buratti et al., 2004; Hicks et al., 2010). A major study has been done on the one of the major SR proteins the SRSF1. The SRSF1 protein has a RRM domain, which interacts with the RRM domain of U1-70K (an auxiliary protein of U1 snRNP), this interaction is controlled by the phosphorylation of the SRSF1 RS domain. In conditions of hyper-phosphorylation the RS domain allows the RRM motif of the SR protein to bind the RRM motif of the U1snRNP, when not phosphorylated the RS domain sequesters the RRM motif and there will be no binding (Cho et al., 2011; Kramer and Utans, 1991).

Interestingly, *in vitro* experiments over-expressing SR proteins demonstrated that it is possible to stimulate the recognition of the 5' splice site even in the absence of U1snRNP (Crispino et al., 1994; Tarn and Steitz, 1994).

In addition to the SR proteins, the hnRNP A1 protein has been shown to modulate the selection of the 5' splice site. In particular, the hnRNPA1 can disrupt binding between U1snRNP and the 5' splice site (Eperon et al., 2000; Pagani and Baralle, 2004).

For hnRNPA1 it has been demonstrated that a mutation in the 5' splice site of exon 3 of NF1 can alter the process of splicing due to alteration of the binding site of U1 snRNP and the binding site of hnRNPH in the presence of the mutation G>C (Buratti et al., 2004).

It has been proposed that in some cases hnRNP A1 and hnRNP H to promote the recognition of the 5'splice site. In these cases hnRNPA1 and H interact with their binding sites on intronic RNA, bringing distal 5' and 3' splice sites closer, favouring the splicing process (Fisette et al., 2010; Pagani and Baralle, 2004).

In conclusion the choice of a perfect 5' splice site depends on different factors, including U1snRNP and U6snRNP during the spliceosome formation, as well as the SR and hnRNP proteins. hnRNP factors are associated with two functions: the stimulation and the inhibition of the selection of splice sites (Fisette et al., 2010).

Finally it has been shown that the presence of secondary structures close to the 5' splice can also affect its recognition. An example is seen in the tau gene where a mutation in the 5' splice site of the exon 10 affects a stem-loop structure (Zhang

et al., 2005). Other diseases associated with the presence of secondary structure are ATM and CFTR. For ATM a deletion in the repressor element causes use of an aberrant 5’ splice site. In CFTR, on the other hand, a new 5’ splice site is created. In both cases RNA secondary structure is the key regulatory element (Manley and Krainer, 2010).

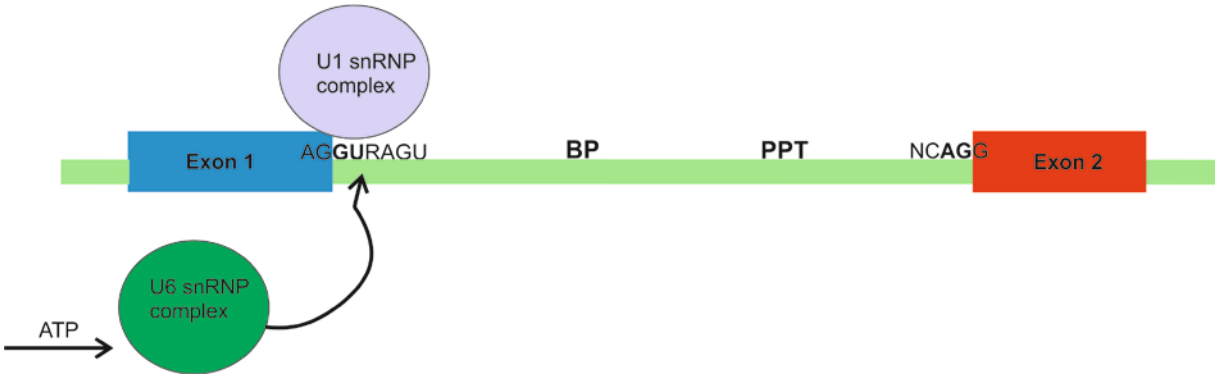


Figure 1.5: Schematic representations of the 5' splice site recognition.

The figure shows a schematic representation of the recognition of the 5' splice site by the U1snRNP complex during the early assemble of the spliceosome complex followed by the recognition of the U6snRNP in presence of ATP. The blue and the red boxes represent the exons and the light green box represents the intron. The green and the violet circle are respectively the U6snRNP and the U1 snRNP.

1.1.5 The 3’ splice site.

The 3’ splice site has three different conserved sequence elements: the branch point, polypyrimidine tract and the AG dinucleotide (Figure 1.6). The 3’ splice site is recognized during the early step of spliceosome assembly by the U2AF35 subunit (Wu et al., 1999), which is part of the auxiliary protein U2AF.

The branch point: In yeast this is a conserved sequence (UACUAAC), while in metazoan, apart from the universally conserved A, it is not highly conserved (Lim and Burge, 2001). However, evidence shows that the mammalian branch point is specified primarily by its proximity to the intron/exon junction, and the general consensus sequence YNYURAC motif (R-Purine, Y-pyrimidine). Most branch points have been mapped within 18-40 nucleotides of the 3’ splice site (Reed and Maniatis, 1985; Ruskin et al., 1985). However, there are exceptions where the branch point is located further than 40 nucleotides from the 3’ splice site but are nevertheless functional and essential for the regulation of alternative splicing.
The branch point is recognized by the SF1 factor during early spliceosome assembly (complex E) (Berglund et al., 1997). Subsequently recognition of the branch site involves binding of U2 snRNP through interaction with SF3a and SF3b in order to form the spliceosome A complex (Newman, 1997; Zhong et al., 2009).
Mutation of the adenine residue involved in the lariat formation strongly reduces splicing efficiency of the downstream exon (Reed and Maniatis, 1988).

The polypyrimidine tract: The polypyrimidine tract is a series of pyrimidines located between the branch site and the terminal AG at the intron/exon junction. The polypyrimidine tract can bind several proteins, such as the auxiliary factor U2AF65-kDa subunit, and the polypyrimidine tract binding protein (PTB) (Gooding et al., 1998; Wagner and Garcia-Blanco, 2001). U2AF65 binds the polypyrimidine tract during the formation of the early (E) spliceosome complex (Kielkopf et al., 2004; Zamore et al., 1992). Binding of the Polypyrimidine Tract

Binding protein (PTB) to the polypyrimidine tract of a 3'splice site can inhibit splicing by directly blocking binding of U2AF65 (Sauliere et al., 2006).
The function of the U2AF65 is to bind the polypyrimidine tract and to bring the 3'splice site and the adenosine branch point closer to each other (Sauliere et al., 2006).
It has been shown that deletion of the polypyrimidine tract abolishes the formation of the lariat intermediate, during the second trans esterification reaction (Mullen et al., 1991; Roscigno et al., 1993)

The terminal AG dinucleotide: The terminal AG dinucleotide defines the 3' border of the intron. This site is characterized by a short YAG/G sequence (Y denotes pyrimidine; the slash indicates the intron-exon boundary and the underlined nucleotides are conserved) (Langford et al., 1984).

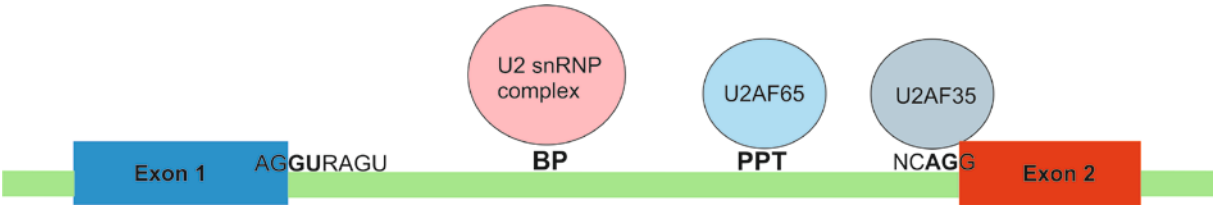


Figure 1.6: Schematic representations of the 3' splice site recognition.

The figure shows a schematic representation of the recognition of the 3' splice site. The exons are represented by blue and red boxes. The U2 snRNP binding the adenosine branch point (BP), the U2AF65 binding the polypyrimidine tract (PPT) and the U2AF35 binds the AG dinucleotide. The blue and the red boxes represent the exons and the light green box represents the intron. The pink and the light blue circle are respectively the U2snRNP and the U2AF complexes (U2AF65 and U2AF35).

1.2 The additional *cis*-acting elements.

The process of splicing is directed by the presence of specific consensus sequences at the exon/inton junction known as the 5' splice site and 3' splice site (Burge and Sharp, 1999). These splicing signals, at the boundaries, are necessary but alone insufficient for the correct splicing process to happen, due to their degenerate nature.

It has been proposed that the recognition of splice sites takes place across the exons through the initial interaction between the 5’ and 3’ splice sites via the exon definition model (Berget, 1995). This initial splice-site recognition across the exon is the result of a combinatorial regulatory mechanism (Smith and Valcarcel, 2000) that uses additional controlling elements. Such additional splicing regulatory elements, can act by increasing or decreasing exon recognition, are named exonic or intronic splicing enhancers (ESE, ISE) or silencers (ESS, ISS), depending on their location and whether they exert a positive (enhancers) or negative (silencer) action on the recognition of splice sites (Cartegni et al., 2002) (Figure 1.7).

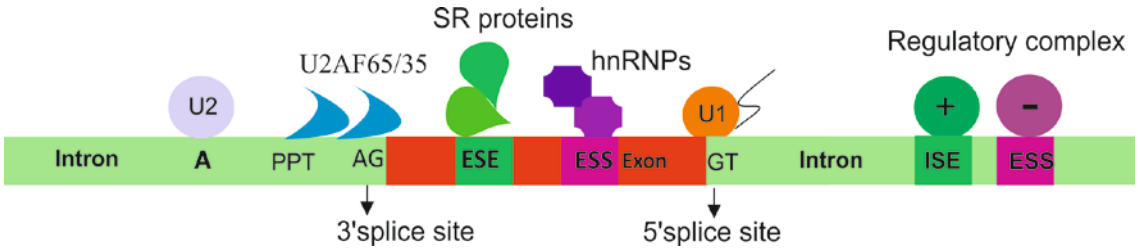


Figure 1.7: Schematic representation of the possible distribution of canonical and additional splicing cis-elements.

The figure shows a schematic representation of the elements involved in the splicing process.
The canonical splicing signal that defines the exon boundaries are: the GT at the 5' splice site, the AG at the 3'splice site and the branch point adenosine.
The U1 snRNP binds to the 5'splice site, the U2 snRNP recognizes the branch site and the U2AF recognizes the polypyrimidine tract and the 3' splice site sequence. Additional enhancer and silencer elements exist in the exon (ESE, ESS) and/or intron (ISS; ISE) allowing correct recognition of the splice sites. Trans-acting splicing factors can interact with enhancers and silencers and are broadly divided into two main groups: the SR family of proteins and hnRNPs. In general, SR proteins binding at ESEs facilitate exon recognition whereas hnRNPs are inhibitory (Twyffels et al., 2011).

1.2.1 Regulatory elements: enhancers and silencers.

ESEs are specifically recognized by various RNA binding proteins, the most prominent of which belong to the serine/arginine-rich (SR) protein family of splicing factors.

The commonest exon splicing enhancers (ESE) contain extended purine-rich sequences (more than 65% purine rich) (Tian and Kole, 1995). ESEs are usually located downstream of suboptimal 3′ splice sites (Fairbrother and Chasin, 2000). A purine-rich ESE usually has one or more binding sites for serine-arginine rich proteins (SR proteins). The purine-rich ESEs recruit or strengthen, through interactions with SR proteins, the binding of U2AF to the upstream 3′ splice site and stimulate spliceosome assembly (Graveley et al., 1998). However, an exonic sequence having one or more binding sites for SR proteins does not necessarily function as an ESE (Zheng et al., 1999); in fact it is possible that an exonic splicing suppressor binds the SR protein and have a negative effect on splicing. Usually the ESS proteins are found downstream from the ESE (Figure 1.8).

Depending on their locations purine rich sequences can function as exonic splicing enhancers or intronic splicing suppressors. In fact, a purine rich element can suppress the recognition of splice sites when located in the intron, and therefore act as a silencer of splicing (Skordis et al., 2003b).

Another class of ESEs is the non-purine-rich ESEs. This class includes the exonic AC-rich enhancer and exonic pyrimidine-rich enhancer. The AC-rich enhancers were first identified by *in vivo* selection experiments and were found to stimulate splicing *in vivo* and *in vitro* (Coulter et al., 1997).

Exonic splicing enhancers (ESEs) are implicated in constitutive splicing events, in addition they were also identified as regulators of alternative splicing (Black, 2003) and implicated in some constitutive splicing events (Lavigueur et al., 1993; Schaal and Maniatis, 1999). In general ESEs, through SR protein binding, create the process of splicing by recruiting splicing factors and/or by antagonizing the action of nearby splicing silencer elements (Cartegni et al., 2002).

Several groups, through the functional systematic evolution of ligands by exponential enrichment (SELEX) (Tuerk and Gold, 1990), have identified other functional ESE sequences. This method was utilized to identify sequences that can act as ESEs in response to specific SR proteins and have characterized SR-protein-specific sequence motifs (Liu et al., 2000; Liu et al., 1998). The ESE motifs identified are usually 6–8 nucleotides long with a degenerate consensus sequences (generally more so than the consensus sequences for the canonical splice sites). The frequencies of individual nucleotides at each position of the consensus were used to derive score matrices in order to predict the location of SR protein-specific putative ESEs (Liu et al., 2000; Liu et al., 1998).

Whilst enhancer elements that are able to promote the inclusion of the exons, silencer elements exist, which usually have an opposite effect. The silencer motifs are less well characterized: the sequence of these can be purine or pyrimidine-rich and bind different splicing factors (Fairbrother and Chasin, 2000). Silencer sequences generally interact with the proteins belonging to the heterogeneous nuclear ribonucleoproteins family (hnRNP) (Figure 1.8) (Cartegni et al., 2002).

A silencer sequence can regulate splicing in different ways: by antagonizing the function of a nearby ESE or by recruiting factors that interfere with the splicing machinery by steric hindrance or through exon looping out of the pre-mRNA or by nucleation and cooperative binding (Cartegni et al., 2002; Matlin et al., 2005). Previous studies based on the exon repression mechanism have shown that inhibition typically occurs during the initial ATP-independent recognition of splice sites. However, the dynamic nature of the splicing machinery suggests that any of the intermediates along the spliceosome assembly pathway are potential targets for biologically relevant regulation (House and Lynch, 2006).
Similar studies have also been done for intronic splicing silencers (ISS). A well-known ISS is the binding site for the poly pyrimidine tract binding protein (PTB) splicing factor. PTB usually binds the polypyrimidine tract in the intron and acts

by having a negative effect on the U2AF65 binding or creating a silencer across the exon (Soret et al., 2006).

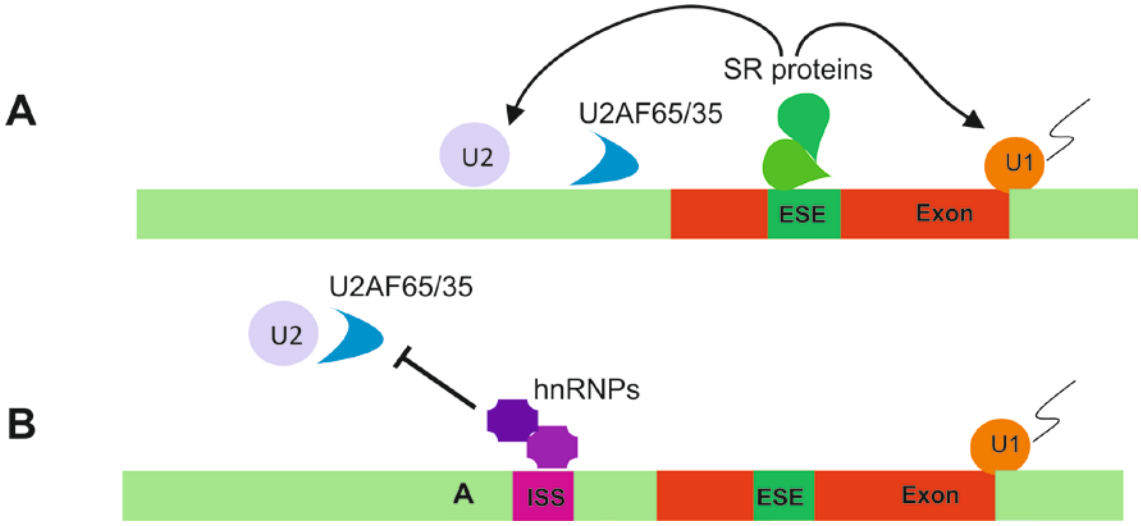


Figure 1.8: Schematic representation of the potential distribution of enhancing and silencing splicing regulatory elements.

The red boxes represent exons; the green boxes represent the introns. U1 and U2 are represented with the orange and violet circle respectively. The U2AF complexes are represented in blue. The SR protein and the hnRNP are designed in green and violet respectively.
(**A**) The SR proteins can function by binding to exonic splicing enhancers (ESEs), where they can recruit U2AF and/or U1 snRNP to the 3' and 5' splice sites, respectively, and can activate the splicing process.
(**B**) Splicing repressors such as: hnRNP proteins can inhibit splicing by binding to intronic splicing silencers (ISSs), where they interfere with the binding of U2AF.

1.3 Additional proteins involved in splicing: Trans-acting factors.

During spliceosome assembly the small nuclear ribonucleoproteins particles (snRNPs) and non-snRNP splicing factors are involved in the splicing process. In addition two families of RNA binding proteins, the serine-arginine rich proteins (SR proteins) and the heterogeneous nuclear ribonucleoproteins (hnRNP factors), have been described as the main components of distinct regulatory complexes with functional specificity in the splicing process and that usually bind the ESEs and ESSs (David and Manley, 2008).

1.3.1 The serine-arginine proteins (SR proteins).

Serine-arginine proteins (SR), also called SR splicing factors (SRSFs) (Mayeda et al., 1999) are families of structurally related RNA binding proteins, highly conserved in metazoan cells. Besides splicing, the SR proteins play multiple roles in all steps of mRNA metabolism such as, mRNA export, stability, quality control and translation (Wu and Maniatis, 1993; Zhou and Fu, 2013). Defects of these functions may contribute to disease (Blaustein et al., 2007).

The serine-arginine proteins were first discovered as splicing factors in the early 1990s. They contain one or two N-terminal RNA-recognition motifs (RRM) and a specific C-terminal domain rich in repeating arginines and serines, the “RS” domain (Ars et al., 2000; Birney et al., 1993). In humans this classification identified 12 SR proteins, now designated as Serine/Arginine rich Splicing Factors (SRSF 1-12) (Figure 1.9).

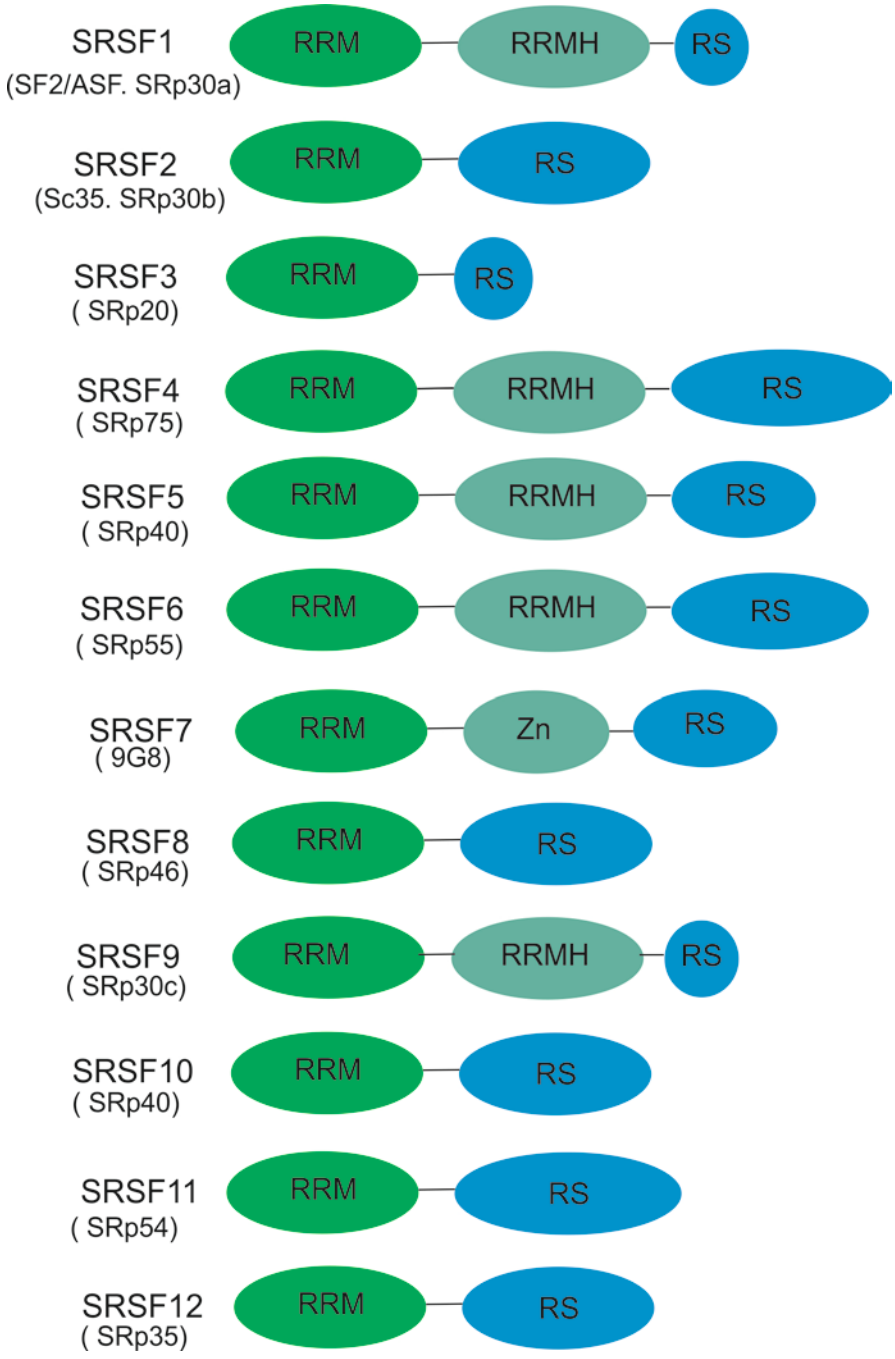


Figure 1.9: The SR protein families.

The figure shows twelve SR proteins (SRSF1-SRF12). RRM is the RNA recognition motif; RRMH is the RNA recognition motif homolog; RS is the arginine/serine rich domains; Zn is the Zinc knuckle (Shepard and Hertel, 2009).

The RRMs mediate sequence-specific binds to the RNA, and so determine substrate specificity (Buee et al., 2000) whereas the RS domain seems to be involved mainly in protein-protein interactions. Other proteins (for example snRNP associated factors such as U1 70K) share this structural organization. SR proteins are essential pre-mRNA splicing factors and critical regulators of the selection of both constitutive and alternative splice sites. They are present in all metazoans but, curiously, are apparently absent from lower eukaryotes such as *Saccharomyces cerevisiae* (Hutton et al., 1998). SR proteins have been implicated in many steps of spliceosome assembly, from recognition of the 5' splice site at the earliest step of splicing to binding and regulation of exonic-enhancer sequences that stimulate the usage of sub optimal splice sites.

There are different models for the mechanism of SR proteins action. The most accepted model for the mechanism of SR protein action is the ability of these splicing factors to bind exonic splicing enhancers (ESEs) and through their RS domain to recruit and stabilize U1 snRNP and U2AF binding to the 5' and 3'splice site respectively and, thus, facilitate recruitment of the spliceosome (Corrionero et al., 2011).

SR proteins have also been suggested to allow protein-protein interactions across introns in a way that brings the 5' splice site and the 3' splice site closer. This intron bridge is mediated by the interaction of the SR protein through the RS domain to the U1 snRNP-associated 70kDa protein (U1 70K) at the 5' splice site and the U2AF at the 3' splice site (Figure 1.10) (Cartegni et al., 2002; Sanford et al., 2005). In addition the RS domain of the ESE bound SR protein has been shown to interact with RNA at the Branch Point (BP) to promote pre-spliceosomal assembly (Shen et al., 2004). SR proteins can also act by antagonizing the negative effect on splicing of an inhibitory protein bound to a juxtaposed exonic splicing silencer (ESS) (Figure 1.10).

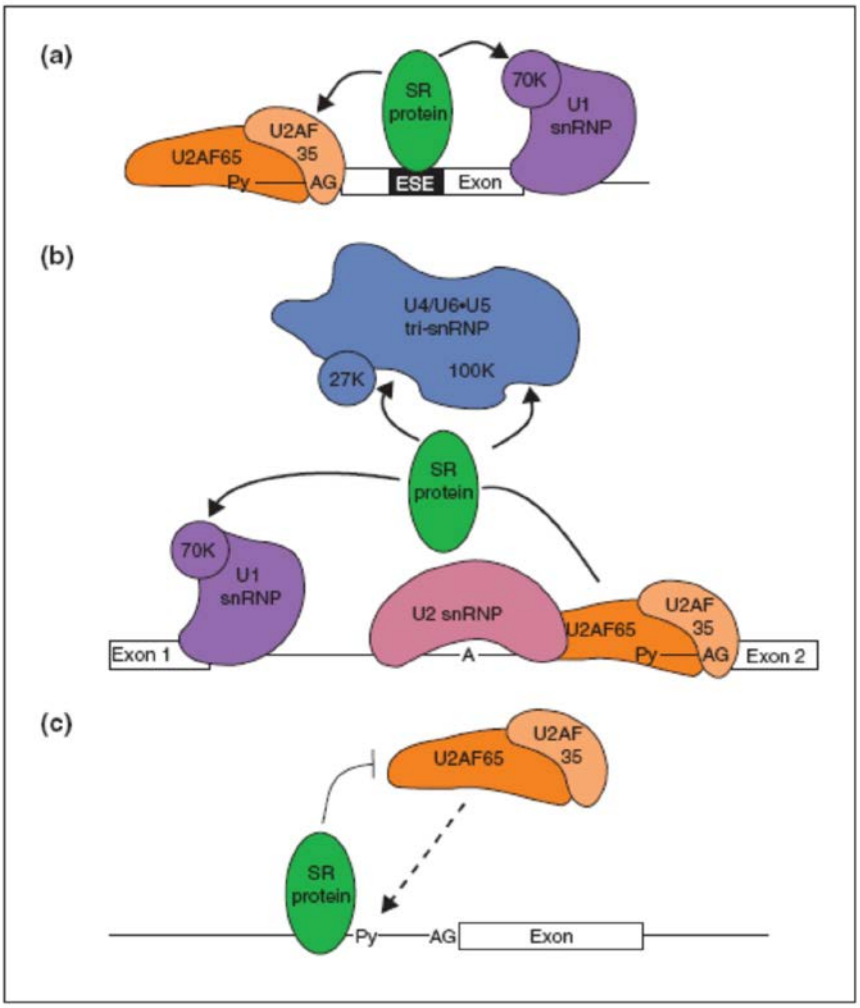


Figure 1.10: Splicing function of SR proteins.

The figure shows three different interaction of SR protein.
(a) SR proteins (green) bound to ESE by interacting with the splicing factors U2AF.
(b) Exon independent functions of SR proteins. SR protein facilitates splice site recognition by interaction with U1snRNP and U2AF. SR proteins also recruit the U4/U6/U5 trisnRNP.
(c) Splicing repression is mediated when SR is associated with intronic sequence (Shepard and Hertel, 2009).

Thus, the structural organization of SR proteins and their ability to interact with other proteins suggest a model for their function, where the RRM(s) mediate sequence-specific binding to the mRNA, and so determine substrate specificity, whereas the RS domain seems to be involved mainly in protein-protein interactions (Cartegni et al., 2002; Valcarcel and Green, 1996), but has also been observed to mediate RNA-protein interaction (Shen and Green, 2004).

SR protein activity is regulated through phosphorylation/dephosphorylating (Stamm, 2008), where the RS domain of SR proteins is phosphorylated and dephosphorylated during the spliceosome maturation by several proteins family such as: the Serine/Arginine protein kinase (SRPKs), the CDC2-like kinase family (CLKs) and the AKT family (Scott et al., 2012).

This post-translational modification is a crucial step for splicing organization inside the cell nucleus by affecting the RNA-binding activity and sub nuclear localization of the SR proteins (Misteli and Spector, 1997). In fact, while localized predominantly in the nucleus, some (but not all) SR proteins shuttle continuously between the nucleus and the cytoplasm (Caceres et al., 1998). The RS domain phosphorylation is required for the translocation of the SR proteins from the cytoplasm to the nucleus and also for the recruitment of these factors from nuclear speckles (“splicing factor compartments”) to the sites of pre-mRNA synthesis (Bourgeois et al., 2004). Phosphorylation is also important for specific RNA recognition, since the high positive charge of un-phosphorylated RS domains masks the specificity of the RNP domains and enhances non-specific binding (Stamm, 2008; Tacke et al., 1997).

1.3.2 Heterogeneous ribonucleoproteins (hnRNP).

hnRNP proteins are located throughout the nucleus and some are extremely abundant, while others are present only in small amounts.

The hnRNP family is a class of diverse RNA-binding proteins that associate with nascent pre-mRNA. These factors remain associated with pre-mRNA until its processing is completed and with mRNAs during export from nucleus to cytoplasm (Izaurralde and Mattaj, 1995). At least 20 major hnRNP proteins exist in human cells, with nuclear weight ranging from 34 to 120 kDA (Dreyfuss et al., 1993).

The structure of hnRNP proteins is modular and consists of one or more RNA binding domain associated with an auxiliary domain often involved in protein-protein interactions (Dreyfuss et al., 1993). The auxiliary domain is usually rich in amino acid, such as glycine, proline. The structure of the members of the hnRNP family is considerably different.

The hnRNP A/B proteins contain two RNP domains at the N-terminus and a Gly-rich auxiliary domain at the carboxyl end.

HnRNP E1-E2 proteins contain three K Homology domains (KH) (Ostareck-Lederer et al., 1998).

The hnRNP H family members contain two (2H9) or three (H, H' and F) RNA recognition motifs (qRRMs) and one or two glycine rich auxiliary domains (Honore et al., 1995).

The hnRNP I (also known as the polypyrimidine-tract binding protein), L and M contain four RNP domains.

The hnRNP C proteins contain a single RNP domain at the N terminus.

hnRNP proteins show general RNA-binding specificity and individual proteins display preference for specific sequences that tend to coincide with sites of functional importance in pre-mRNA processing. However, hnRNP proteins generally do not bind to specific sites exclusively but recognize different RNA

sequences with different affinities. This RNA binding ability is further modulated by cooperative protein-protein interactions (Dreyfuss et al., 2002; Dreyfuss et al., 1993). hnRNP proteins frequently mediate splicing repression, particularly through binding to exonic splicing silencer elements or by sterical interference with other protein interaction (Cartegni et al., 2002). Nevertheless, depending on their pre-mRNA binding position, some hnRNPs can also associate with enhancer elements and help exon inclusion (Caputi and Zahler, 2002). The hnRNPA-B subfamily acts predominantly as repressors and the hnRNP F/H subfamily acts as activators as well as repressors (Talukdar et al., 2011).
Although many of the hnRNPs are localized in the nucleus, some of the hnRNPs shuttle between the nucleus and cytoplasm, which suggest a role in nuclear export and in other cytoplasmic processes (Martinez-Contreras et al., 2007; Pinol-Roma and Dreyfuss, 1992).

1.4 Alternative Splicing.

Genetic information flows between DNA to RNA after which it is translated into protein. Alternative splicing plays a central role between transcription and translation.

Alternative splicing (AS) is an important mechanism that allows the production of a large number of mRNA and proteins from a single gene (Maniatis and Tasic, 2002). Alternative splicing in flies is crucial for sex determination; where the pre-mRNA of DSCAM (Down syndrome cell adhesion molecule) gene can produce more than 38000 different mature transcripts by alternative splicing.
Alternative splicing affects 95% of mammalian genes, and has been implicated in other regulatory process such as chromatin modification and signal transduction. There are different ways in which alternative splicing can occur; including exon skipping, different splice site choice and intron retention (IVs) (Cartegni et al., 2002; Ladd and Cooper, 2002). In addition the 5’-terminal exons of an mRNA can

be switched through the use of alternative promoters and alternative splicing. Similarly, the 3'-terminal exons can be switched by combining alternative splicing with alternative polyadenilation sites. These individual patterns can be combined in a single transcription unit to produce a complex array of splice isoforms. Moreover, changes in alternative splicing can modulate transcript expression levels subjecting mRNAs to nonsense-mediated decay (NMD) by creating a stop codon within the coding sequence or by altering the structure of the gene product by inserting, or deleting, protein parts (Faustino and Cooper, 2003) (Figure 1.11). Splice site strength and binding of splice factors to enhancer and silencer elements play an important role in alternative splicing.

Alternative splicing is not only regulated by splicing factors but also by other processes that involve the transcription machinery. In addition to the splicing process there are the capping and polyadenylation processes that together with the splicing process modify the mRNA respectively at the 5' ends and the 3' ends.

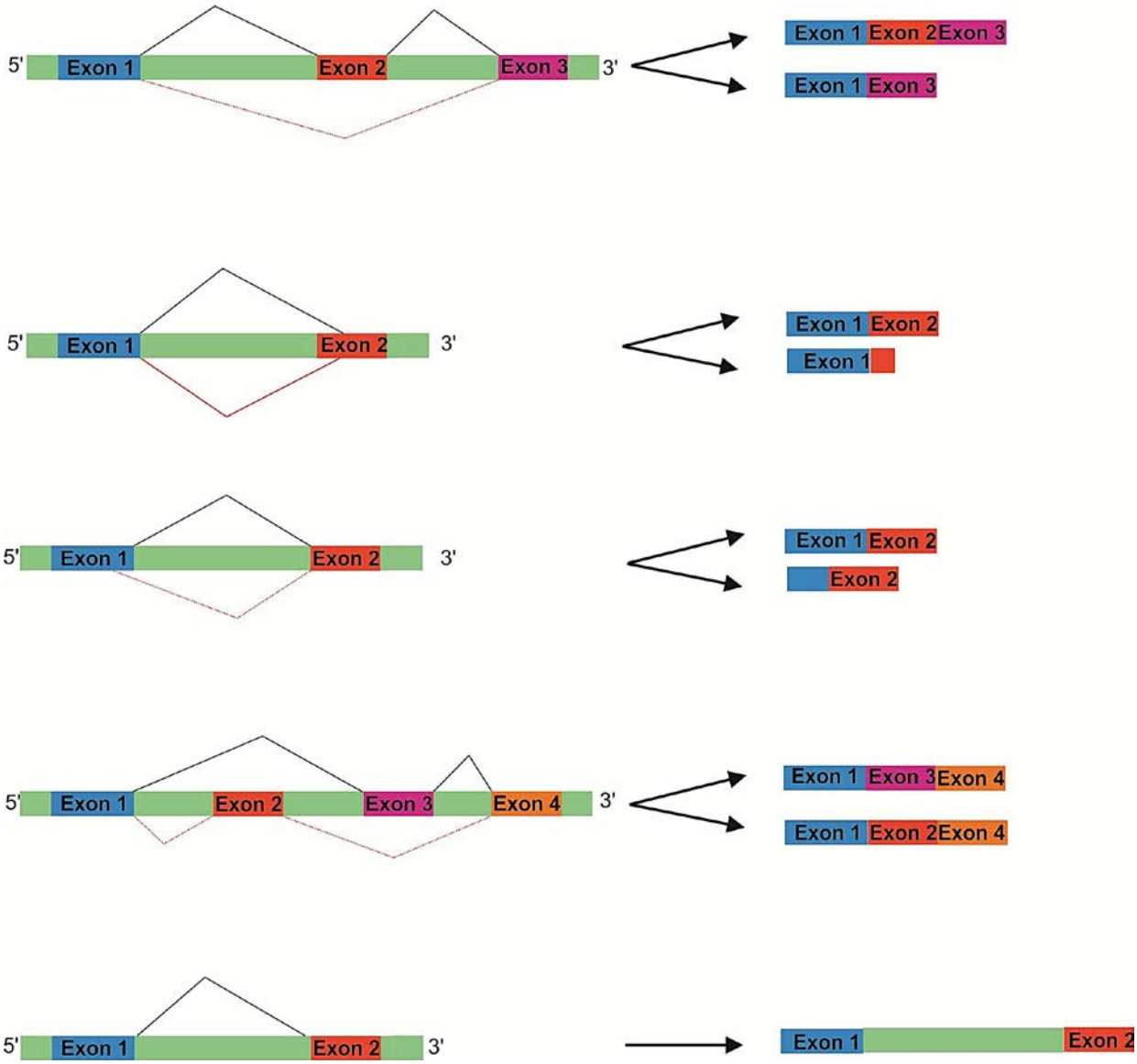


Figure 1.11: Models of alternative splicing.

The figure shows different models of alternative splicing. In each case, one alternative splicing event is indicated in black and the other in red. In the last example the alternative pathway corresponds to constitutive splicing.

1.4.1 Splicing and Disease.

Mutations in regulatory sequences that affect alternative splicing are known to cause human hereditary disease and cancer (Chakarova et al., 2002).
Mutations can disrupt a binding site for an enhancer or silencer or can create a new binding site. This is then associated with the formation of different mRNAs and different encoded proteins. In addition nucleotide changes that may create a premature stop codon (nonsense mutation) or an amino acid change (missense mutation), can actually be acting at the pre mRNA level by affecting the elements that are involved on the splicing process. The same is true for silent sequence variants. These do not alter the encoded amino acid of a protein but can disrupt a crucial splicing regulatory sequence.
Mutations at the 3' and 5' splice site are the most frequently reported splicing mutations at *cis* splicing regulatory sequences. But it is also possible to have mutations in *trans* acting factors.
Alteration of splicing has been shown to contribute to the development of several diseases such as: cystic fibrosis (CF), hereditary non polyposis colorectal cancer (MLH1), neurofibromatosis (NF1), spinal muscular atrophy (SMA), front temporal dementia with Parkinson, retinitis pigmentosa, chronic lymphocytic leukaemia (CLL) and many others. These mutations are either an alteration of the *cis* or *trans* acting elements that cause inclusion or exclusion of exons.

In the following section are brief explanations of a few well studied diseases caused by an alteration of the process of splicing, where the nucleotide change can cause exclusion of the specific exon in the final transcript.

Frontotemporal dementia and Parkinsonism linked to chromosome 17 (FTDP-17) is an autosomal dominant disorder caused by mutations in the *MAPT* gene which encodes for the tau proteins (McKie et al., 2001).
The tau protein is essential for microtubule assembly. In particular the tau protein isoform with 4R (4 repeat domain). The presence of this 4R is determined by the inclusion of exon 10 which encodes the last four microtubule binding domains. The normal ratio between 4R and 3R (where there is exclusion of exon 10) is one; mutations in FTDP17 alter this ratio. The ratio between 4R and 3R is maintained by the presence of different ESEs that facilitate the inclusion of exon 10. The majority of FTDP-17 mutations affecting the inclusion of exon 10 do so due to the presence of intronic changes adjacent to the 3' end of exon 10. The increase of inclusion of exon 10 creates an accumulation of the 4R isoforms in the brain of the patient (Makarova et al., 2002; Vithana et al., 2001).
However not all mutations in FTDP-17 increase inclusion of exon 10, in fact one mutation, a 3 nucleotide deletion situated inside the 5' ESE (Δ K280) gives skipping of exon 10 (Wang et al., 1995).

Cystic fibrosis (CF) is an autosomal recessive disorder caused by mutations in the cystic fibrosis trans membrane conductance regulator (CFTR) gene. Alternative splicing of CFTR pre-mRNA gives rise to a number of CFTR isoforms. The Cystic Fibrosis Mutation Database lists 226 mutations (Gonzalez-Santos et al., 2002) that have been linked with an alteration of the splicing process.
These mutations can also either alter a *cis* acting element or *trans* splicing factor. An example of an alteration of a cis acting element are two polymorphisms in the CFTR gene located at the 3' end of intron 8 which affects splicing of exon 9.
One variant contains 5, 7 or 9 uridines in the polypyrimidine tract of intron 8, the other variant has a polymorphic poly (UG) locus situated immediately upstream of the (U) tract. An individual with these alterations expresses a small fraction of CFTR mRNA lacking exon 9 and expresses a non-functional protein (Wyatt et al., 1992).

Retinitis pigmentosa (RP) is an autosomal dominant disease and it is characterized by sequence alterations of different genes*: PRPF31, HPRP3 and PRPC8* (Cartegni and Krainer, 2002; Kashima and Manley, 2003; Terns and Terns, 2001).

All these three genes are involved in the function of the U4/U5-U6 spliceosome complex. Prp31 is an important splicing factor and is part of the U4/U6snRNP-associated proteins which promotes the interaction between U4/U6 and U5snRNP. Mutations in Prp31 create accumulation of the U1 and U2 complex, preventing the interaction between the U4/U6 and the U5 snRNP (Kashima et al., 2007).

Mutations in *HPRP3* are missense mutations in exon 11. *HPRP3* is a component of U4/U6 complex. The effect of the HPRP2 mutations on the function of the protein is still unknown (Danglot et al., 1995; Gutmann et al., 1995).

PRPC8, encodes for the protein (PRP8) that is a core component of U5snRNP. Mutations in PRPC8 have been associated with a severe form of the disease (Zhou et al., 2011).

Spinal muscular Atrophy (SMA) is the second most common autosomal recessive disorder. SMA is caused by the loss of the *SMN1* gene that encodes the SMN protein, which regulates snRNP assembly (Park et al., 1998). In humans there is another copy of this gene termed SMN2, which differs from SMN1 by only 5 nucleotides but produces a truncated protein due to skipping of exon 7. Different studies have looked at the mechanism of the splicing defect in SMN2. A silent mutation +6 C>T in exon 7 of SMN2 has been found, which causes skipping of exon 7. Two models have been proposed to explain the mechanism of splicing behind the exclusion of exon 7. The first model proposes disruption of a binding site for SRSF1 in the presence of the variant. (Quesada et al., 2012). The second model highlights the presence of a binding site for hnRNPA1 in the presence of the sequence variant +6 C>T. Binding of hnRNPA1 blocks the use of the 3’ splice site (Best et al., 2013; Grosso et al., 2008). The presences of these truncated non-functional proteins are useful when considering the use of chemical

compounds that can increase inclusion of exon 7 as potential therapy for the disorder. Chemical compounds and oligonucleotides for this aim have been developed (Introduction 1.4.2).

Neurofibromatosis type I (NF-1) is an autosomal dominant disorder. NF-1 disease is characterized by loss of the neurofibromin protein, which is similar to the Ras GTPase activating protein (GAP) family. The NF-1 gene is composed of 60 exons; studies of alternative splicing on NF-1 detected the presence of different sites of alternative splicing in particular three different alternative splicing events involving exon 9br (predominately expressed in brain), exon 23a (present with a different concentration in all tissue) and exon 48a (specifically present in muscle) (Chao et al., 2003; Manoharan, 1999). The mechanisms of action of these exons are unknown, with the exception of the splicing isoform 23a, which is regulated by Hu proteins, TIA-1 and CEFL splicing factors (Dunckley et al., 1998).

Another important alternative splicing event that has been associated with NF-1 disease is skipping of exons 29 and 30. Three isoforms have been analysed: skipping of exon 29, 30 or skipping of both exons. All three isoforms alter the process of splicing introducing a stop codon (Raponi et al., 2009).

In disease it is also possible to have a mutation in the components of the splicing process. Examples are the mutations in SF3B1. SF3B1 has been demonstrated to be the most frequently mutated gene in chronic lymphocytic leukaemia (CLL) and in myelodysplastic syndrome (MDS). SF3B1 has an important role in the splicing process; as it is an essential component of U2 snRNP (Wan and Wu, 2013).
Not only mutations but also alteration in the level of splicing regulatory proteins can cause disease and in particular cancer (Best et al., 2013; Vaughn et al., 1996). One example of cancer related with alternative splicing is breast cancer.

1.4.2 Oligonucleotides and therapy.

An innovative strategy to correct alternative splicing defects involves using small molecules that target transacting factors and/or their binding sites and thus restore normal splicing.

This new therapeutic approach includes the use of modified antisense RNA oligonucleotides (AONs), modified bifunctional RNA oligonucleotides, modified U1 small nuclear RNA targeting sequences situated downstream of the 5' splice site and the use of trans splicing strategy (SMART, spliceosome mediated RNA trans splicing) that creates a chimeric RNA where the aberrant splicing region is replaced with a healthy region. An example of SMART was demonstrated by Chao et al to correct the haemophilia A phenotype of FVIII (Chao et al., 2003).

Antisense oligonucleotides (AONs) are synthesized from DNA and are able to induce RNase H-mediated digestion of target RNA. The oligonucleotides used to modify the splicing process do not require the activity of RNase H, because it is important to preserve mRNA for translation. Therefore the oligonucleotides used are chemically modified to inhibit RNase H activity.

The most used oligonucleotide modifications are those that have a modification at the 2'position of the ribose, that inhibits RNAseH activity (Aartsma-Rus, 2010) and are: the 2'o-methyl (2'-O-Me) (Monaco et al., 1988; Muntoni and Wood, 2011), the 2'O-methoxyethyl (MOE), (Hua et al., 2010) the 2' O-aminopropyl and bridged based where the 2'-O and 4' positions are connected through a methylene group (Griffey et al., 1996; Teplova et al., 1999) (Figure 1.12).

In addition to modification to the ribose ring it is possible to have an oligonucleotide where the backbone phosphodiester is modified with a sulphur atom (the phosphorothioate Ps).

Oligonucleotides with a combination of modification in a ribose and backbone are often used. An example of this is the 2'OMePs/MOEPs and morpholino phospohorodiamidate (PMO) oligonucleotides. These oligonucleotides have been shown to alter the process of splicing.

A further two types of oligonucleotides that have been used in pre-mRNA splicing alteration are the peptide nucleic acid (PNA) oligonucleotides and the locked nucleic acid (LNA) oligonucleotides. PNA has a nucleic base binding to the polyamide backbone of N (2aminoethyl) glycine units (Egholm et al., 1993) whereas the LNA has 2'O-4'C methylene binding to bicyclic ribofluranosyl nucleosides (Petersen et al., 2000).

Duchenne Muscular Dystrophy (DMD), *Bcl-x* in lung metastases and *SMN* are good examples of disease where the modified oligonucleotides have been used to try to correct the disease.

DMD is an X linked disease, caused by a mutation in the *DMD* gene. This disease affects around 1 in 3500 live male births and it is characterized by disruption of the *DMD* reading frame resulting in the production of non-functional dystrophin protein (Taylor et al., 1999).
The AONs used for DMD disease have been shown to correct the aberrant dystrophin transcription through exclusion of exon 44 (Liu et al., 2001). In this case it is possible to generate a shorter but functional protein (Goyenvalle et al., 2009). The use of these AONs can lift DMD to the milder Becker Muscular Dystrophy (BMD).
The first oligonucleotide used to correct DMD in the dystrophic *mdx* mouse was the 2-OmethylPhosphorothioate RNA (2'-OMePS). The 2'-OMePS are able to target the 3'splice site and the 5'splice site of exon 23 causing skipping of exon 23. After three weeks of oligonucleotide injection, skipping of exon 23 was induced and restored the expression of dystrophin in all skeletal muscles. This oligonucleotide recreated high levels of dystrophin, but unfortunately there was no high expression of dystrophin in cardiac muscle (Mann et al., 2001).
For this reason the oligonucleotide was redesigned with the use of PMO modifications. The new oligonucleotide was injected into the muscle of the *mdx* mouse and in Canine X-linked muscular dystrophy (CXMD) (Moulton and Moulton, 2010).

Skipping of exon 51 in *DMD* has already been used in clinical trials. In this case the oligonucleotide 2'-OMePs (PRO051) was injected directly into the muscle of a patient with DMD disease (van Deutekom et al., 2007). This oligonucleotide induced the expression of the dystrophin protein to around 16%. In addition the muscular injection of PMO (AVI-4658) stimulated dystrophyn exon 51 skipping in humans, causing an increase of the expression of the protein of around 27% which is an adequate amount of dystrophin for muscular function.

Antisense oligonucleotides have also been used also to increment the inclusion of an exon in cases when that exon normally would have been skipped due to the presence of a mutation. The prime example of this approach has been used to treat SMA (Spinal muscular atrophy) (Skordis et al., 2003a).

As stated before SMA is an autosomal recessive neurodegenerative disorder of motor neurons. The gene involved is called *SMN1*. Humans have a second copy of the gene which differs from the original SMN by only 5 single bases in the coding sequence. This 2nd human specific version is called SMN2.

The substitution C>T in exon 7 of SMN2 causes skipping of this exon and the protein product is not functional. With the C>T substitution the binding site for the splicing factor SRSF1 is disrupted (Cartegni and Krainer, 2002). In addition this substitution in SMN2 allows binding of hnRNPA1 which negatively regulates the use of splice sites (Cartegni et al., 2006).

A valid treatment for SMA is the use of antisense oligonucleotides (ASOs), with the aim of inducing inclusion of exon 7 in SMN2 and restores some functional SMN protein. Different oligonucleotides have been used to do this, 2'-O-MePs, 2'O-Me, 2'-O- MOE and PNA (Cartegni and Krainer, 2003; Owen et al., 2011a; Skordis et al., 2003b).

The difficulty with this therapy is the ability of this therapy to be delivered. For SMA patients the oligonucleotide would have to cross the blood brain barrier (BBB). Experiments in mouse have shown that the AOs can reach the central nervous system through intrathecal injection (Passini et al., 2011).

Therapeutic approaches using modified oligonucleotides have been also used to alter the ratio of alternatively spliced Bcl-x pre-mRNA. The Bcl-x pre-mRNA can generate two transcripts coding for either a long Bcl-xL or for a short Bcl-xS. The long form (Bcl-xL) has been shown to inhibit apoptosis; and the short form (Bcl-xS) has a pro-apoptotic function (Boise et al., 1993; Minn et al., 1996). In this case the use of a 2'-O- MOE oligonucleotide increases the Bcl-xS form through the inhibition of the 5'splice site in exon 2 of Bcl-xL RNA thereby restoring the normal splicing pattern (Taylor et al., 1999).

Another therapeutic approach is the use of molecules that promote the inclusion of the exon (Cartegni and Krainer, 2003). A construct termed ESSENCE (exon specific splicing enhancement by small chimeric effectors) was created where an antisense peptide nucleic acid (PNA) is fused with ten RS repeats at the 3'end. This construct bound exon 7 of *SMN2*. Adding this construct in an *in vitro* splicing experiment showed an increase of exon 7 inclusion (Cartegni and Krainer, 2003).
ESSENCE has also been tested in a BRCA1 mini-gene carrying a mutation at position 6 of exon 18, this mutation is a nonsense mutation and causes skipping of the exon, due to disruption of a SRSF1 binding site (Liu et al., 2001). The use of these synthetic molecules induced the inclusion of the exon 18.

A similar approach has been used by Skordis using a bifunctional 2'O-methyl oligonucleotide complementary to the SMN2 exon 7 and with a GGA repeat tail, which has a positive effect on the inclusion of the exon 7 (Skordis et al., 2003b).
This demonstrated that the presence of SR protein close to the exon splicing enhancer is sufficient to promote inclusion of the exon.

Long term efficiency of therapy is also an issue with bifunctional oligonucleotides. This could be possible using viral vectors. In fact it has been demonstrated that the use of small nuclear RNA (such as U7 and U1) carrying an

antisense sequence have the advantage of long term efficiency and eliminated the need of repeated injections of oligonucleotide.

A universal U7 snRNA carrying sequence complementary to the exon and a tail for hnRNPA1/A2 (repressor of the process of splicing) was developed. The U7snRNA with tail is able to induce skipping of a particular exon having an opposite effect of U7snRNA without tail. The U7snRNA with the tail has been used in DMD disease giving efficient skipping of the exon 51 (Goyenvalle et al., 2009).

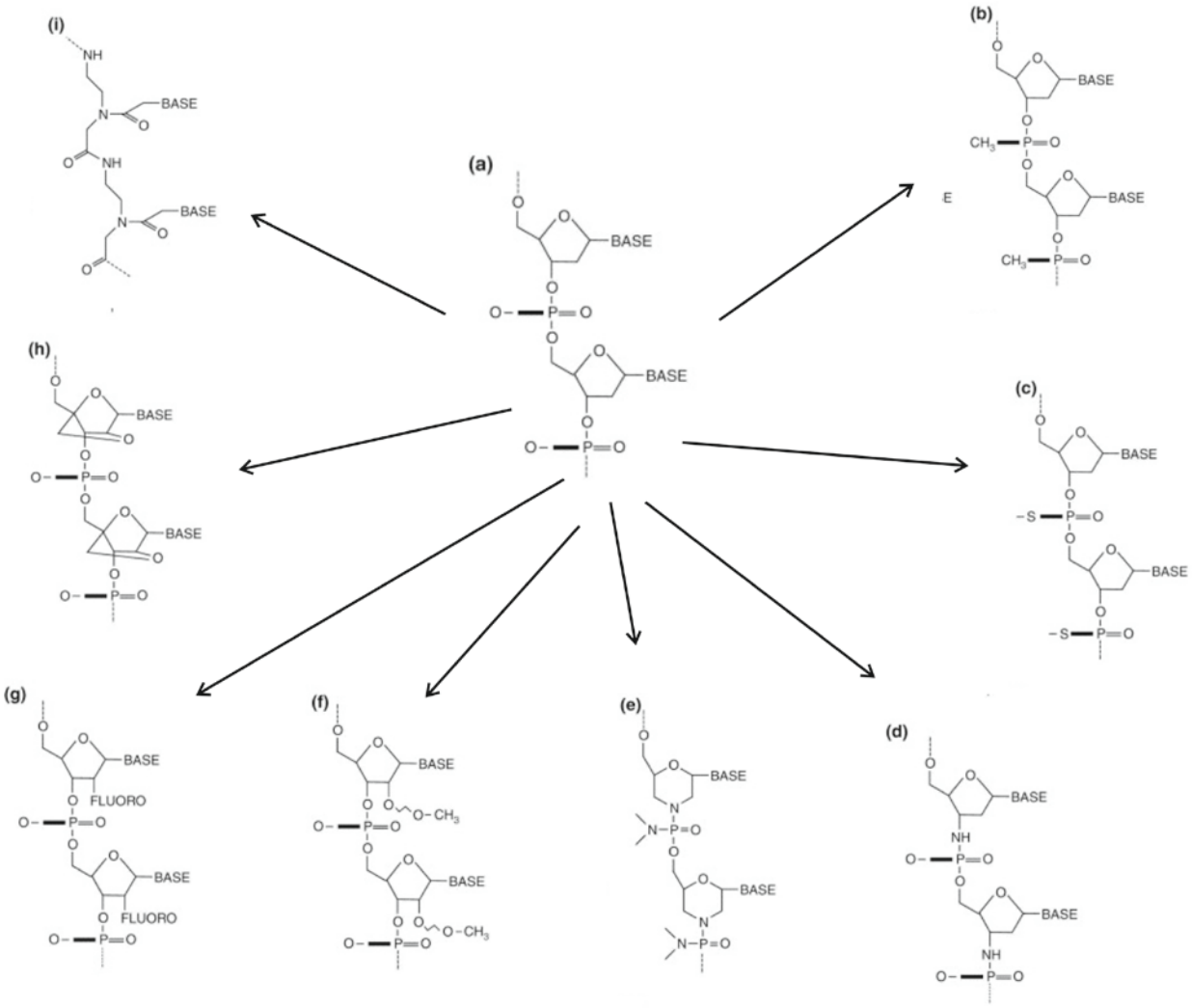


Figure 1.12: Chemical structure of RNA oligonucleotide.

A.Unmodified phosphodiester oligonucleotide. **B**. 2'-O-Methyl (2'-O-Me_methyl group at position 2' of the ribose ring) **C**. Peptide Nucleic Acid (PNA). **D**. Morpholino (PMO, morpholino group linked through phosphorodiamidate linkage). **E**. N3-P5'Phosphoriamidate (contains N3-P5' phosphoramidate linkage in place of a bridging oxygen atom). **F**.Methylphosphosphonate (methyl group inplace of a non bridging oxygen on each backbone phosphorous). **G**. Phosphorothiodate (ps-Sulphur in place of a non-bridging oxygen atom on each backbone phosphrous. **H** Locked Nulcleic Acid (LNA - joined 2'-*O*,4'-C-methylene linked ribose ring). **I** 2′-O-methoxyethyl (MOE - 2' Methoxyethyl addition on to the ribose ring). Figure adapted from (Dias and Stein, 2002; Estibeiro and Godfray, 2001).

1.5 BRCA1.

Breast cancer is one of the most common diseases that affect women in the world. The risk of Breast Cancer is associated with different factors: mutations in one of the tumour suppressor *BRCA1* and *BRCA2* genes, other lower penetrance genes and environmental/lifestyle factors (tobacco, radiation and xenoestrogen).

The *BRCA1* gene is located on chromosome 17q21.31 and was first identified on the basis of its linkage to early onset breast and ovarian cancer in women in the 1990. In 1994 the *BRCA1* gene was cloned for the first time (Miki et al., 1994).

Inherited mutations of *BRCA1* are responsible of 40-45% of hereditary breast cancer and these mutations are responsible for 2-3% of all breast cancer, since the *BRCA1* gene is rarely mutated in sporadic breast cancer. *BRCA1* has been classified as a "tumour suppressor gene". Mutations in this gene cause the loss or reduction of the BRCA1 protein, and cells can progress to cancer. Women with mutations in *BRCA1* can have up to 80% probability of developing cancer during their life (Neuhausen and Marshall, 1994). BRCA1 has roles in: cell cycle progression, DNA repair, DNA damage responsive cell cycle check-point (allows the cell to repair the damage before mitosis), regulation of specific transcriptional pathways and apoptosis. These functions involve different portions of the BRCA protein. The *BRCA1* gene contains 24 exons (22 coding and 2 non-coding), encodes a protein of 1,863 amino acids, and contains three distinct protein interaction regions: the ring domain, the RAD51 interaction domain and the BRCT domain (Figure 1.13) (Miki et al., 1994; Pavlicek et al., 2004).

Figure 1.13: A schematic diagram of BRCA1.

The figure shows the BRCA1 polypeptide which has the N terminal Ring Motif (red), the Nuclear localization signal (NLS, violet) and the two C terminal BRCT domains (blue).

The N-terminal Ring finger domain interacts with different proteins, including formation of heterodimers with BARD1 (BRCA1-associated ring domain protein) (Hashizume et al., 2001). BRCA1 and BARD1 act as tumour suppressors preventing cells from growing and dividing too rapidly or in an uncontrolled way. Another function of BARD1 is to maintain the BRCA1 protein in the nucleus.
The BRCA1 C-terminal has a coil domain together with the tandem repeat globular domains termed BRCT (Koonin et al., 1996). These domains are critical in DNA repair (O'Donovan and Livingston, 2010).
Both domains, C-terminal and N-terminal, are critical for DNA repair and DNA damage response signal.
BRCA1 is implicated in all phases of the cell cycle and in the regulation of orderly events during cell cycle progression (Deng, 2006b).
Consequently loss-of-function mutations of *BRCA1* result in an increase of apoptosis, defective DNA damage repair, defective G2/M cell cycle checkpoint, chromosome damage, and aneuploidy (Venkitaraman, 2002). *BRCA1* cancer-predisposing truncating and missense mutations, are commonly found within the two C-terminal BARCT motifs and to a lesser extent in the critical Zn2+ binding residues within the N-terminal RING finger, indicating that these regions are critical for tumour suppressor function (Friedman et al., 1994). It is proposed that mutations in *BRCA1* do not directly result in tumour formation, but instead cause genetic instability thereby exposing the cells to a high risk of malignant transformation (Kinzler and Vogelstein, 1997).

1.5.1 The function of BRCA1 in the cell cycle.

The cell cycle consists of four different stages: Gap 1 (G1), that is the gap before DNA replication; synthesis (S) during which DNA is replicated; Gap 2 (G2), the gap phase following the DNA replication, and mitosis (M) in which chromosomes are segregated and the cell division take place (Hartwell and Weinert, 1989).

Cell cycle progression is regulated by the activity of protein kinase complexes consisting of a cyclin and a cyclin-dependant kinase (Cdk). As implied in the name, Cdks are tightly regulated through their association with cyclin proteins.
A large number of proteins are responsible for maintenance of genome integrity and the functions of these proteins include DNA synthesis, DNA damage repair, and regulation of cell cycle checkpoints (Hartwell and Weinert, 1989).
The cycle checkpoints are used to maintain the integrity of the genome, and respond to DNA damage. These include the G1/S-phase checkpoint, the S-phase (or intra-S) checkpoint, and the G2/M checkpoint. BRCA1 plays an important role in coordination of the cell cycle (Deng, 2006a; Ruffner and Verma, 1997; Vaughn et al., 1996).
BRCA1 was shown to interact with hypo-phosphorylated Retinoblastoma (RB) to inhibit cell proliferation and induce G1-arrest (Deng, 2006a). When RB is hypo-phosphorylated it interacts with E2F to prevent transcription of downstream genes required for S-phase progression of the cell cycle and consequently there is inhibition of cell proliferation (Aprelikova et al., 1999). Therefore BRCA1 maintains RB in a hypo-phosphorylated state to have growth arrest. BRCA1 C terminal domain (BRCT) forms a complex with RB binding proteins and histone deacetylase 1 and 2 (HDAC1 and HDAC2) and this complex with RB and histone deacetylase is thought to suppress transcription of E2F (Yarden and Brody, 1999).
It has been shown that BRCA1 forms a heterodimer with BARD1 (BRCA1 associated RING domain protein), and the heterodimer is required to maintain the stability of BRCA1 and BARD1. The BRCA1-BARD1 complex has been shown to be required for ATM-mediated phosphorylation of Chk2 and p53 at Ser15 which is required for G1/S-phase arrest via transcriptional induction of p21(Fabbro et al., 2004).
In summary BRCA1 induces the arrest of the G1/S cell cycle through different protein interactions including RB, p21, p53, ATM and BARD1.

BRCA1 also controls different genes associated with the regulation of the G2/M checkpoint.

BRCA1 represses Cyclin B, the cyclin responsible for activation of the cdc2 kinase which allows passage between the G2 to M phase (Deng, 2006b). BRCA1 also regulates the expression of Wee1 Kinase, which inhibits the phosphorylation of cdc2 and consequently inhibits the cyclinB-cdc2. BRCA1 regulate the 14-3-3 chaperon protein, which targets the cdc25C and sequesters this protein into the cytoplasm, where the cdc25C cannot activate cyclinB-cdc2 (Hutchins and Clarke, 2004; Muslin and Xing, 2000).

It has been shown that the protein GADD45 controls the cell cycle G2-M checkpoint through the inhibition of the cd2 and consequently regulates complex cyclin B and CDC2 (Wang et al., 1999), BRCA1 can regulate GADD45 and regulate G2/M cell cycle arrest.

BRCA1 induce the arrest of the G1/S cell cycle and the G2/M cell cycle, a schematic representation of the model is showed in Figure 1.14.

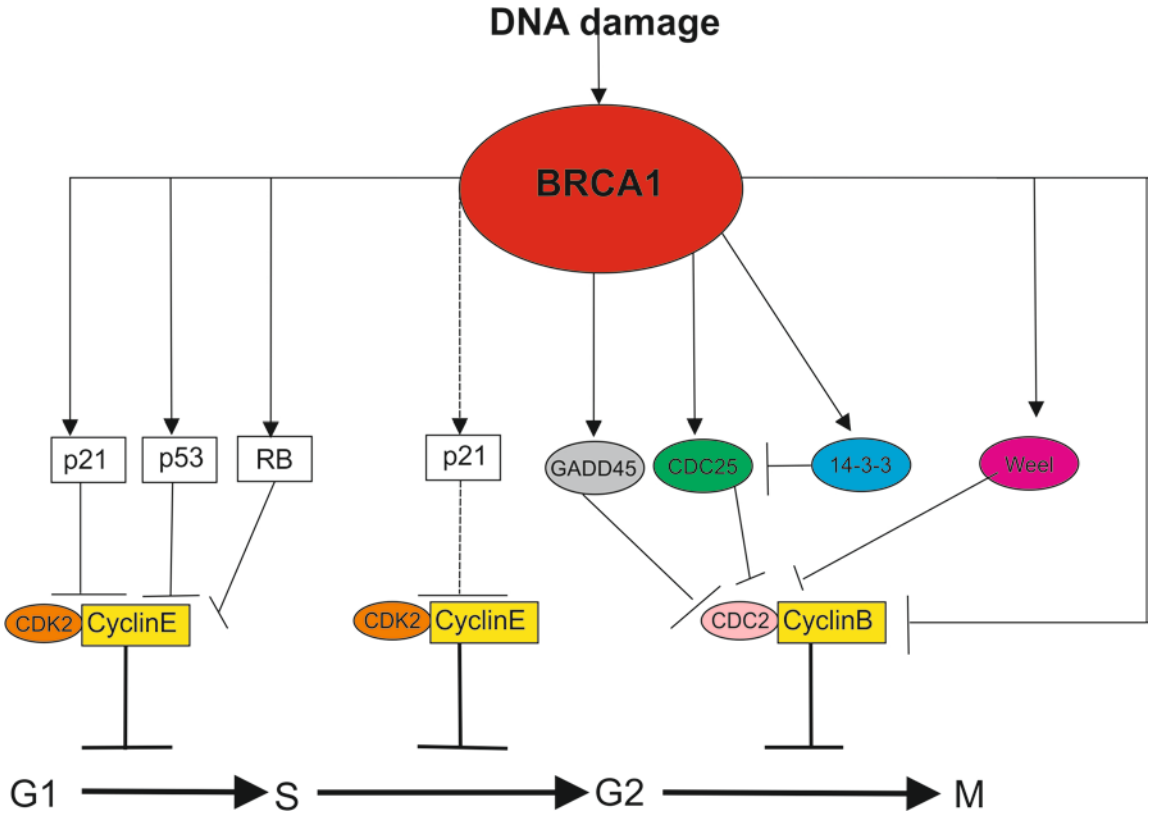


Figure 1.14: Schematic representation of BRCA1 transcriptional regulation and cell cycle control in response to DNA damage.

BRCA1 through different mechanism can stimulate the transcription of p21, p53 and RB leading to G1/S arrest. It stimulates the transcription of different G2/M checkpoint regulatory genes: GADD45, CDC25, 14-3-3 and Weel, which inhibit the cdc2 Cyclin B (Mullan et al., 2006).

1.5.2 BRCA1 regulation and DNA repair.

Accurate repair of damaged DNA is essential for maintenance of genomic integrity.

BRCA1 plays an important role in maintaining genomic integrity by protecting cells from DNA double strand breaks (DSB). DSBs are generated endogenously, through consequence for example of collapsed replication fork, DNA replication and DNA repair, and by exogenous agents such as ionizing radiation (IR) and other genotoxic compounds (Friedberg, 2008).

The genomic instability resulting from persistent or incorrectly repaired DSBs can lead to activation of oncogenes, inactivation of tumour-suppressor genes, or loss of heterozygosis (LOH) at specific loci (Raynard et al., 2008).

In response to DNA damage agents there are three different pathways involved in the repair of DNA double strand breaks: **Non-homologous end joining (NHEJ)**, **single-strand annealing (SSA)**, and **homology directed repair (HDR).**

Non-homologous end joining (NHEJ) requires three enzymes to repair the DNA double strand breaks; in particular it requires a nuclease to identify the DNA damage; a polymerase to fill the gap with new DNA and the ligase to restore DNA integrity. This pathway has an opposite effect compared to homologous repair (HR) as the break ends are directly ligated without the use of a homologous template. The use of NHEJ pathway it is often accompanied by modification of the region around the broken site (Kass et al., 2013). The NHEJ pathway is favoured by cells in G0/G1, where there is no homologous template available, and involves the direct re-ligation of the broken DNA ends; this regulation is accomplished by the cycling–dependent kinase Cdk1.

The NHEJ is considered an inherently error prone process due to the loss of the terminal bases that are usually removed to allow ligation to occur in an efficient manner. NHEJ does not rely on extensive sequence recognition and thus has the capacity to ligate DNA ends from non-homologous chromosomes resulting in increased chromosomal aberrations (Lieber, 2010).

Single-strand annealing (SSA), is also an inherently error prone process, and promotes DNA double strand break repair by annealing short regions of sequence homology to the DNA break site (West et al., 2000). The process is quite simple, when there is the DNA damage single stranded regions are created adjacent to the break, in this way the complementary sequences can anneal to each other. This intermediate annealing can be processed by digesting the single strand tail and filling the gap.

Homology directed repair (HDR) is the most accurate of the three repair mechanisms, predominates over the other DNA double strand breaks repair pathways during the S and G2 phases of the cell cycle in which sister chromatids are readily available (Kass et al., 2013)
The homologous recombination pathway requires an identical sequence, which is used as template for repairing the DNA break. The central mechanism of HDR is the use of the recombined enzyme Rad51 (West et al., 2000). Rad51 forms a nucleoprotein filament on the recently processed ssDNA and then interacts with the sister chromatid, which acts as a template for DNA synthesis to repair the DSB (Li and Heyer, 2008).

BRCA1 plays essential roles also in homology repair and nucleotide excision repair (NER) and is able to mediate these functions through interaction with components of the DNA repair machinery, and through the regulation of the expression of genes involved in the DNA damage repair pathways (Deng and Wang, 2003). Different studies suggest that BRCA1 has a role in damage repair by acting as a link between the elements that cause damage and the that repair the DNA damage (Boulton, 2006).

BRCA1 has a critical role in responding to DNA double strand breaks through its function in homology repair. BRCA1 recruits BRCA2 through a mediator protein PalB2, the interaction between *BRCA1* and PalB2 occurs via the coiled-coil

domain (situated in the C-terminal of BRCA1). Mutations found in BRCA1 coiled-coil domain abolish the PalB2 binding activity consequently compromise the HR activity, these mutations have been found in BRCA1 tumors (O'Donovan and Livingston, 2010; Sy et al., 2009). The binding of BRCA1 and BRCA2 facilitates Rad51 filament formation on the ssDNA (Zhang and Powell, 2005). Rad51 catalyzes the invasion of the homologous sequence on the sister chromatid, which is then used as template for accurate repair of the broken DNA ends. A recent study has shown the importance of BRCA1 in DNA damage repair using the HDR pathway in mouse (Kass et al., 2013). During HDR repair it has been suggested that ATM is important in the generation of Replication Protein A (RPA)-coated ssDNA which is an essential early intermediate in HDR process. HDR is able to recruit Rad 51 to repair the DNA (Kass et al., 2013). BRCA1 co-localizes with Rad50, a member of the MRN complex, following the induction of DNA damage; Mre11 encodes nuclease activity which resects flush ends of DSBs to generate ssDNA tracts. BRCA1 binds DNA directly and inhibits this Mre11 activity regulating the length and the persistence of ssDNA generation at sites of DNA damage (Haber, 1998; Wang et al., 2000). As ssDNA is a substrate for DNA repair by HR, it appears that BRCA1 might play an essential role in HR-mediated repair of DSBs through its inactivation of Mre11. In addition to it somewhat unclear roles in DSB repair, BRCA1 has also been found to be a constituent member of a large nuclear protein complex named the BRCA1-associated surveillance complex (BASC). This complex contains the MRN complex, DNA-mismatch repair proteins MSH2, MLH1 and MSH6, DNA helicase BLM, ATM, RFC and PCNA (Wang et al., 2000). Many of these proteins are involved in the sensing and repair of abnormal DNA structures, and have been linked with the repair of replication-associated DNA damage (Haber, 1998).

Taken together this evidence suggests that BRCA1 might function as a coordinator of multiple processes required for the maintenance of genome integrity during the process of DNA replication and DNA-replication associated repair (Gudmundsdottir and Ashworth, 2006).

1.5.3 BRCA1 splicing isoforms.

BRCA1 gene produces different splicing isoforms (Table 1.1) that have different expression patterns in different tissues (Orban and Olah, 2003). There is evidence that more than thirty mRNA splicing isoforms are associated with the *BRCA1* gene, most of which maintain the original open reading frame and consequently have the possibility to code for a functional protein, the most common ones are represented in the Table 1.1.

Five m-RNA isoforms, full length, Δ(9,10), Δ(11q), Δ(9,10,11q), and the Δ11 show alternative splicing between exon 9 and exon 11, are expressed in a variety of tissues under different conditions, and are called the predominant splicing variants (Lu et al., 1996; Orban and Olah, 2001; Xu et al., 1997).

In particular exon 11 has three functional splice sites: the exon 11 3' splice site (at intron 10/exon 11 junction), the exon 11 5' splice site (at the exon 11/intron 11 junction), and an additional 5' splice site (11q 5' splice site), which is situated 117 nucleotides downstream from the exon 11 3'splice site. The different use of this three splicing isoform determinate the production of the three splicing isoform such as the full isoform, the Δ11q and the Δ11 isoform (Figure 1.15).

Table 1.1: The known BRCA1 splice isoforms in humans.

Name of variants	Keep the open reading frame	Tissues
Full length BRCA1	Yes	Breast, Ovary Testis,Thymus
with exon 1a	Yes	Breast, Ovary Testis,Thymus
with exon 1b	Yes	Placenta
Δ(2-10)	Yes	Breast, Lymphocytes
Δ(5)	Yes	Breast, Lymphocytes, Ovary
Δ(5q,6)	Yes	Breast, Lymphocytes
Δ(9,10)	Yes	Breast, Lymphocytes, Ovary
Δ(9-10,11q)	Yes	Breast, Lymphocytes, Ovary
Δ(9-10,11)	Yes	Breast, Lymphocytes
Δ(11q)	Yes	Breast, Lymphocytes, Ovary
Δ(11)	Yes	Ovary, Thyroid
Δ(14-17)	Yes	Breast, Lymphocytes
Δ(14-18)	Yes	Breast, Lymphocytes
Δ(3)	Yes	Breast, Lymphocytes
Δ(3,5q)	Yes	Lymphocytes
Δ(6,7)	Yes	Breast, Lymphocytes, Ovary
Δ(9)	Yes	Lymphocytes
Δ(15-17)	Yes	Breast, Lymphocytes

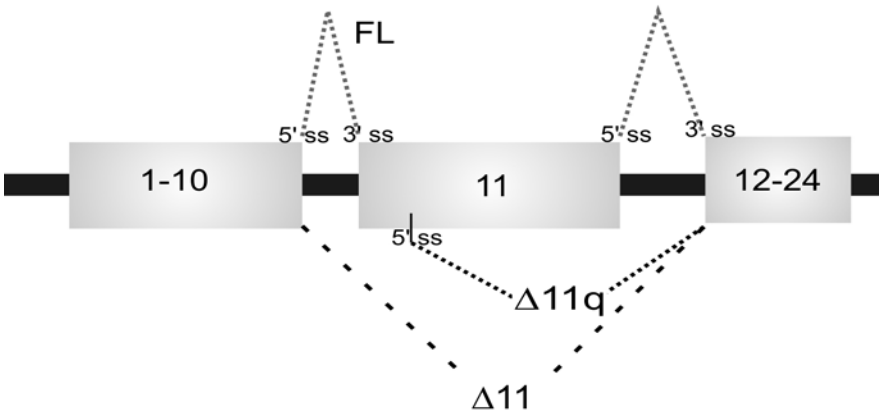


Figure 1.15 Schematic representation of splice site in *BRCA1* exon 11.

The grey boxes represent exons; the black lines represent introns and the dotted lines represent the splicing outcome according with the use of exon 11 splice sites. The splice sites of exon 11 are represented in figure with the abbreviation ss.

The full length isoform includes all the exons, from exon 1 to exon 22, and is predicted to encode a protein of 1863 amino acids. This full isoform is expressed in human cells lines and contains all the characteristic domains of BRCA1, and plays all of the functions of BRCA1 such as: DNA repair and DNA damage-responsive cell cycle checkpoints. BRCA1 is able to link BARD1 through the N-terminal RING domain (Fabbro et al., 2002) and has been shown to be an alternative mechanism for the nuclear transport of BRCA1 splice isoforms. For example Δ11q and Δ11 lack the NLS but have the RING domain.
Another mechanism that allows BRCA1 Δ11 to enter into the nucleus has also been reported. Qin et al proposed a piggy-back mechanism of BRCA1 translocation into the nucleus through the binding of BRCA1 to ubiquitin conjugation enzyme 9 (Ubc9) (Qin et al., 2011). Ubc9 is a SUMO-E2 conjugating enzyme, which is transported into the nucleus through importin B (Figure 1.16). Recent publications have outlined the presence of another element that can help a BRCA1 that lacks NLS to enter into the nucleus. There are two splicing isoforms of BRCA1 (Δ11 and Δ11q isoform) which are able to enter into the nucleus even if there are not the NLS. These two splicing isoforms have 39aa of exon 11 which could be responsible for their transport into the nucleus. Consequently these 39 amino acids function like a NLS (Korlimarla et al., 2013).

The Δ11 isoform is composed of 21 coding exons arising from in-frame splicing between exon 10 and exon 12 where the exon 11 is excluded. It retains the highly conserved amino-terminal RING finger and carboxyl-terminal BRCT domains found in full-length BRCA. The Δ11 isoform lacks the nuclear localization signals previously identified in exon 11(Huber et al., 2001). Studies of cells homozygous for the allele BRCA1 Δ11 have shown that this isoform is inefficient for binding the protein Rad51, suggesting that exon 11 is required for recruitment of Rad51. The inability to localize Rad51 may compromise the capacity of those cells homozygous for Δ11 to repair double-stranded breaks, thereby contributing to the defective G2/M checkpoint observed in response to ionizing radiation in these cells (Huber et al., 2001).

BRCA1 is located in the centrosome and interacts with a variety of proteins that are involved in centrosome duplication. The interaction between BRCA1 and the protein is mediated by amino acid domain situated in BRCA1 exon 11 spanning from 508 to 803aa, and this information provides information that can determine the different effects of BRCA1 full length and BRCA1 Δ11 on centrosome duplication (Kim et al., 2006).

Another important isoform of *BRCA1* is the Δ11q also termed BRCA1 Δ11b (Kim et al., 2006). This isoform derives from alternative splicing in which most of the exon 11 (from nucleotide 905-4215) but not all, is excluded from the mRNA. The BRCA1 Δ11q protein appears to be functionally different from the full length in several aspects. First, as a consequence of loss of the nuclear localization signal (NLS) encoded in exon 11q, BRCA1 Δ11q is predominantly cytoplasmic. Although there is the possibility of transporting BRCA1 Δ11q isoform into the nucleus using the alternative pathway described above (Korlimarla et al., 2013; Qin et al., 2011). However the Δ11q isoform plays a role in tumour formation (Maniccia et al., 2009). In fact overexpression of BRCA1 Δ11q isoform causes an increase of this isoform in the nucleus and therefore it is possible to have an increase of apoptosis (McEachern et al., 2003). On the other hand, when there is nuclear depletion of the Δ11q isoform (cytoplasmic retention) it is possible to induce cell proliferation.

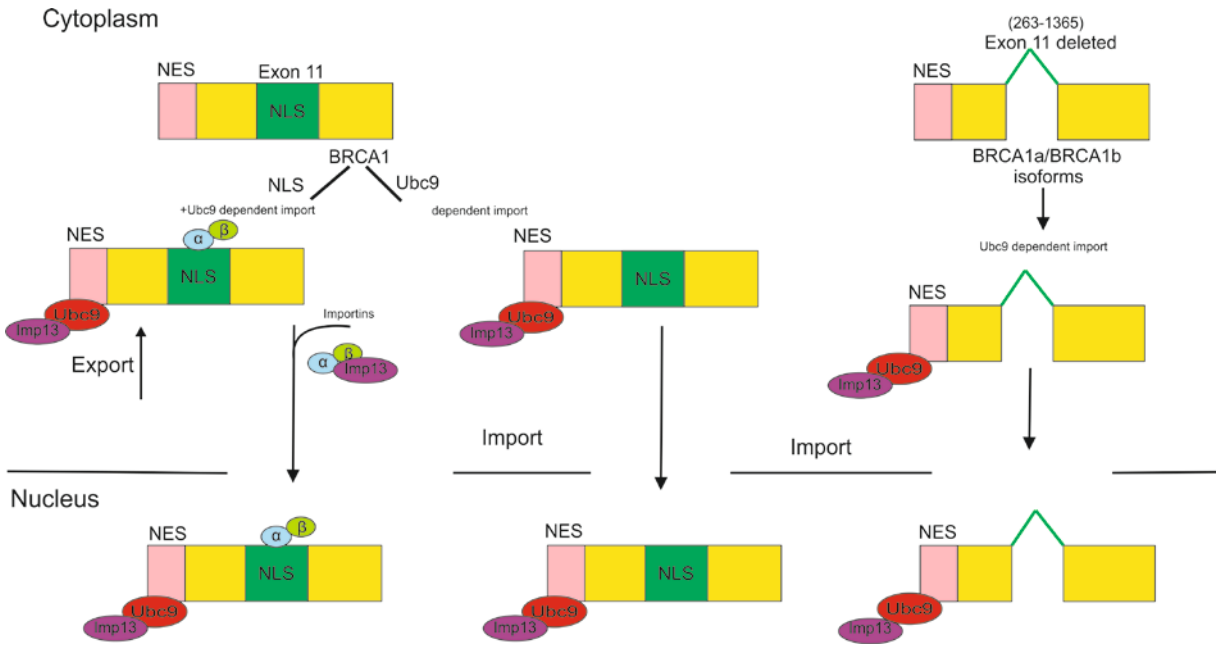


Figure 1.16: Model shows how Ubc9 is an important pathway to shuttle *BRCA1* to the nucleus.

BRCA1 enter in the nucleus through an important alpha and beta pathway and or Ubc9 BRCA1a/1b proteins can enter the nucleus by binding to Ubc9 (Qin et al., 2011).

1.6 Hypothesis and aims.

BRCA1 gene has different isoforms. Alteration of this isoform proportions have been observed in cancer. The alterations may be due to mutations in splicing regulatory sequences.
Balancing these isoforms in cancer is fundamental for obtaining tumour suppression.

Aims:

1. Development of a mini-gene construct to analyse sequence variants and splicing mechanism;
2. Determination of the sequence variants affecting BRCA1 splicing;
3. Identification of the regulatory sequence/regions affected by these mutations;
4. Identification of splicing factors involved;
5. Developing antisense-based strategies to modify aberrant alternative splicing patterns.

Chapter 2

Materials and methods.

2 Chemicals and reagents.

All chemicals used, unless otherwise stated, were obtained from Sigma, Fischer Scientific and Bio-Rad.

2.1 Molecular biology: DNA manipulation.

2.1.1 Bacterial cultures.

Bacterial cultures were used for DNA preparation (Methods 2.2.2) and bacteria transformation (Methods 2.1.2).

Cultures were originated from single colonies grown and selected in solid LB agar plates (10 g/L bacto tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar) containing the appropriate selective antibiotic (50 µg/mL ampicillin). All bacterial cultures were grown shaking at 250 rpm in LB broth in the presence of the appropriate selective antibiotic. Unless otherwise stated, bacteria growth was performed at 37 °C. Plasmids were maintained in the short term as single colonies on agar plates at 4 °C.

All the cultures were handled in sterile conditions.

2.1.2 Bacteria transformation.

Transformations of competent cells were performed in DH5α *E. Coli* cells (Invitrogen or Promega) bacteria which were transformed to prepare DNA for the cloning protocol (Methods 2.2.6).

To transform DH5α *E. Coli* cells, 50 µL of competent bacteria were defrosted on ice, mixed with 1-100 ng of DNA and incubated on ice for 30 minutes. Cells were subjected to heat shock at 42°C for 45 s, incubated on ice for 2 minutes, mixed with 200 µL of super optimal culture (SOC) solution (20 g/L bacto tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose), and grown at 37°C for 1 h. Bacteria were then plated on solid LB agar containing the appropriate selective antibiotic and incubated at 37 °C o/n.

2.2 Constructs.

2.2.1 Mini-gene constructs.

In this study a mini-gene for splicing assay has been designed.

The mini-gene used in this thesis is the pB1 wild type mini-gene that has been created in pCDNA3 (+) vector (Invitrogen) (Figure 2.1) The polylinker of pCDNA3 (+) has been replaced with another polylinker that contains the appropriate restriction sites for cloning the regions of interest (the alpha-globin exon1/intron1 and part of the genomic region of *BRCA1* gene between exon 8 and exon 12 with the flanking intron regions).

Exon 1 of alpha globin (α-globin) is the first exon of the mini-gene providing a strong splice donor site at its 3' end and an ATG start codon at the 5'end. Using specific oligonucleotides and the two-step PCR mutagenesis method (Senapathy, 1988), a stop codon was created in exon 12 and a single nucleotide insertion was created in exon 8 in order to maintain the correct reading frame. Exon/intron 1 of α-globin was amplified from human genomic DNA by polymerase chain reaction (PCR reaction) (Methods 2.2.7), with the primers Alfa Reverse and **Bam** Alfa Forward (Figure 2.1). The PCR product of 220bp was purified (Methods 2.2.10).

A DNA fragment of 340bp containing exon 8 along with of intron 8 (117 nt) was amplified using a PCR reaction (Methods 2.2.7), using oligonucleotides 8**Xho**Reverse and EX8 Forward. The PCR product was purified (Methods 2.2.10). The primer EX8 Forward inserted an additional nucleotide (cytosine) in position 3 of exon 8; this step was necessary to maintain the correct open reading frame in the mini-gene construct (from ATG start codon in alpha globin exon 1 to the TAG stop codon created in BRCA1 exon 12). The primers EX8 Forward and Alpha Reverse contain specific tails that allowed annealing of the two PCR products in a second PCR step. In this second PCR step the two PCR products (containing α-globin exon/intron 1 and BRCA1 exon/intron 8) were mixed together and amplified with the primer **Bam** Alpha Forward, used for the amplification of the α-globin, and the primer 8**Xho**Reverse, used for the amplification of the exon 8. In this way it was possible to obtain the amplification of the two inserts joined together. The PCR product was purified and digested with Xho and BamH1 restriction enzyme (Methods 2.2.6). The DNA fragment of ~560bp was cloned in to pcDNA3.1 (Figure 2.1).

DNA fragment of 528bp containing exon 9 along with part of the intronic flanking regions was amplified using a PCR reaction from genomic DNA using oligonucleotides Ex 9 **Sal1** forward and Ex9**Xho** Reverse. The PCR product was purified and digested with SalI and Xho restriction enzyme. The DNA fragment was cloned into pcDNA3.1 exon 8 α (Figure 2.1).

A DNA fragment of 660bp containing exon 10 along with intronic flanking regions was amplified using a PCR reaction from genomic DNA using oligonucleotides Ex10**Xho** Forward and **HindIII**10R Reverse. The PCR product was purified and digested with Xho and HindIII restriction enzymes. The DNA fragment was cloned into pcDNA3.1 exons 8 and 9 (Figure 2.1).

A DNA fragment of 1973bp containing part of exon 11 along with intron 10 flanking regions was amplified using a PCR reaction from genomic DNA using oligonucleotides 11HindIII forward and 11 Reverse. The PCR product was purified and digested with KpnI and HindIII restriction enzymes. The DNA fragment was cloned into the pcDNA3.1 plasmid containing exons 8, 9 and 10 (Figure 2.1).
A DNA fragment of 2255bp containing the end of exon 11 and exon 12 along with intron 11 was amplified using a PCR reaction from genomic DNA using oligonucleotides 12 forward and Xba12Stop Reverse. The PCR product was purified and digested with KpnI and Xba restriction enzymes. The DNA fragment was cloned into pcDNA3.1 plasmid containing exons 8, 9, 10 and 11 (Figure 2.1). This final mini-gene is called pB1 (Figure 2.1).

The sequence and the respective associated restriction sites of each primer are specified in Table 2.1. All constructs were made using standard PCR methods (Methods 2.2.7) with Taq polymerase Roche.

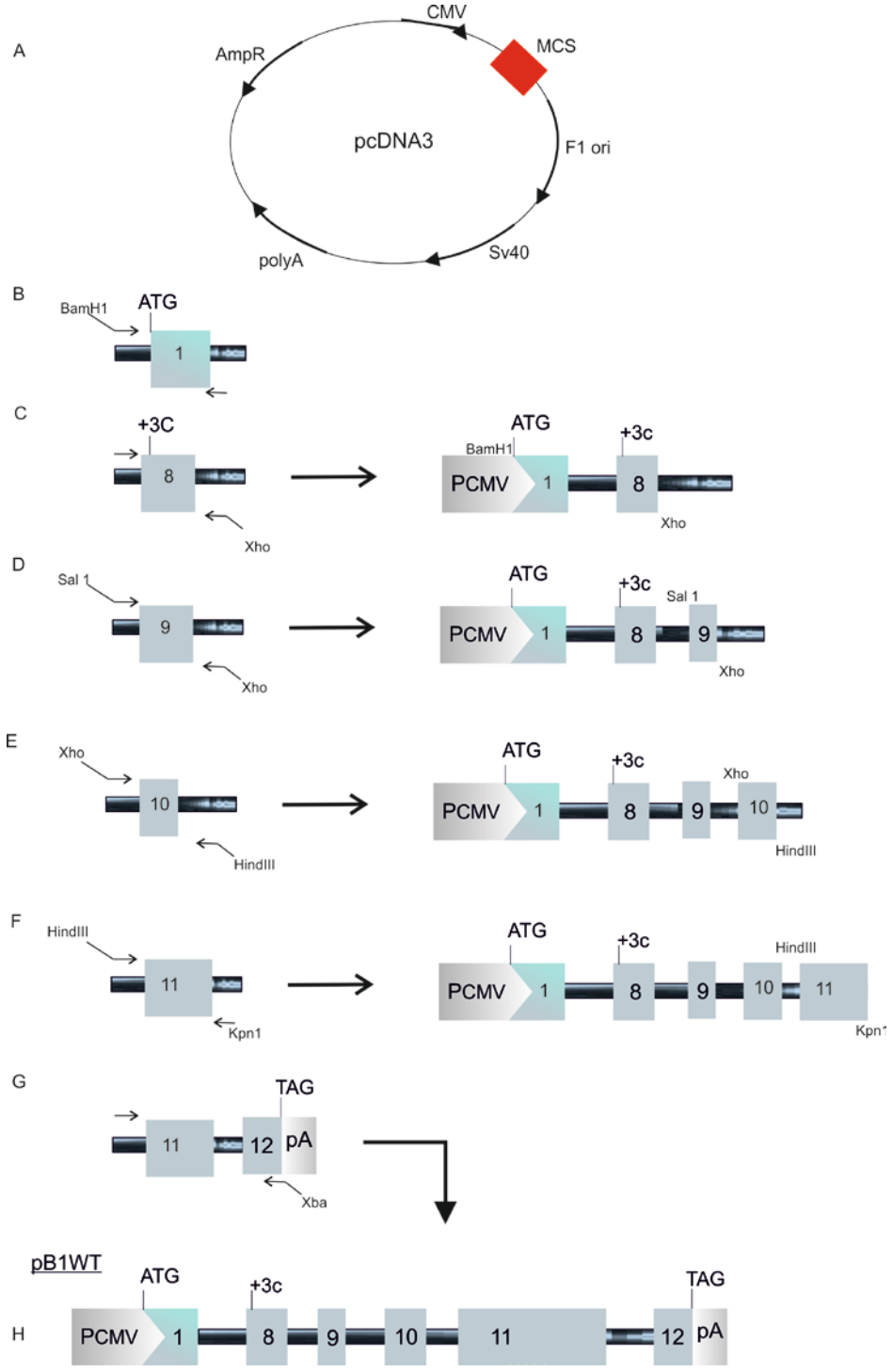


Figure 2.1: Mini-gene constructs.

Diagram of pCDNA3 vector (5446 bp), showing characteristic elements of the plasmid (Ampr, F1 ori, SV40, PolyA, CMV and MCS) and each of the six PCR

products (Alpha globin, Exon8, 9, 10, 11 and 12) cloned into the vector multiple cloning sites (MCS) to produce pB1WT vector (B-H). The blue box represent the exons and the green line the introns.

Table 2.1: Primers for the mini-gene constructs.

Primers utilized in this study to produce the pB1 mini-gene. For each primer, the restriction sites incorporated are also indicated and are in bold.

Primer name	Primer sequence	Restriction Site
Alfa Reverse	5' TGGTTGTCCTGCGGGGAGAA 3'	
Bam Alfa Forward	5' TT**GGATCC**ATGGTGCTGTCTCCT 3'	**BamH I**
8**Xho**Reverse	5' AAAA**CTCGAG**AGGTGTGAGCCA 3'	**XhoI**
EX8 Forward	5'CCCGCAGGACAACCAGTCTCAGT 3'	
Ex 9 Reverse	5' GTTATAAGCGGCCTCACTAC 3'	
Ex 9 Reverse	5' AAACTCGAGAAAGAAAGCAAACA 3'	
Ex10**Xho** Forward	5' CGT**CTCGAG**CAACCATTTCATTTC 3'	**XhoI**
10**HindIII** Reverse	5'TTT**AAGCTT**TCCCGTTGTTAGTC 3'	**HindIII**
11**HindIII** Forward	5' TAAGT**AAGCTT**AATGGCCAGTAA 3'	**HindIII**
11 Reverse	5' ACTTTCCTGAGTGCCATAATCA 3'	
12 Forward	5' TAGAGAGTAGCAGTATTTCATT 3'	
Xba12Stop Reverse	5' AAA**TCTAGA**TGGTAAAATGTCA 3'	**Xba**

2.2.2 DNA preparation.

Different DNA preparations were utilised depending on the requirements of the experiment, the concentration, level of purity and quantity required.
DNA required for ligation, transfection, PCR and plasmid manipulations was prepared through mini-prep purifications.

To perform mini-prep purifications various kits were used through the course of this study, including Promega, QIAGEN, and Fermentas.
In general, after DNA transformation a single colony from the LB agar plate was grown o/n in 5 mL of LB containing the appropriate selective antibiotic. The culture was centrifuged at 12000 x g, 4 °C for 5 minutes and the resultant pellet was resuspended in 250 µL of cell resuspension buffer (50 mM TrisHCl pH 8.0, 10mM ethylenediaminetetraacetic acid (EDTA), 100 µg/mL RNaseA). Bacteria were lysed in 250 µL of cell lysis solution (200 mM NaOH, 1% sodium dodecyl sulphate [SDS]) and the lysate was incubated for 2 minutes at room temperature (RT). To neutralise the lysate 350 µl of the Neutralization Solution (3.0M potassium acetate) was added and the solution mixed gently, following this it was centrifuged at 12000 x g for 5 minutes. The supernatant containing plasmid DNA was added to a spin column and centrifuged at 12000 g for 1 minute. The DNA bound to the column was washed twice: each time 500 µL of Wash Solution (1 M NaCl, 50 mM 3-[N-morpholino] propane sulfonic acid (MOPS) pH 7.0, 15% isopropanol) was added to the column and the column was centrifuged at 12000 g for 1 minute. To avoid residual ethanol the column was centrifuged at 12 000 g for an additional minute. 50 µL of H_2O was added to elute the DNA form the column and the column was centrifuged at 12000 g for 1 minute. The DNA was quantified through UV spectrophotometry at 260 nm wavelength (λ).
The compositions of the different buffers used and the main steps followed in the DNA purification protocols are summarized in Table 2.2 and Table 2.3.

Table 2.2: DNA preparation.

Table summarising the main steps followed to obtain DNA preparation by mini-prep protocols.

	Miniprep
Starting culture	1 colony into 5ml LB, o/n
Pelletting	12000 g, 4 °C, 5 min
Cell Resuspension	250 µL resuspension buffer
Cell Lysis	250 µL lysis buffer 2 min RT
Neutralisation	350 µL neutralisation buffer
Centrifugation	12000 g, 4°C, 5 min
DNA binding	supernatant added to column
Washes	12000 g, 1 min twice: 500 µL wash buffer 12000 g, 1 min
DNA elution	50 µL H_2O
DNA quantification	spectrophotometry,260 nm λ

Table 2.3: Buffer for DNA preparation.

Table summarising the different buffers utilised to obtain DNA preparations.

Buffer	Buffer Composition
Cell resuspension buffer	50 nM TrisHCl pH 8.0 10 nM EDTA 200 µg/ml Rnase A
Cell lysis buffer	200 nM NaOH 1% w/v SDS
Neutralising and Precipitating buffer	3 M potassium acetate
Equilibration buffer	750 nM NaCl 50 nM MOPS pH 7.0 15% Isopropanol 0.15% Triton X-100T
Washes buffer	1 M NaCl 50 nM MOPS pH 7.0 15% v/v Isopropanol
Elution buffer	1.25 M NaCl 50 nM TrisHCl pH 8.5 15% v/v Isopropanol

2.2.3 DNA quantification.

DNA obtained from mini-preps (Methods 2.2.2) was quantified through spectrophotometry. Sample concentration was estimated by its light absorption at 260 nm λ (OD_{260}). DNA quality and purity was assessed by comparing the light absorption values of the sample at 230 nm (OD_{230}) compared to OD_{260} and at 280 nm (OD_{280}) compared to OD_{260}. Ratio OD_{260}/OD_{230} greater than 1.5 indicated negligible DNA contamination by organic compounds or guanidium salt (present in DNA purification columns); ratio OD_{260}/OD_{280} greater than 1.8 was the index of protein-free samples.

2.2.4 DNA digestion.

DNA was digested by restriction enzymes to perform restriction analysis or to produce intermediate fragments for cloning (Methods 2.2.6) and to eliminate DNA template in site directed mutagenesis.

Digestions were performed in a 20-100µL reaction volume and contained 1-5 µg of DNA, 1-5 U of restriction enzyme(s), 1x reaction buffer specific for the enzyme and 100 µg/mL of bovine serum albumin (BSA) when required by the enzyme(s) used. When double digestions were performed, the restriction buffer enabling highest activity of both enzymes was employed. Most of the digestions were performed by incubating the reaction volume at 37 °C for two or three hours and inactivated by incubating at 65 °C for 20 minutes; different incubation and inactivation temperatures were used as required by specific enzymes.

2.2.5 DNA electrophoresis.

DNA electrophoresis was performed to check plasmid preparations (Methods 2.2.2), to identify successful clones (Methods 2.2.8), to visualise PCR products (Methods 2.2.7), or to isolate and purify specific DNA fragments from restriction digestions or PCR reactions.
DNA samples were mixed with DNA loading buffer (0.05 % bromophenol blue, 40 % sucrose, 1 M EDTA pH 8.0, 0.5 % SDS). Electrophoresis was performed at room temperature (RT) an agarose gels (0.8-2 % agarose, Tris-borate-EDTA buffer) in Tris-borate-EDTA.
(TBE) buffer (40 mM TrisBase, 1 mM EDTA pH 8.0, 20 mM acetic acid), applying a voltage of 80-100 V.

2.2.6 Cloning.

Cloning protocols were utilized to produce the different constructs used in this thesis.
Specific primers which included 5` and 3` restriction enzymes were designed to amplify regions of interest by PCR reaction (Methods 2.2.7). The vector and the PCR products were digested (Methods 2.2.4) with the appropriate enzyme. Digested vector and PCR product were analysed by DNA electrophoresis (Methods 2.2.5), quantified (Methods 2.2.5) and gel extracted (Methods 2.2.9). Insert (digested PCR product) and the digested vector DNA were ligated (Methods 2.2.11) and transformed into NEB Express High Efficiency *E. Coli*.
Mini prep purification (Methods 2.2.2) was utilized to isolate plasmid DNA form bacteria colonies. DNA was then screened by restriction analysis (Methods 2.2.4) and sequenced (Methods 2.2.12) to ensure it corresponded with the product of interest.

2.2.7 Polymerase chain reaction (PCR reaction).

PCR reactions were performed to obtain amplification of single exons and flanking introns for the mini-gene preparation, to perform site directed mutagenesis, deletion, insertion and for cDNA amplification.

To accurately amplify the exons and analyse the splicing process the PCR reaction was performed using Fast Start Taq Polymerase (Roche), in 25 µL or 100 µL volume Table 2.4.

Every PCR reaction consisted of an initial phase of DNA denaturation, followed by a variable number of cycles including three phases: 1) DNA denaturation; 2) Primer annealing and 3) DNA elongation. Lastly, the reaction was completed by a phase of final DNA elongation.

The duration of each phase and the temperatures chosen for DNA denaturation and elongation were determined by the type of DNA polymerase employed, by the size of the final amplification product and by the oligos melting temperature. Annealing temperature (Ta) was a function of the characteristics of the primers and was optimised by gradient PCR for each pair of primers and their relative template. The number of cycles performed was 25 for PCR employed in mutagenesis and amplification of inserts for cloning; while 35 cycles were used for PCR amplification of cDNA products.
Table 2.5 summarises the cycling conditions used for each PCR reaction for the creation of the mini-gene.
In addition the Table 2.6 and Table 2.7 summarizes the cycling conditions used for site direct mutagenesis and for the entire mutant used.

Table 2.4: PCR reaction volume.

Table summarizing the composition of the reaction buffer used to perform PCR.

DNA polymerase	Fast start Taq DNA polymerase (Roche)	Fast start Taq DNA polymerase (Roche)	LongRange Taq DNAPolymerase (Qiagen)
Purpose	cDNA amplification	PCR products	cDNA amplification
Reaction Volume	25 µL	100 µL	25 µL
Template	1-10 ng cDNA	1-10 ng DNA	50-500 ng cDNA
dNTPs	200 µM each	200 µM each	500 µM each
Primers	0.25 µM each	1 µM each	0.4 µM each 10X
DNA polymerase buffer	1x Fast start DNA Buffer	1x Fast start DNA Buffer	LongRange PCR Buffer with Mg_2 10X
DNA polymerase	2U/µL Fast start Taq DNA polymerase	2U/µL Fast start Taq DNA polymerase	2U/µL of Long RangePCR Enzyme Mix.
Water	to make the final volume	to make the final volume	To make the final Volume

Table 2.5: PCR cycling conditions for the mini-gene.

Table summarising the cycling conditions to perform PCR to create the mini-gene constructs.

PCR product to obtain	αglobin ,ex 8 and ex 9	exon 10	exon 11	exon 12
DNA polymerase	Taq DNA polymerase (Roche)	Taq DNA polymerase (Roche)	Taq DNA polymerase (Roche)	Taq DNA polymerase (Roche)
Initial Denaturation	3 min, 95 °C	3 min, 95 °C	3 min, 95 °C	3 min, 95 °C
Number of cycles	35	35	35	35
-denaturation	30 sec 95 °C	30 sec 95 °C	30 sec 95 °C	30 sec 95 °C
-annealing	3 min, 56 °C	3 min, 58 °C	3 min, 56 °C	3 min, 54 °C
-elongation	30 sec 72 °C	1:30 min, 72 °C	2min, 72 °C	1:30 min, 72 °C
Final extension	5 min, 72 °C	5 min, 72 °C	5 min, 72 °C	5 min, 72 °C

Table 2.6: PCR cycling conditions for site directed mutagenesis.

Table summarising the cycling conditions to perform PCR which has been used to create site-directed mutagenesis.

PCR products	PCR 1	PCR 2	PCR 1+2
DNA polymerase	Taq DNA polymerase (Roche)	Taq DNA polymerase (Roche)	Taq DNA polymerase (Roche)
Initial denaturation	95 °C 3 min	95 °C 3 min	95 °C 3 min
Number of cycles	35 cycles	35 cycles	10 cycles
Denaturation	95 °C 40 sec	95 °C 40 sec	95 °C 40 sec
Annealing	60 °C 45 sec	58°C 45 sec	48 °C 45 sec
Elongation	72 °C 1:30 min	72 °C 1:30 min	72 °C 1:30 min
Number of cycles			10 cycles
Denaturation			95 °C 40 sec
Annealing			50 °C 45 sec
Elongation			72 °C 1:30 min
Number of cycles			5 cycles
Denaturation			95 °C 40 sec
Annealing			52 °C 45 sec
Elongation			72 °C 1:30 min
Final Elongation	72 °C 5 min	72 °C 5 min	72 °C 5 min

Table 2.7: PCR of the mini-gene with the mutation.

Setting employed for the PCR reactions used to create the mini-gene SF2/Del and 1923A. Temperature and duration of each phase are indicated.

PCR products	PCR 1	PCR 2	PCR 1+2
DNA polymerase	Taq DNA polymerase (Roche)	Taq DNA polymerase (Roche)	Taq DNA polymerase (Roche)
Initial denaturation	95 °C 3 min	95 °C 3 min	95 °C 3 min
Number of cycles	35 cycles	35 cycles	10 cycles
Denaturation	95 °C 40 sec	95 °C 40 sec	95 °C 40 sec
Annealing	48 °C 45 sec for1923A 52 °C 45 sec for Sf2 56 °C 45 sec for Del	48 °C 45 sec for1923A 52 °C 45 sec for Sf2 56 °C 45 sec for Del	52 °C 45 sec for1923A 50 °C 45 sec for Sf2 54 °C 45 sec for Del
Elongation	72 °C 1:30 min	72 °C 1:30 min	72 °C 1:30 min
Final Elongation	72 °C 5 min	72 °C 5 min	72 °C 5 min

2.2.8 PCR mediated mutagenesis.

Different mutations were introduced in the pB1 mini-gene (688G>A, 689A>C, 690G>T, 691A>C, 692C>T, 692C>A, 693G>A, 693G>C, 693G>T, 694G>C, 694G>T); a deletion (between codon 688 to 694), an insertion (between codon 688 to 694) and a double mutant (c.689A>C and 693G>A).

The mutated mini-genes were generated through a two-step PCR overlap extension using the pB1 WT constructs as a template.
The first step includes two PCR reactions. One PCR reaction uses a forward primer (i10F; with a non complementary tail for a restriction site that will be used for the digestion of the insert) and a mutagenic reverse primer. The other PCR reaction uses a forward mutagenic primer and a reverse primer (NewsEx11Reco) that, presents the restriction site that will be used for the digestion of the insert.
The mutagenic primers carry the specific mutation and also present a non complementary tail to the DNA template. The tail was designed to allow annealing, in the second PCR step, of the two first step PCR products.
The second PCR step mixes the two PCR products from step one. These PCR products are the template; i10F and NewsEX11Reco are the forward and the reverse primer respectively.
After these two steps a new insert carrying the mutation is obtained. The primers used are summarized in Table 2.8, Table 2.9 and Table 2.10.

Table 2.8: Primers for the mini-gene constructs.

Table indicating the name and the sequence of all primers utilised for the mutagenesis.

Constucts	Primers Sequence
688 G>A	18BRCA1F (Forward) 5'AATTTTCT**Y**AGACGGATGTAACAAA 3' 18BRCA1R (Reverse) 5' ATTTGTTACATCCGTCT**R**AGAAAA 3'
689 G>T	19BRCA1F (Forward) 5' ATTTTCTG**Y**ACGGATGTAACAA 3' 19BRCA1R (Reverse) 5' ATTTTGTTACATCCGTC**R**CAGAA 3'
690 G>T	20 BRCA1 F (Forward) 5' TTCTGA**Y**ACGGATGTAACAAATAC 3' 20BRCA1R (Reverse) 5' TCCGT**R**TCAGAAAATTCACAAG 3'
691 A>C	231 2F (Forward) 5' TTCTGAG**Y**CGGATGTAACAAATAC 3' 231 2R (Reverse) 5' TACATCCG**R**CTCAGAAAATTCACAA 3'
692C>T 692 C>A	231 1F (Forward) 5' TTCTGAGA**R**GGATGTAACAAATAC 3' 231 1R (Reverse) 5' TACATCC**Y**TCTCAGAAAATTCAC 3'
693G>C 693G>T	231 GTC F (Forward) 5' TCTGAGAC**Y**GATGTAACAAATACTG 3' 231 GTC R (Reverse) 5' TTACATC**R**GTCTGAGAAAATTCAC 3'
694G>C 694G>T	231 +1F (Forward) 5' CTGAGACG**Y**ATGTAACAAATACTGAAC 3' 231+R (Reverse) 5' GTGAATTTTCTGAGACG**R**ATGTAAC

Table 2.9: Primers for the deletion and insertion constructs.

Table indicating the name and the sequence of all primers utilised for the deletions and insertion.

Constructs	Primers Sequence
SF2/ASF Insert	SF2F (Forward) 5' TTCTGAAGAAGATGTAACAAATACTGAACATC 3' SF2R (Reverse) 5' TTGTTACATCTTCTTCAGAAAATTCACAAGCA 3'
1923A	1923AF (Forward) 5' CTGCGACAGATGTAACAAATACTGAACA 3' 1923AR (Reverse) 5' TTGTTACATCTGTCGCAGAAAATTCAC 3'
Del	DelF(Forward) 5' GTGAATTTTCTATGTAACAAATACTGAACATC 3' DelR (Reverse) 5'TTTGTTACATAGAAAATTCACAAGCAG 3'

Table 2.10: Primers used for the second step of PCR in the site direct mutagenesis.

The table summarise the primer forward and reverse used in the step 2 of the site direct mutagenesis. The sequences of primers are showed.

.

Primer name	Primers Sequence
i10F	5' TGTCGTGGTATGGTCTGATA 3'
NewEx11Reco	5' AGGCTGGTTTGCTTTTATTACAG 3'

2.2.9 DNA gel extraction.

Specific DNA products were separated by agarose gel electrophoresis (Methods 2.2.5) and extracted and purified from agarose gels using the Promega Gel Extraction Kit or the Fermentas gel extraction kit. When the Promega Gel Extraction Kit was employed, the band of interest was visualized under UV light, cut from the agarose gel and transferred to a tube. 10µl Membrane Binding Solution per 10mg of gel slice was added to the DNA. The mixture was vortexed for 30 seconds and incubated at 50–65°C vortexing every 2 minutes; until the gel slice was completely dissolved. The dissolved gel mixture was transferred to the Mini-column and incubated at room temperature for 1 minute. It was then centrifuged at 16000 x g for 1 minute and 700 µl Membrane Wash Solution (ethanol added) was added to the column.
The solution was centrifuged at 16000 g for 1 minute and the process was repeated with 500 µl Membrane Wash Solution. A further centrifugation was performed at 12000 g for 5 minutes, the tube was emptied and the column centrifuged for 1 minute with the micro centrifuge lid open to allow evaporation of any residual ethanol. The mini-column was transferred into a clean 1.5 ml micro centrifuge tube and 50 µl of Nuclease-Free Water added. This was then incubated at room temperature for 1 minute and centrifuged at 16000 g for 1 minute. The eluted DNA was quantified through UV spectrophotometry at 260 nm wavelength.

Alternatively, when the Fermentas kit was employed the band of interest was visualized under UV light, cut from the agarose gel and incubated at 65°C for 10 minutes with 1:1 volume of Binding Buffer. The dissolved gel mixture was transferred to the GeneJet Purification Column and centrifuged at 16000 g for 60 sec. 700 µl Wash buffer with ethanol was added to the column and centrifuged for 1 minute. The tube was emptied and the column centrifuged for 1 minute with the micro centrifuge lid open to allow evaporation of any residual ethanol. The mini-columns were transferred into a clean 1.5 ml micro centrifuge tube and 50 µl of

Elution Buffer was added. This was then incubated at room temperature for 1 minute. The solution was finally centrifuged at 16000 g for 1 minute.

2.2.10 PCR purification.

PCR products for cloning protocols were purified using Wizard Genomic PCR Purification Kit. The 100 µL PCR reaction was mixed with 100 µL of Membrane Binding Solution and the solution was transferred to a column and centrifuged at 16000 g for 1 minute. The column was washed with 700 µL of membrane wash solution and centrifuged at 16000 g for 1 minute. The column was then washed with 500 µL of membrane wash solution. The PCR product was eluted by adding 50 µL of ultra-pure water to the column, incubating 1 minute and centrifuging at 16000 g for 1 minute.

2.2.11 DNA ligation.

Ligation was performed to clone insert (restriction digested PCR products) into the vector (restriction digested pB1 mini-gene; restriction digested vector pCDNA3 [+]).

Vector was digested with the appropriate restriction enzymes, isolated by agarose gel electrophoresis (Methods 2.2.5) and purified by gel extraction (Methods 2.2.9). The insert of interest was obtained by PCR reaction (Methods 2.2.7), identified by electrophoresis, isolated and purified by gel extraction (Methods 2.2.9). The ligation was set up in a reaction volume of 20 µL containing 100 ng of total DNA, 1 U of T4 DNA Ligase (NEB) and 1 x of T4 DNA Ligase buffer (NEB). The amount of DNA in the reaction was calculated to have a fixed molar ratio between the vector and the insert; molar ratios of 1 to 1, 1 to 3 or 1 to 6 were tested each time. The ligations were incubated o/n at room temperature Table 2.11 summarizes ligation protocols.

Table 2.11: Ligation.

Table summarizing the main steps for DNA ligation.

Type of insert	Insert from gel extraction
Reaction volume preparation	100ng total DNA +H_2O +1X Ligase Buffer +200U T4 DNA Ligase
Incubation	Overnight RT

2.2.12 DNA sequencing.

All new constructs, PCR products (Methods 2.2.9) cloning (Methods 2.2.8) and site direct mutagenesis products were sequenced by Gene service - Source BioScience plc.

Plasmid DNA was supplied to the company at the concentration of 100 ng/µl diluted in water; PCR products were purified and supplied at the concentration of 1 ng/µl for every 100 bp of product; sequencing primers were supplied at the concentration of 3.2 pmol/µl.

2.3 Molecular biology: RNA manipulation.

2.3.1 RNA extraction.

To analyse the splicing process total mRNA was extracted from MCF7 breast cancer cells (Methods 2.7.1).
RNA was collected after transfection of cells (Methods 2.7.3) with the appropriate mini-gene construct (Methods 2.2.1). 48 hours after transfection the RNA was extracted using the RNeasy Plus Mini kit.
RNA was collected from cells after 48 hours from the transfection. The cells were washed with PBS and RLT Plus Buffer was added to lysate the cells (Table 2.12).
The lysate was homogenised by pipetting up and down and transferred to a gDNA Eliminator spin column (to eliminate genomic DNA) and centrifuged for 30 seconds at 10000 x *g*. RNA purification was performed by adding 1 volume (usually 350 µl or 600 µl) of 70% ethanol and mixed.
The sample was transferred into an RNeasy spin column and centrifuged for 15s at 10000 x *g*. After that 700 µl of RW1 buffer was added to the RNeasy spin column and it was centrifuged for 15s at 10000 x g to wash the spin column membrane. The membrane was washed twice with 500 µl of RPE buffer; 50 µl of nuclease free water was added and centrifuged for 1 min at 10000 x *g* to elute the RNA. The RNA was stored at - 80 °C (Table 2.13).
RNA was quantified through UV spectrophotometry at 260 nm wavelength.

Table 2.12: Volumes of Buffer RLT Plus for lysing pelleted cells.

Number of pelleted cells	Volume of Buffer RLT Plus
< 5 X10^6	350 µl
5 X10^6 - 1 X10^7	600 µl

Table 2.13: RNA extraction.

Table summarising the main steps for RNA extraction from cell lines.

Cell homogenisation	350 µl RLT < 6 cm dish centrifugation >8000 g at 30 sec
RNA Isolation	70% Ethanol centrifugation >8000 g at 30 sec 700 µl RWI Buffer centrifugation >8000 g at 30 sec
RNA Purification	two washes 500 µl RPE Buffer centrifugation >8000 g at 30 sec
RNA Elution	50 µl RNAse free water centrifugation >8000 g at 1 minute

2.3.2 RNA extraction from blood.

To analyse the splicing process total mRNA was extracted from blood.
Blood was collected in PAXgene RNA tubes, inverted for 10 times centrifuged for 10 minutes at 3500 x g at room temperature and the supernatant was removed. 4ml RNase free water was added to the pellet. The pellet was vortexed until visibly dissolved and centrifuged for 10 minutes at 3500 x g at room temperature. The pellet was air dried for 10 minutes and resuspended in 350 µl of BR1 buffer. Once the pellet was completely dissolved 300 µl of BR2 buffer and 40 µl proteinase K was added, the sample was vortexed and incubated for 10 minutes at 55 °C using a shaker incubator at 1400 rpm. The sample was centrifuged for 3 minutes at 13000g. The supernatant was carefully transferred to a fresh 1.5 ml micro centrifuge tube. 350 µl of 100% ethanol was added 700 µl of the sample was added to a PAXgene RNA spin column placed in a 2ml processing tube, the sample was centrifuged to maximum speed and 350 µl of Buffer BR3 was added and again centrifuged to maximum speed.
The DNase stock solution was prepared by dissolving DNase I (1500 Kunitz units) in 550 µl of RNase-free water (1500 Kuntz units/0.55 mls). This was mixed by inversion and stored at -20ºC. 10.5 µl DnaseI solution and 73.5 µl of Buffer RDD was added to each sample and mixed gently by flicking the tube. This was centrifuged briefly to collect residual liquid from the sides of the tube, and then incubated at room temperature for 15 minutes. 350 µl of Buffer BR3 was added to the PAXgene RNA spin column and it was centrifuged for 1 minute at 13000 x g.
500µl of Buffer BR4 was added to the PAXgene RNA spin columns and centrifuged for 3 minutes at maximum speed. The columns were placed in a new 1.5ml elution tube and 40 µl of buffer BR5 was added and centrifuged for 1 minute. The eluted RNA was quantified through spectrophotometry at 260 nm λ and stored at -80 °C.

2.3.3 Reverse Transcription polymerase chain reaction (RT-PCR).

To analyse the splicing pathway of transfected mini-genes the total RNA extracted from the cells was retro-transcribed to cDNA using a QIAGEN Long Range 2step RT-PCR. cDNA was then amplified through PCR reaction. A reaction volume of 20 µL was prepared by mixing 4 µL of LongRange RT Buffer 5X, 2 µL of dNTPmix (10mM each), 1 µL of pCSR primer (final concentration 1 µM), 0.2 µL of LongRange RNase inhibitor (4U/µL) 1 µL of LongRange Reverse Transcriptase (Qiagen), 1 µg of total RNA. The reaction volume was brought to 20 µL by adding RNase free water. The reaction was incubated for 90 minutes at 42 °C and the enzyme was inactivated by heating the reaction at 85 °C for 5 minutes.

To discriminate between the mini-gene and the endogenous the pCSR (5′ GCAACTAGAAGGCACAGTCGAGG 3′) primer has been used. The pCSR binds to a specific region of the mini-gene and is located 105 nucleotides upstream the polyadenilation signal.

The cDNA obtained was ready to be analysed by PCR amplification reaction (Table 2.14).

To analyse the splicing pathway of endogenous *BRCA1* the total RNA extracted from the cells (or blood) was retro-transcribed to cDNA using ImProm-II Reverse Transcription System (Promega), see Table 2.14.

A mix of RNA and random primers was incubated for 5 minutes at 70 °C, for denaturation, and immediately chilled in ice. The reverse transcription reaction mix was added and incubated for 5 minutes at 25 °C and one hour at 42 °C. The reverse transcription reaction mix was prepared as shown in Table 2.15. The reverse transcriptase enzyme was inactivated by heating the reaction at 70 °C for 15 minutes. The cDNA obtained was ready to be analysed by PCR amplification reactions (Method 2.2.7). The PCR condition required for each cDNA synthesis kit are summarised in Table 2.16.

The primers used for PCR amplification were: the 9 10 Forward (5′ ACTTATTGCAGTGTGGGAGA 3′) and a mix of two reverse primers; the 12/10/11qReverse (5′ CCAGATGCTGCTTCACCCT 3′) and 11 full (5′ GGAGTCCTATCATTACAT 3′).

The 9,10Forward primer binds the junction between exon 9 and exon 10. The primer 12/10/11qReverse is situated at the beginning of exon 12 and presents two nucleotide complementarity with the end of exon 10 and the end of exon 11q. The 11 full Reverse primers binds exon 11 in a region downstream the Δ11q donor site and is therefore able to detect the 11 full length isoform only (Figure 2.2).

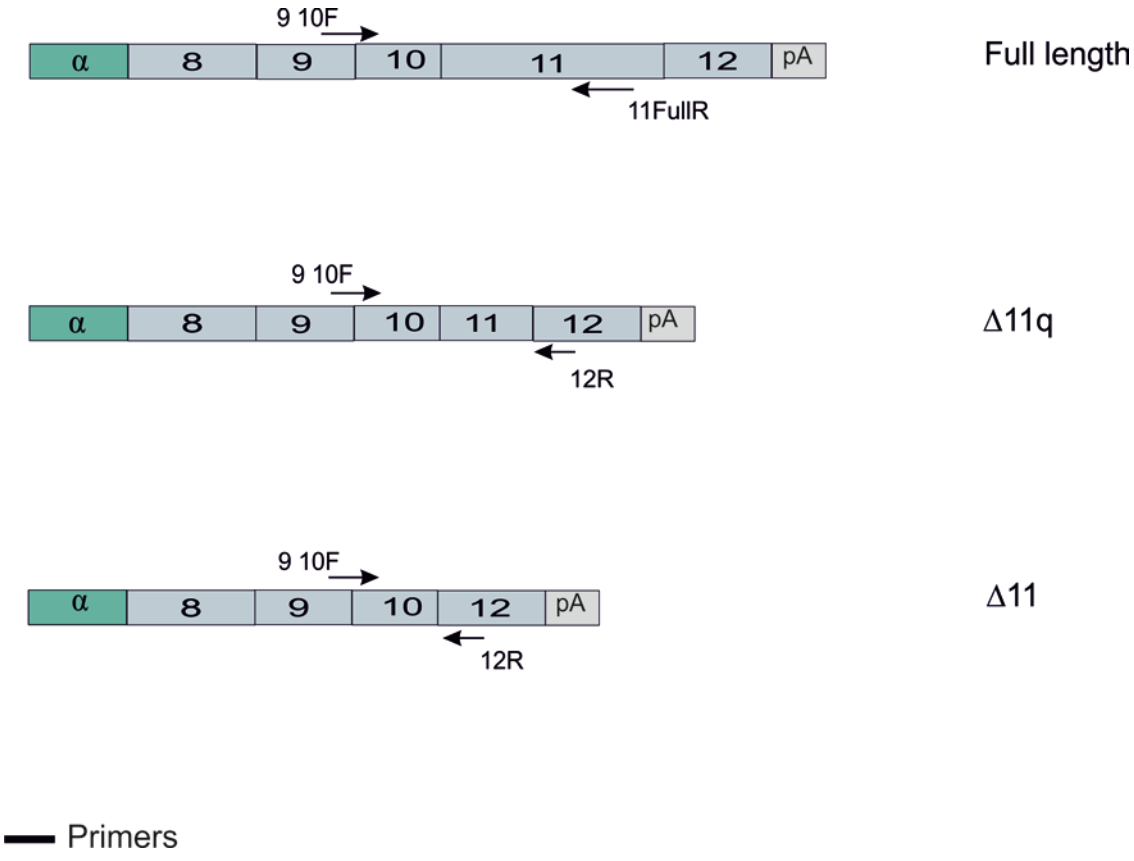


Figure 2.2: Schematic representation of the primer used for the RT-PCR.

The diagram shows the position, in the mature spliced transcripts, of the primer (indicated with the arrow) used to detect the full length, the Δ11q isoform and the Δ11 isoform. The grey boxes represent BRCA1 exon 8 to exon 12. The green box represents αglobin exon 1. pA=polyadenylation.

Table 2.14: Composition of Reaction Mix for Long-Range PCR (0.1–10 kb).

Table summarises the main steps for the RTPCR using Qiagen kit.

Component	Volume in each reaction	Final concentration
LongRange PCR Buffer with Mg2+, 10X	5 µl	1 X 2.5 mM Mg2+
dNTP mix (10mM each)	2.5 µl	500 µM of each dNTP
Primer A	1µl	0.4µM
Primer B	1µl	0.4µM
Long Range PCR Enzyme mix	5 µl	2 units for 50 µl
Template cDNA	5 µl	50-500ng
RNAse free water	Variable up to 50µl	

Table 2.15: Composition of Reaction Mix for Promega Kit.

Table summarises the main steps for the RTPCR using Promega kit.

Component	Volume in each reaction	Final concentration
RNA	X µl	up to 1 µg/reaction
Random Primers	1 µl	0.5 µg/reaction
Final Volume	5 µl	0.4µM
Incubation	70 °C 5 minute	
ImProm II Reaction Buffer	5 µl	1x; 2.5mM Mg2+
dNTP mix (10mM each)	1 µl	500µM of each dNTP
$MgCl_2$	1.2 µl	1.5-8.0mM
Recombinant RNAse ribonuclease Inhibitor	0.5 µl	20 u
ImProm II Reverse Trascriptase	1 µl	
Incubation	25 °C 5 minute 42 °C 60 minute	
Inactivation enzyme	70 °C 15 minute	

Table 2.16: PCR conditions used to analyse the cDNA.

Settings employed for the PCR reactions used throughout this study. Temperature and duration of each phase are indicated, depending on the PCR product to amplify.

	cDNA using Qiagen Kit	cDNA using Promega Kit
DNA Polymerase	Long Range PCR Mix	Taq DNa Polymerase
Initial Denaturation	93 °C 3 minute	93 °C 3 minute
Cycle	35 Cycles	35 Cycles
Denaturation	93 °C 30 sec	95°C 40 sec
Annealing	54 °C 45 sec	56 °C 45 sec
Elongation	68 °C 40 sec	72 °C 45 sec
Final Extension	72 °C 5 minute	72 °C 5 minute

2.4 Biochemistry: Protein analysis.

2.4.1 Nuclear extraction.

To perform a biochemical analysis, the protein samples were extracted from HeLa cells. The extraction was performed using a CelLytic Nuclear Extraction Kit (Sigma).

Nuclear extract was extracted from a large number of cells (10^8-10^{10} cells ~1ml of packed cell volume PCV) were collected and centrifuged for 5 minutes at 450g. 500 ml of 1X Lysis Buffer (including DTT and protease inhibitors) was added to the packed cell (volume 100ml). The cells were incubated on ice for 15 minutes, allowing swelling. In order to assess the degree of lyses before centrifugation trypan blue solution was added to a cell sample.

The crude nuclei pellet was resuspended in ~70 ml (2/3X PCV) of Extraction Buffer containing a 1 ml of 0.1 M DTT solution and 1 ml of protease inhibitor cocktail. This was vortexed at medium speed for 15-30 minutes, and centrifuged for 5 minute. The supernatant was then transferred to a clean tube and snap frozen in aliquots with liquid nitrogen and stored at -80 °C.

2.4.2 Protein quantification.

Protein preparations from cells were quantified through colorimetric assays by using a *DC* Protein Assay kit (Biorad) based on the Lowry method (Lowry et al., 1951).

To prepare the calibration curve as an internal reference, 5 protein samples of 5 µL each were analysed and prepared by diluting BSA at different known concentrations from 0.1 mg/mL to 10 mg/mL. Samples of 5 µL each were also prepared from the total cell lysates to be quantified. Two different dilutions, 1:1

and 1:5 were used. To allow red/ox reactions, 25 µL of alkaline copper tartrate solution (98 % proprietary buffer A, 2 % proprietary buffer S) was added to each sample, the solution was mixed and 200 µL of buffer containing folin (proprietary buffer B) was then added. The reaction volume was incubated at RT for 15 minutes to allow the reduction of the folin and the oxidation of the aromatic residues in the proteins samples, mediated by the reduction and oxidation of the copper. The concentration of the reduced folin in the sample was quantified through UV spectrophotometry at 750 nm wavelength. Being proportional to the reduced folin concentration, the total cell lysate protein concentration was deduced by referring to the calibration curve.

2.4.3 Protein electrophoresis.

Protein electrophoresis was performed to detect proteins with the pull down assay (Methods 2.6.1) and to separate protein prior to Western Blot analysis (Methods 2.5.4).

Protein samples were mixed with 4x loading buffer (8 % SDS, 250 mM TrisHCl pH 6.8, 35 % glycerol, bromophenol blue, 50 mM DTT), incubated at 95 °C for 5 minutes and centrifuged at 12 000 x g for 1 minute (Table 2.17).

Polyacrylamide gels were prepared by polymerising a short 4 % polyacrylamide layer on top of a main 7-15% layer.
Preparation of 4% polyacrylamide (40% acrylamide/ bis-acrylamide (Biorad), 0.125M TrisHCl pH 6.8 10 % SDS, 0.1 % tetramethylethylenediamine (TEMED), 10% ammonium persulfate [APS]), and the preparation of the 7-12 % polyacrylamide layer (40% acrylamide/ bis-acrylamide (Biorad), 0.375 M TrisHCl pH 8.8, 0.1 % SDS, 0.05 % TEMED, 0.05 % APS (Table 2.18) was performed without protein contamination and all devices were previously cleaned with 70% ethanol.

Samples were separated by SDS gel electrophoresis (SDS-PAGE) on 7.5-15% polyacrylamide gels. The short layer of 4% polyacrylamide gel on the top, allowed the protein present in the sample to be compacted. The SDS-PAGE gel was run in a running buffer (25 mM Tris, 192 mM glycine, 0.1 % SDS) by applying voltage 60-100V. Proteins on the gel were either transferred to a nitrocellulose membrane for western blot analysis (Methods 2.5.4) or stained (Methods 2.5.1 and Methods 2.5.2) to be visualised and eventually identified by mass spectrometry *(*Methods 2.5.3).

Compositions of the different buffers and reagents used for protein electrophoresis are summarized in Table 2.18.

Table 2.17:Protein electrophoresis protocol.

Summary of the protocol used for protein electrophoresis.

Protein Pre-treatment	95°C 5 minute in loading buffer 12000 g in 1 minute
Protein electrophoresis	10-12% polycrylamide gel running buffer 60-100 V

Table 2.18: Protein electrophoresis buffers.

Composition of the buffers and reagents employed for protein electrophoresis.

Buffer	Buffer Composition
4X Loading Buffer	8%SDS 250mM TRIS HCl pH 6.8 35% glycerol Bromophenol blue
4% Polyacrylamide gel	40% Acrylamide Bis-solution 0.5mM TRIS Hcl pH 6.8 10% SDS 10% APS 0.05% Temed
7-15% Polyacrylamide gel	40% Acrylamide Bis-solution 0.5mM TRIS Hcl pH 8.8 10% SDS 10% APS 0.05% Temed
Running Buffer	25 mM TRIS 192 mM glycine 0.1% SDS

2.5 Protein staining.

2.5.1 Coomassie blue staining.

To verify the presence of bands in a SDS page gel, the gels were stained using a Coomassie based protocol.
Coomassie based staining was performed by incubating the SDS gels in the Coomassie Blue staining solution (Biorad) for 30 minutes, the solution was removed and Coomassie distaining solution was added until the gel became clear. Gels were rinsed with water and visualised. Another way used to visualize the protein in SDS page was to use the Instant-Blue solution (Expedeon) for 15-30 minutes, after that the solution was removed and the gel rinsed with water and visualised.

2.5.2 Silver staining.

To distinguish and isolate proteins resulting from binding assays SDS gels were silver stained to achieve a better and more sensitive band resolution.

Silver staining was performed by using the ProteoSilver PROTSIL1 kit (Sigma). The gel was fixed by incubation in 100 ml of fixing solution (50% ethanol, 10% acetic acid) for 20 minutes and washed firstly in 100 ml of 30% ethanol for 10 minutes, then in 200 ml of water for 10 minutes. The gel was incubated in 100 ml of proprietary sensitizing solution for 10 minutes and washed in 200 ml of water for 10 minutes twice. A silver equilibration was performed by incubating the gel in 100 ml of proprietary silver solution for 10 minutes, after which the gel was washed in 200 ml of water for 1 minute. The gel was developed by incubation in 100 ml of proprietary developer solution for 3-6 minutes, until the bands were

clearly visible but not too strongly stained. Finally, the reaction was terminated by adding 5 ml of proprietary stop solution to the developer solution and by incubating for 5 minutes. The gels was washed in 200 ml of water for 15 minutes and visualised.

2.5.3 Protein identification.

After protein electrophoresis (Methods 2.4.3) and gel staining (Methods 2.5.1-2) the bands of interest were excised from the polyacrylamide gel, and transferred into a tube and kept in water. The proteins were identified by mass spectrometry by the Medical Biomics Centre at St. George's, University of London, United Kingdom.

2.5.4 Western Blot.

A western blot analysis was performed to detect proteins of interest after protein electrophoresis (Methods 2.4.3.).

Following protein electrophoresis, polyacrylamide gels were transferred to a nitrocellulose membrane (GE Healthcare). The transfer chamber was assembled inside a plastic perforated cassette by stacking the polyacrylamide gel and the nitrocellulose membrane and placing them between 2 stratified layers, each of them composed internally of 2 filter papers and externally by a sponge. Protein transfer was performed was performed at 250 mA, 4ºC for 2 h, in transfer buffer (20 mM tris, 153 mM glycine, 20 % methanol). Proteins remaining in the polyacrylamide gels were visualised by Coomassie blue staining.

Proteins transferred to the nitrocellulose membrane were stained by incubating the membrane in Ponceau red solution (1 % acetic acid, 0.1 % ponceau red) for

5 minutes, after which the solution was removed and the membrane was rinsed with water.

The membrane was incubated in blot solution (0.5% of blocking reagent solution [Roche]) diluted with Tris buffer saline (50 mM Tris base 150 mM NaCl pH 7.5) for 1 h or overnight. When antibodies detecting phosphorylated proteins were utilised, the membrane was diluted in blocking solution. The membrane was incubated from 2 h to o/n at 4 ºC with the primary antibody diluted in Tris buffer saline solution (TBS) (50 mM tris-HCl pH 7.4, 150 mM NaCl, 0.1 %), plus blot solution (Roche). After hybridation, to wash the antibody in excess, the membrane was incubated 3 times at room temperature for 10 minutes each with Tris buffer saline tween (TBST) solution (50 mM tris-HCl pH 7.4, 150 mM NaCl, 0.1 %, 0.1% tween 20%).

The secondary antibody, horse radish peroxidise (HRP) conjugated, diluted in blot solution, was applied to the membrane for 1 h. Washes of 10 minutes each were performed 3 times with TBST solution to eliminate the antibody in excess. The antibodies used and their working dilutions are specified in section Methods 2.5.5. After the last wash in TBST solution, the membrane was incubated for 5 minute with enhanced chemiluminescence (ECL) solution (GE Healthcare). The solution was prepared by mixing the proprietary solution A with the proprietary solution B in equal volumes. Protein signal was detected either with the bio-imager ChemiDoc XRS (Biorad), or by photographic film. The treated membrane was exposed to a light that is able to capture the image, the exposition is for a variable period of time during which images were registered every 30 s to determine the best exposure duration, using the ChemiDoc XRS.

When using photographic film, a film was exposed to the membrane for a time variable from 30 seconds to o/n depending on the signal intensity. After exposition, the film was incubated shacking for 2-5 minutes in developer solution (Kodak), rinsed in water, and incubated shacking for 5 minutes in fixing solution (Kodak) and rinsed in water before being air-dried. Images of dried films were acquired by scanning.

The main steps followed to perform Western blot analysis are summarised in Table 2.19. Composition of different buffer and reagent used for Western blot analysis are summarised in Table 2.20.

Table 2.19: Western blot protocol.

Summary of the protocol used for western blot analysis, with main steps outlined.

Blotting	on nitrocellulose membrane transfer buffer 70V, 2h, 4 °C
Labelling	-1h or o/n in blotting solution (Roche) -2h or o/n with primary antibody in: blot solution or PBST + 2.5% Milk -3x 10 min in: TBST solution or PBST -1h with secondary antibody in: blot solution + 0.1% Tween -3x 10 min in: TBST solution or PBST
Detection	5 min in ECL solution 2s-o/n film exposition 2-5 min in developer solution rinse in H_2O 5 min in fixing solution rinse in H_2O air-drying

Table 2.20: Buffers for western Blot analysis.

Composition of the buffers and solutions employed to perform western blot analysis.

Buffer	Buffer Composition
Transfer Buffer	20mM TRIS 153mM glycine 20% Methanol
Ponceau red solution	1% Acetic Acid 0.1% ponceau red
Blot solution	0.5% of Blocking Solution (Roche) 50mM TRIS Base 150mM NaCl pH 7.5
TBST Solution	50 mM TRIS HCl pH 7.4 150mM NaCl 0.1% Tween 20
PBST Solution	PBS1X 0.1% Tween 20
PBS 2.5-5% Solution	PBS1X 0.1% Tween 20 2.5-5% Milk

2.5.5 Antibodies.

To detect proteins of interest while performing western blot analysis, different primary and secondary antibodies were used, as listed below. Table 2.21 summarises all antibody used in this study with host organism and working dilution for the techniques they were employed for.

Table 2.21: Antibodies.

Table reporting all antibodies utilised in this study and their dilution.

Antibody	Specifications	Host	WB diluition
α TDP43		rabbit	1:1000
α hnRNPA1		rabbit	1:1000
α SF2/ASF	Invitrogene 32-4500 Millipure MABE163	mouse	1:1000
α 9G8	Abcam 72616	mouse	1:1000
α Tra2β	Abcam 31353	mouse	1:1000
α 1H4	Invitrogene 33-9400	mouse	1:1000
α hnRNP L	Abcam 6106	mouse	1:2000
α Hsc70	Santa Cruz SC 7298	mouse	1:2000
α rabbit HRP	DAKO P0448	goat	1:2000
α mouse HRP	DAKO P0447	goat	1:2000

2.5.6 Quantification using Image J.

Image of films obtained following Western Blot analysis were quantified using the software Image J (http://rsb.info.nih.gov/ij/). A square box including the biggest band to quantify was drawn; the same box was placed around each of the bands to quantify. Intensity of each band was measured by calculating the integrated density over the area defined by the box. Background value was subtracted. The intensity of each band was then normalized against the intensity of normalization protein blotting (http://lukemiller.org/index.php/2010/11/analyzing-gels-and-western-blots-with-image-j/).

The splicing isoform has been calculated using Image J programme. A box was designed around each splicing isoform and the relative quantity of each isoform is presented as a percentage of the total transcript levels.

2.6 In vitro RNA method

2.6.1 RNA preparation for pull down analysis.

The pull down analysis is schematically shown in Figure 2.3. One nanomole of RNA oligonucleotide

(WT short 5'UUUCUGAGACGGAUGUAACAA3'),
(WT Long 5'UGAAUUUUCUGAGACGGAUGUAACAA3'),
(Syn 5' UUUCUGAGACAGAUGUAACAA 3'),
(Syn Long 5' UGAAUUUUCUGAGACAGAUGUAACAA 3'),

was placed in a 400 µl reaction mixture containing fresh 0.1 M NaOAc, pH 5.0, and 5 mM sodium *m*-periodate (Sigma).

The RNA oligonucleotides were treated with sodium m-*periodate* in order to oxidize its 3' end making it able to bind the adipic dehydrazide-agarose beads in a covalent manner. For each reaction the same amount of RNA control hnRNP L (contains specific binding sites for the protein hnRNP L) has been added.

Reaction mixtures were incubated for 1 h in the dark at room temperature. The RNA was ethanol-precipitated and resuspended in 100 µl of 0.1 M NaOAc, pH 5.0. Then, 300 µl of adipic acid dehydrazide-agarose bead 50% slurry (Sigma) was washed four times in 10 ml of 0.1 M NaOAc, pH 5.0, and pelleted after each wash at 4000 rpm for 5 min in a 15 ml falcon tube. After the final wash, 300 µl of 0.1 M NaOAc, pH 5.0, was added to the beads. The slurry was then divided into equal aliquots that were mixed with each periodate-treated RNA sample and incubated overnight at 4 °C on a rotator. The beads with the bound RNA were then pelleted and washed three times in 1 ml of 2 M NaCl and three times in 1.5 ml of RNA washing buffer (1M HEPES-KOH, pH 7.5, 0.5 M EDTA, 100mM dithiothreitol, v/v glycerol, 1 M KCl). They were incubated in 1X RNA binding buffer (RNA washing buffer plus 600 µg of HeLa nuclear extract, and 5µg/µl of heparin) for 30 min at room temperature in a 500 µl final volume, pelleted by centrifugation at 1000 rpm for 5 min, and washed five times with 1 ml of RNA

washing buffer. Numerous washes have been done to allow the elimination of all the nuclear proteins with no affinity for our RNA. After the final centrifugation 60 µl of SDS-PAGE sample buffer were added to the beads and heated for 5 min at 90 °C before loading the surnatant containing proteins onto a 12% SDS-PAGE gel.

The SDS-PAGE was stained using a Coomassie Blue stain, and the bands of interest were excised and identified by mass spectrometry (Methods 2.5.3). Alternatively, a western blot analysis was performed using specific antibodies to detect protein of interest.

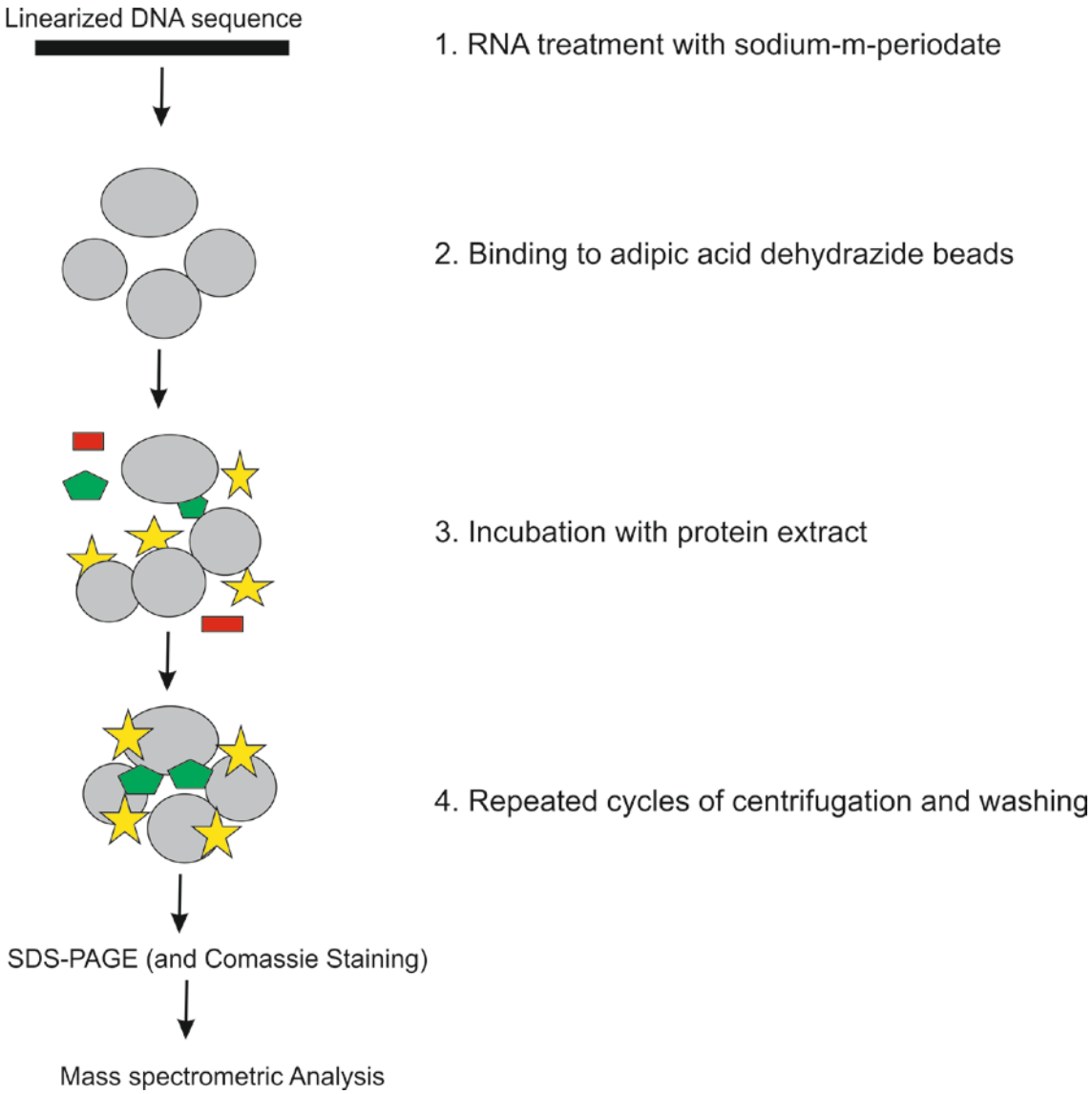


Figure 2.3: Schematic representation of the pull down analysis procedure.

The steps of the pull-down are indicated on the right of the figure and in the text. Following separation of the proteins in SDS-PAGE acrylamide gel the protein bands can be visualized by Coomassie staining or recognized by western blot analysis and incubation with specific antibodies.

2.6.2 Modified Oligonucleotides.

The sequences of the modified RNA oligonucleotide are shown in Table 2.22. Oligonucleotides Syn NT, Del1, Del 1 NT, SF2D7a, and Scramble were synthesised by Eurogenetec. The oligonucleotide Syn Oligo was a gift from Francisco Baralle, ICGEB Trieste, Italy.

Oligonucleotide abbreviations: s=phosphorothioate, o=2'-O-methyl.

Table 2.22:RNA Modified Oligonucleotide.

Summary of the modified oligonucleotide.

Bifunctional oligo	Sequence
Syn oligo	(*)A(*)G(*)G(*)A(*)GGACGGAGGACGGAGGACAmUmAmC(*)mA(*)mU(*)mC(*)mU(*)mGmUmCmU(*)mCmA
Syn NT	[U*A*C*A*U*CUG*U*C*U*C*A]
Del 1	(A*G*G*-A*G*G-ACG-GAG-GAC-GGA-GGA-CA)[C-AG*A*-A*A*A*-UUC-AC*A*-A*G]
Del 1 NT	[C*A*G*A*A*AAUUC*A*C*A*A*G]
SF2D7a	(A*G*G*A*G*GACGGAGGACGGAGGACA)[UUC*C*U*U*G*UCACUC*A*G*A]
Scramble	[A*C*C*C*U*GUCUU*A*G*G*U]

KEY
* = Phosphorothioate [ ]= 2'O Methyl m = 2'O Methyl (RNA)

2.7 Tissue culture.

2.7.1 Maintenance and analysis of cells in culture.

Cell cultures were used for transient transfection of mini-genes (Methods 2.2.1) or to prepare nuclear extracts.

Different cell lines were cultured, which included human epithelial carcinoma cells (HeLa), breast cancer cell lines (MCF7), human mammary epithelial cell (HMEC), and Lymphocyte cell line.

Cultures were initiated from cells frozen at -80 °C in freezing medium (90% fetal bovine serum (FBS) 10% DMSO). HeLa and MCF7 cells were cultured at 37ºC in standard culture medium composed by Dulbecco`s Modified Eagle Medium (DMEM) (Invitrogen) with 10% Fetal Bovine Serum (Invitrogen), 100 u/ml Penicillin (Invitrogen) and 100 μg/ml Streptomycin (Invitrogen). The Lymphocyte cell line was derived from the patient: heterozygote for the synonymous mutation c.693G>A by HPA cultures. The Lymphocyte cells were grown in suspension in RPMI 1640 Medium supplemented with 10% fetal bovine serum (Euro Clone) and 100 u/ml Penicillin (Invitrogen) and 100 μg/ml Streptomycin (Invitrogen).

A human Mammary Epithelial Cell line was grown in HMEC (HMEC Basal Serum Free Medium, HMEC Supplement, Bovine Pituitary Extract, Invitrogen).

Cells were maintained in standard growing conditions of 20% oxygen and 5% CO_2 unless otherwise stated. Cells were cultured on 75 cm^2 flask, 10 cm-Petri dishes or 6 wells plates.

Plates containing a confluent monolayer of cells were treated with Trypsin (Invitrogen) as follows. Cells washed with PBS solution, were incubated at 37° C with 1 ml of 1X trypsin solution for 2 minutes or until cells were dislodged. After adding 10 ml of media, cells were pelleted by centrifugation and resuspended in 5 ml pre-warmed medium. 1-2 ml of this cell suspension was added to 10 ml medium in a fresh plate and was gently mixed before incubation.

All cultures were handled in sterile conditions.

2.7.2 Cell counting.

The cells were counted using a haemocytometer. Two different readings were taken for each count and the average were multiplied by 1 x 10^4 to calculate the number of cells per millilitre. The average number was multiplied by 2 x 10^4 if trypan blue was added to discriminate dead cells.

2.7.3 Transfections.

The DNA used for transfection was prepared with Promega purification kit (Promega) or Fermentas kit as previously described (Method 2.2.2).
2 x 10^5 cells were plated in 6 well plates 24 h before transfection. 1.5μg or 0.5 μg of DNA was diluted in 100 μl of DMEM without FBS serum and without antibiotic, and then 4 μl of FuGene 6 (Promega) transfection reagent was added. After 15 minutes of incubation at room temperature, the transfection mix was added to the cells. Cells were incubated at 37° C, 5% oxygen and after 48h the RNA was extracted (Methods 2.3.1).

siRNA transfection was performed in 6 well plates where 2 x 10^5 cells were plated 24 h before transfection (Table 2.23). Different concentration of siRNA were diluted with DMEM without FBS and serum, 4 μl of Interferin (Polyplus transfection) was added to the diluted siRNA and after 20 minutes (incubation) the mix was added to the cells.
When the siRNA were transfected together with the DNA (mini-gene), 1.5μg or 0.5μg of DNA were diluted in 200μl of 150nM NaCl (Jet Prime buffer) to this solution 4 μl of jetPrime Reagent was added. After 15 minutes of incubation the mix was gently added to the cells.

Table 2.23 siRNA sequence.

Schematic representation of the siRNA used.

siRNA name	Sequence
Luciferase	CGUACGCGGAAUACUUCGA
SRSF1	AGUUAUGGAAGAUCUCGAU
SRSF6	1 GCAGAUCUAAGGAUGAGUA 2 AAGAUGAGGCUCUAAGGAA
SRSF9	AGAGGAUGCUAUUUAUGGA

2.8 In-silico analysis.

In silico analyses were performed to predict splicing signals using bioinformatics tools.

2.8.1 Human Splicing finder (HSF).

The Human Splicing Finder (http://www.umd.be/HSF/) is a tool to predict the effect of mutations on splicing signals or to identify splicing motifs within nucleotide sequence (Desmet et al., 2009). The splicing Finder was queried with sequence of interest to analyses mutations creating or destroying putative splicing regulatory sequences.

Chapter 3

Results

Analysis of synonymous variant c.693G>A in *BRCA1* exon 11.

3 Introduction.

It is important in diagnostic testing to look at unclassified sequence variants (UVs) and to assess their pathogenicity (Baralle et al., 2009; Tosi et al., 2010).
A proportion of these unclassified variants have been linked with an alteration of the process of splicing, however it is difficult to reliably predict their effect.
A more reliable test would be RNA analysis direct from the patient samples; however this is not always possible. For example, some genes are expressed only in a particular tissue, their regulation of splicing may differ between tissues, or RNA samples may not be available.
In this study I analysed an unclassified sequence variant which had been found in a patient with a strong family history of breast cancer, identified at the genetic diagnostic service laboratory in Salisbury. The sequence variant is a synonymous substitution in *BRCA1* exon 11 at codon position 231 (p.231 Thr>Thr; c.693G>A).
Synonymous variants change the DNA sequence without directly affecting the protein sequence; the encoded amino acid residues remain unchanged. However, it is possible for a synonymous change to mutate exonic sequences necessary for splicing regulation. In this case the synonymous variant should not be considered neutral as it may cause aberrant splicing and therefore be potentially pathogenic.

The c.693 G>A variation was chosen because it lies at codon position 231 which was reported by Hurst et al. to be in a critical region and was suggested by Orban et al. to be a particular hot spot for splicing regulation (Hurst and Pál, 2001; I. Orban and Olah, 2001).
Hurst et al. demonstrated that in the *BRCA1* gene there is an area which they termed the “critical region” between the end of BRCA1 exon 10 and the beginning of exon 11 (in particular around codons 200-300), where the ratio of synonymous substitutions - termed the Ks, is low compared with the rest of the gene, when comparing human/dog and mouse/rat (Hurst and Pál, 2001).

In the critical region different codons (codons 195, 196, 215, 231, 244 and 313), were predicted to be a hot spot for synonymous mutations; because the ratio Ka/Ks at these codon positions was higher than 1. Ka is the ratio of nonsynonymous substitutions per non-synonymous site. These hotspots may highlight regulatory sequences and therefore be important for splicing; the synonymous variant c.693G>A (on the hotspot codon 231) was analysed for its effect on splicing using different approaches which included bioinformatic analysis, RNA and mini-gene splicing assays.

3.1 Analysis of synonymous variant c.693G>A in comparison to wild type using the Human Splicing Finder website.

Bioinformatic algorithms are useful tools to predict the impact of nucleotide variants on the splicing process. Websites are available to predict the consensus sequence disrupted or created with a mutation that include ESEFinder, SpliceAid, and The Human Splicing Finder.

ESEFinder 3.0 is a web tool that analyses putative exon splicing enhancers responsive to the human SR proteins SF2/ASF (SRSF1), SC35 (SRSF2), SRSF5 (SRSF5) and SRp55 (SRSF6). It is able to predict if a mutation could disrupt or create these elements. The ESEFinder performs searches for putative ESEs by using weight matrices that are linked with the motifs for the four different human SR proteins.
These ESE elements were identified using the SELEX (Systematic Evolution of Ligands by Exponential enrichment) approach. The score is considered a high score when it is higher than the threshold value defined in the input page, the thresholds for the SR proteins are:

SRSF1 1.956, SRSF2 2.383, SRSF5 2.670, and SRSF6 2.676 (Cartegni et al., 2003).

A second bioinformatic tool available is SpliceAid. This is a collection of RNA sequences experimentally proven to bind splicing proteins (Piva et al., 2009). SpliceAid2 has also been developed over the last year updating the splicing factors binding the pre-mRNA sequence, and also updating information about tissue specificity (Piva et al., 2012).
More recently a new web site called SpliceAid-F has been developed that gives detailed information from literature about expression, functional domain, protein and chemical interactions, RNA binding and no-binding site of splicing factors (Giulietti et al., 2013).

In this thesis I have mainly used Human Splicing Finder, (http://www.umd.be/HSF/credits.html). This tool is free on-line and predicts which splicing factor motifs will be destroyed or created given a mutation. It incorporates many know algorithms used for splicing prediction (e.g. ESEFinder 3.0) and it also uses new algorithms derived from the Universal Mutation Database (UMD) (Desmet et al., 2009). Moreover, the Human Splicing Finder allows easy comparison of a mutated sequence with a wild type sequence.

Two sequences were analysed for the effect of the variant c.693G>A on splicing.

The wild type sequence:

agctgcttgtgaattttctgagacggatgtaacaaatactgaacatcat

The mutant sequence (the synonymous change c.693 G>A is shown in red):

agctgcttgtgaattttctgagacagatgtaacaaatactgaacatcat

The sequences included 24 nucleotides upstream and downstream from the variant to allow prediction of splicing regulatory sequences whose length may vary from 4 to 8 nucleotides (ESE/ESS) to more than 10 (for example the 3'

splice site). The two sequences were run in the Human splicing Finder programme choosing the option “analyse mutations” from the drop down menu.
The main differences predicted were the disruption of the binding site for SRSF1, SRSF5 and SRSF7 within the mutated sequence (Table 3.1and Table 3.2).

In presence of the nucleotide change in position c.693G>A the binding sites for SRSF1 and SRSF5 are predicted to be completely destroyed. Whereas, the two SR proteins are predicted to bind the wild type sequence with threshold respectively of 82.93 for SRSF5 and 85.09 for SRSF1 (Table 3.1). For this prediction, Human Splicing Finder used the same concept of Position Weight Matrices used in ESEFinder but in order to simplify the interpretation of results obtained it uses a normalized range of scale from 0 to 100 as default thresholds.
The Human Splicing Finder used new matrices to predict SRSF7 binding based on experimental data. Comparison between the wild type sequence and the sequence carrying the mutation c.693G>A showed that SRSF7 was predicted to bind the wild type sequence with a score value of 71.34 (where the standard threshold is 59.245), but not the mutant sequence (Table 3.2).
Therefore analysis with Human Splicing Finder suggests that the variant c.693G>A is a potential splicing mutation since it is predicted to alter the binding of positive regulatory proteins of splicing (SR proteins SRSF1, SRSF5 and SRSF7).
Therefore the patient heterozygous for the variant c.693G>A was recruited in order to collect a blood sample for further analysis.

Table 3.1: Human Splicing Finder using ESE Finder matrices.

Table shows that in position 17 and 18 of the wild type sequence begin the ESE motifs for the binding of SRSF5 and SRSF1 respectively which are predicted to be destroyed in the presence of the variation c.693G>A. The thresholds used are indicated.

Sequence Position	Linked SR Protein	Reference Motif (value 0-100)	Linked SR protein	Mutant Motif (value 0-100)	Variation
17	SRSF5	tgagacg (82.93)			Site Broken -100
18	SRSF1	gagacgg (85.09)			Site Broken -100

Enhancer motifs
ESE Finder matrices for SRSF5, SRSF2, SRSF1 and SRSF6 proteins

Thereshold values:

-SRSF1:72.98
-SRSF5:78.08
-SRSF2:75.05
-SRSF6:73.86

Table 3.2: Human Splicing Finder using experimental data.

Table shows that in position 20 of the wild sequence begins a predicted binding site for SRSF7 which is predicted to be destroyed in presence of the variation c.693G>A. The thresholds used are indicated.

Sequence Position	Linked ESE protein	Reference Motif (value 0-100)	Linked ESE protein	Mutant Motif (value 0-100)	Variation
20	SRSF7	gacgga(71.34)			Site Broken -100

ESE motifs from HSF - Experimental

Thereshold values:

-SFRS10: 75.964
-SRSF7: 59.245

3.2 RNA analysis of c.693G>A.

Whilst bioinformatic tools are available online to predict the effect on splicing of particular mutations, this is not enough evidence alone and can give false-positives or miss weaker associations. For this reason it is important to complete further investigation.

This can be done using either patient RNA (not always available) or by mini-gene analysis (Vibe-Pedersen et al., 1984). Both these approaches have been used here to confirm that the synonymous variant c.693G>A affects splicing of BRCA1 exon 11.

Firstly the effect of the synonymous variant c.693 G>A, on the splicing process, was evaluated through RNA extraction (Methods 2.3.2) from a whole blood sample of the patient and from which cDNA was made (Methods 2.3.3). The splicing products were amplified by PCR reaction using specific primers. The primers were designed to analyse the three splicing isoforms: the full length 11, (inclusion of exon 11); the Δ11q (with inclusion of partial exon 11) and the Δ11 isoforms (exon 11 is completely excluded). A schematic representation of the primers used relative to the exons and isoforms studied is shown Figure 3.1.

Comparing the splicing products resulting from the patient blood sample (patient heterozygous for the variant c.693G>A) with the normal control blood sample, a difference is evident in the proportion of the three splicing isoforms full 11, Δ11q and Δ11 (Figure 3.2).

Both normal control and patient express the three splicing isoforms. However, the patient (heterozygous for c.693G>A) shows a decrease of full 11 and Δ11q isoform and an increase of Δ11 isoform. This indicates that the synonymous c.693G>A variant affects BRCA1 splicing.

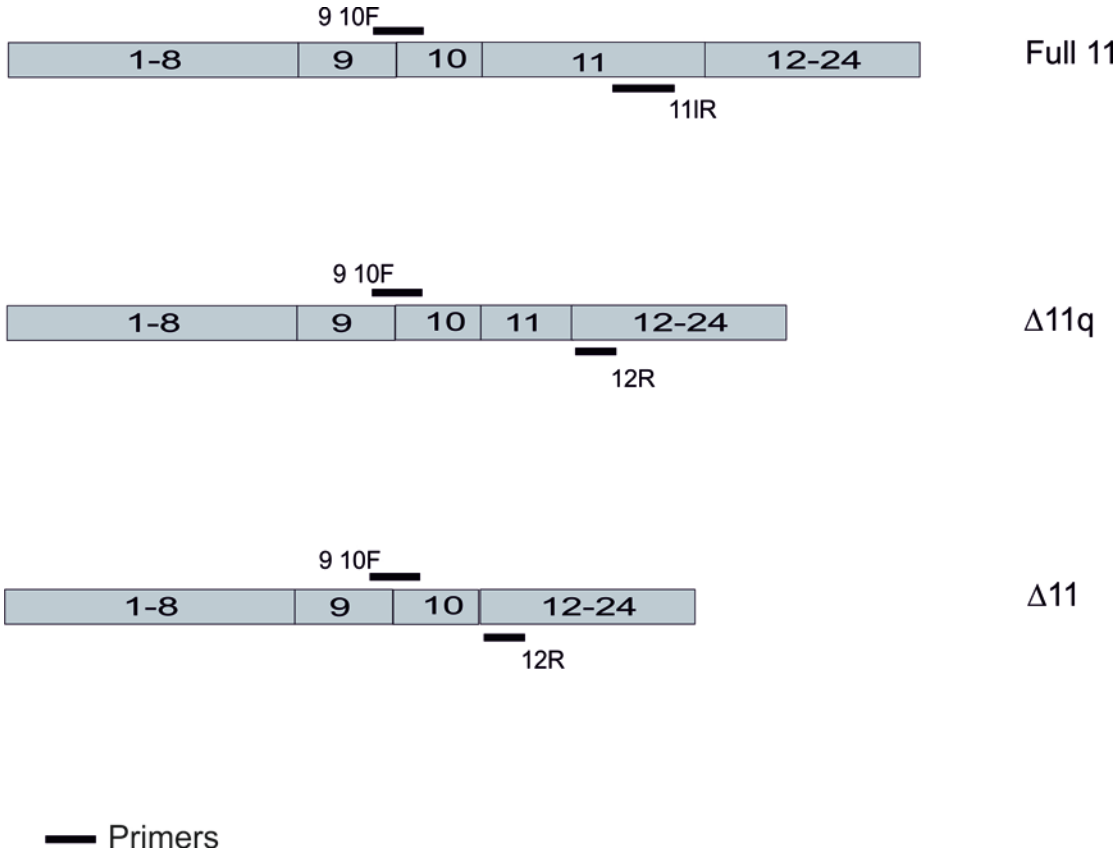


Figure 3.1: Schematic organization of the primers used for PCR amplification following RTPCR.

Diagram showing a schematic representation of the three isoforms (Full 11, Δ11q, Δ11) detected using the forward primers 9,10F and the two reverse primers 12R and 11R (black bars)

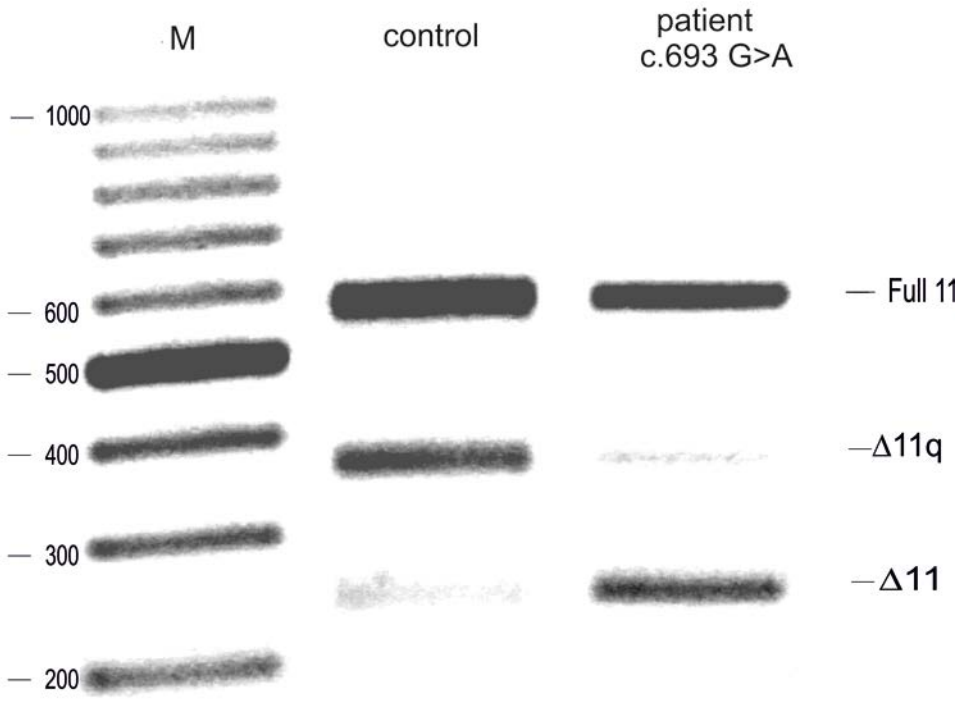


Figure 3.2: RTPCR from blood of a patient carrying the variant c.693 G>A.

Figure shows the result of RT-PCR of RNA from blood of the patient with the synonymous substitution. The PCR products were loaded on 1.5% agarose gel for electrophoresis. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11. The sizes of these bands are 572bp for the full length isoform, 374bp for the Δ11q and 297bp for the Δ11 isoforms. M is the 100bp marker; the corresponding size is reported next to each band. The experiment was performed in triplicate using each time a control sample from a different donor.

3.2.1 Development of mini-gene to analyse splicing of *BRCA1* sequence variant.

In order to test the effect of the variant in a cellular context a mini-gene for splicing assay has been used. The mini-gene assay has the advantage that sequence variants can be investigated in appropriate cell lines (for instance a case breast cancer cell line).
The mini-gene is a simple version of a gene, and contains the genomic region of interest that usually includes the alternatively spliced exons and flanking intronic sequence. A genomic region of interest containing a specific mutation that could effect splicing mechanism, can be amplified from normal and affected individuals and cloned into a mini-gene. The mini-gene is then transiently transfected into cell lines of interest, where it is transcribed by RNA polymerase II. The resulting pre-mRNA is processed to obtain the mature mRNA. A schematic representation of the mechanism of the mini-gene is shown in Figure 3.3. The mRNA splicing pattern is analysed by RT-PCR and PCR amplification with appropriate primers designed to discriminate endogenous transcript from those of the mini-gene.

An 11kb mini-gene was created that included *BRCA1* genomic sequence from exon 8 to exon 12 in the pCDNA 3(+) vector Figure 3.4. The mini-gene construct was termed pB1WT.

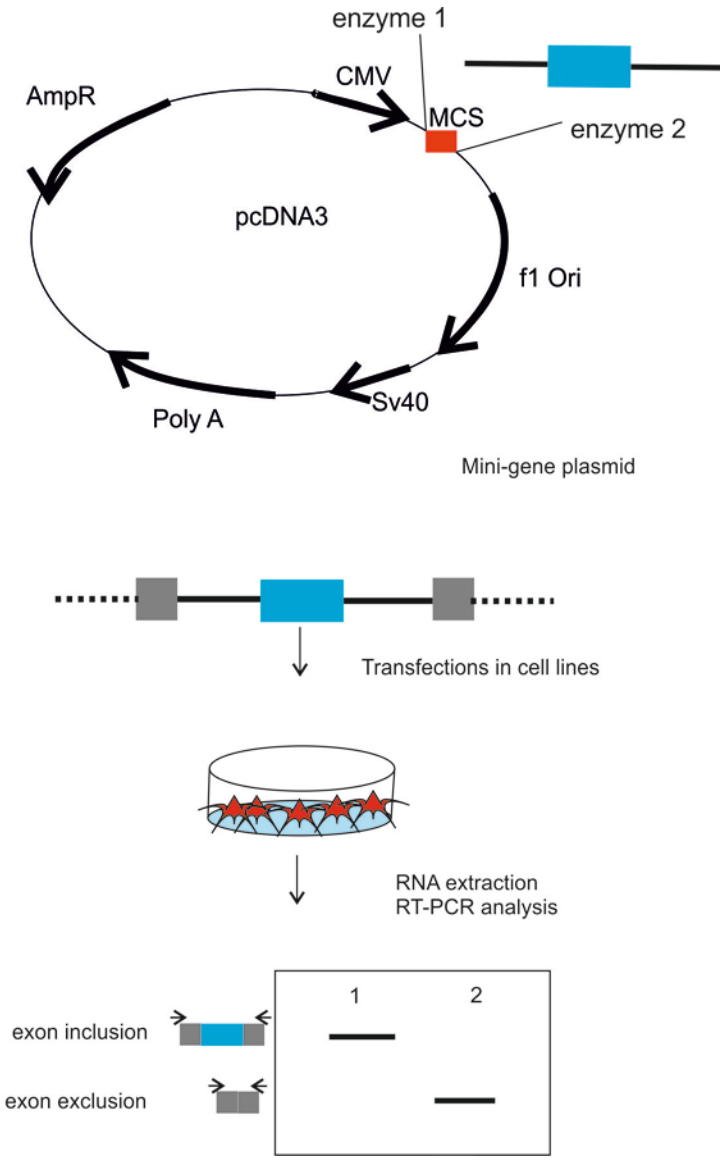


Figure 3.3: Schematic representation of a mini-gene system.

The mini-gene for splicing assay is a eukaryotic expression plasmid (for instance pCDNA3) that contains a simple version of a gene. In this case the reported mini-gene is comprised of 3 exons (shown as 2 grey boxes and 1 blue). The mini-gene is transfected in specific cell lines, followed by RNA extraction and RT-PCR and gel electrophoresis. In the example the lane 1 represents the inclusion of the "blue exon" whereas the lane 2 corresponds to its exclusion.

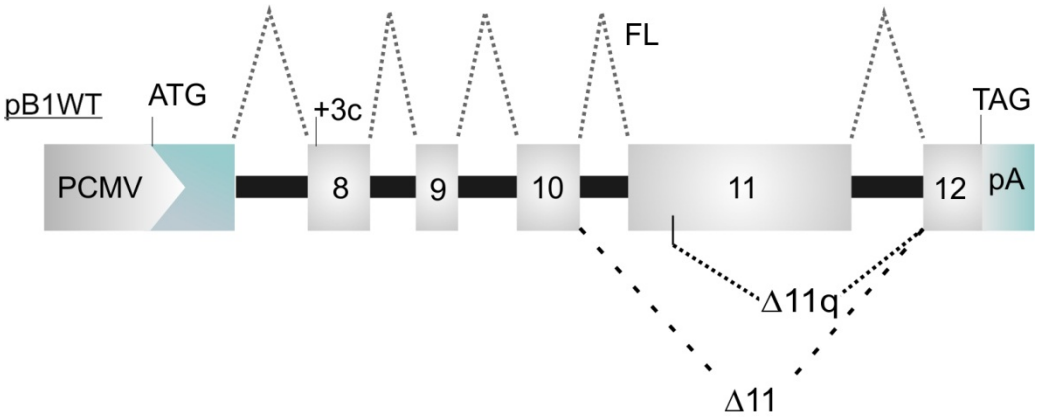


Figure 3.4: Schematic representation of the pB1WT mini-gene.

The grey boxes represent exons and black line introns. PCMV is the Cytomegalovirus promoter of the pCDNA3 vector. ATG is the start codon, TAG the stop codon and the +3c is the insertion of a cytosine into exon 8 to maintain the reading frame. Dotted lines represent the possible splicing events. pA is the polyadenilation signal.

The pB1WT mini-gene was tested for transcript to be produced for each of the three splicing isoforms (Full 11, Δ11q and Δ11) when transfected into the breast cancer cell line MCF7. Following transfection, total RNA was extracted (Method 2.3.1) and the splicing products analysed by RT-PCR and PCR amplification (Method 2.3.3).

To discriminate between splicing products originating from the mini-gene and endogenous *BRCA1*, a specific primer (pcsR) was used for the synthesis of mini-gene cDNA. This primer binds to a specific region of the mini-gene corresponding to a pCDNA3 sequence located 105 nucleotides upstream of the polyadenylation signal (Methods 2.3.3) (Figure 3.5).

The results show that the pB1 wild-type (WT) mini-gene produces all three splicing isoforms to different degrees (Figure 3.6 lane pB1 WT). Endogenous *BRCA1* expression was also analysed which showed that the pB1WT mini-gene isoform expression pattern to be similar (Figure 3.6 lane endogenous WT).

In addition the mini-gene carrying the synonymous variant c.693 G>A was tested in a splicing assay following transfection in MCF7 cells (Figure 3.5 lane pB1 c.693G>A). The results showed that transcripts with the synonymous variant had decreased of the Δ11q isoform and increase of the Δ11 isoform similar to the pattern found in the patient (Figure 3.6 and Figure 3.2). To test another cell type the mini-gene was transfected in mammary epithelial cell line (HMEC) and gave similar results (Figure 3.7). Together these results strongly suggest that the synonymous variant c.693G>A affects splicing in breast tissue.

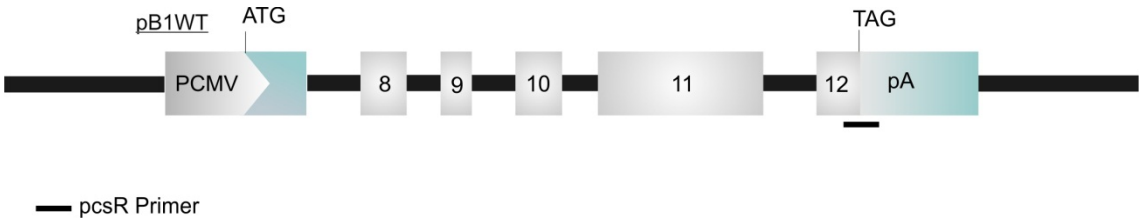


Figure 3.5: Schematic representation of the pcsR primer.

Figure show the pB1 mini-gene, the grey box represent the exons and the black lines between exons represent the introns inserted into the modified vector pCDNA3 (+). The black lines on either side are vector sequences. The pcsR primer used is represented with a short black bar.

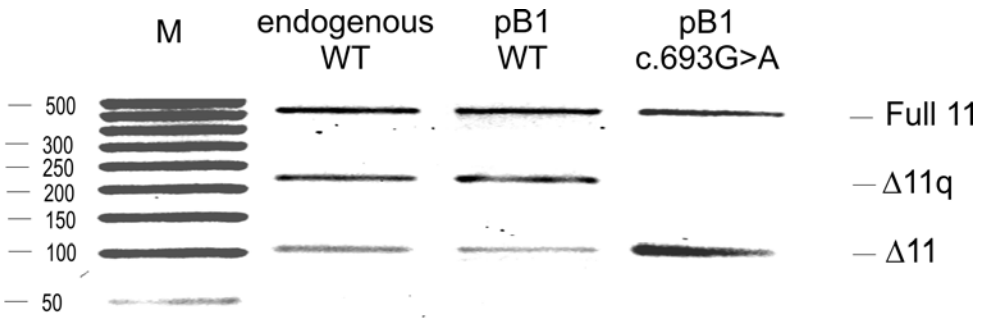


Figure 3.6: Splicing products of the endogenous *BRCA1 gene* and the mini-gene transfection in the breast cancer cell line MCF7.

The figure shows splicing products for the endogenous *BRCA1* gene, the pB1WT mini-gene and the mini-gene pB1c.693 G>A in the breast cancer cell lines, MCF7). The amplified RT-PCR products are shown on a 1.5% agarose gel. To the right of the gel is reported the isoform corresponding to each of the three bands: Full length 11 (Full 11), Δ11q, and Δ11 isoforms. M is the 50bp DNA ladder marker and the corresponding molecular weight for each band is shown on the left.

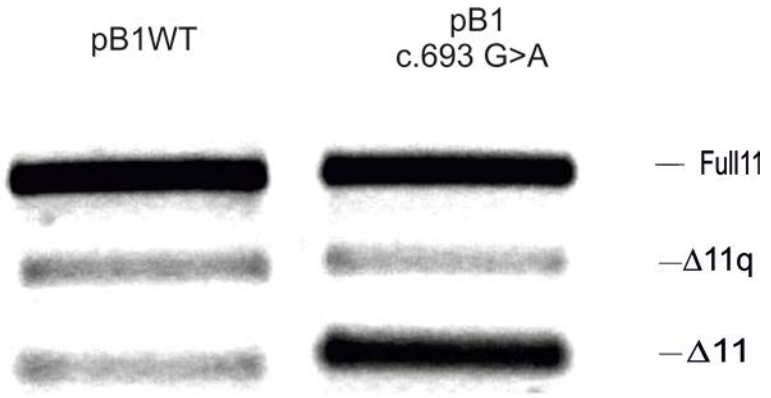


Figure 3.7: Splicing product of the mini-gene transfection in the mammary epithelial cell line (HMEpC).

The figure shows the splicing product for the pB1WT mini-gene and the mini-gene carrying the variant c.693 G>A transfected in to a mammary epithelial cell line (HMEC). On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11) the Δ11q and the Δ11. The RTPCR splicing products are shown on a 1.5% agarose gel.

3.3 Site directed mutagenesis, deletion and manipulation of a putative regulatory region.

The mini-gene splicing assay is also a useful system to study the mechanistic effect of splicing. Several studies have been able to identify splicing regulatory sequences using site directed mutagenesis of positions surrounding splicing mutations in mini-gene systems. Examples of mini-genes already used for mechanistic studies include BRCA1 exon 18 where different mutations created binding sites for silencer proteins. In particular the nonsense change c.5199G>T, initially predicted to disrupt a putative binding site for a SR protein, SRSF1, has been shown to create a silencer which binds hRNPA1 and DAZI (Deleted in Azoospermia-Associated protein 1) (Goina et al., 2008). Several other variants were detected in exon 18 of BRCA1 that cause skipping of the exon due to the creation of a silencer. One of the most common is the missense c.5123C>A (new nomenclature), creating a specific binding site for hRNPA1 and hnRNPH/F (Millevoi et al., 2010).
On the other hand, a mutation situated in an exon can also destroy a binding site for enhancers. An example of this is the presence of two synonymous substitutions in exon 5 of *PDHA1* gene (c.483 and c.498). These two mutations cause disruption of the binding site for SRSF6 exonic splicing enhancer motifs which are essential for the inclusion of exon 5 (Boichard et al., 2008).

Human diseases are often correlated with exon sequence changes, where the association with the disease phenotype is unclear. For instance nucleotide changes in exon 12 of the *CFTR* gene may cause skipping of exon 12 and are associated with cystic fibrosis. In exon 12 of *CFTR* gene site directed mutagenesis demonstrated the presence of overlapping enhancer and silencer regions which were termed the composite regulatory element of splicing (CERES) (Pagani et al., 2003b).

In order to investigate putative splicing regulatory elements that could be affected by the synonymous variant c.693 G>A, site directed mutagenesis of the region surrounding this nucleotide position (c.693) was performed (Figure 3.8).

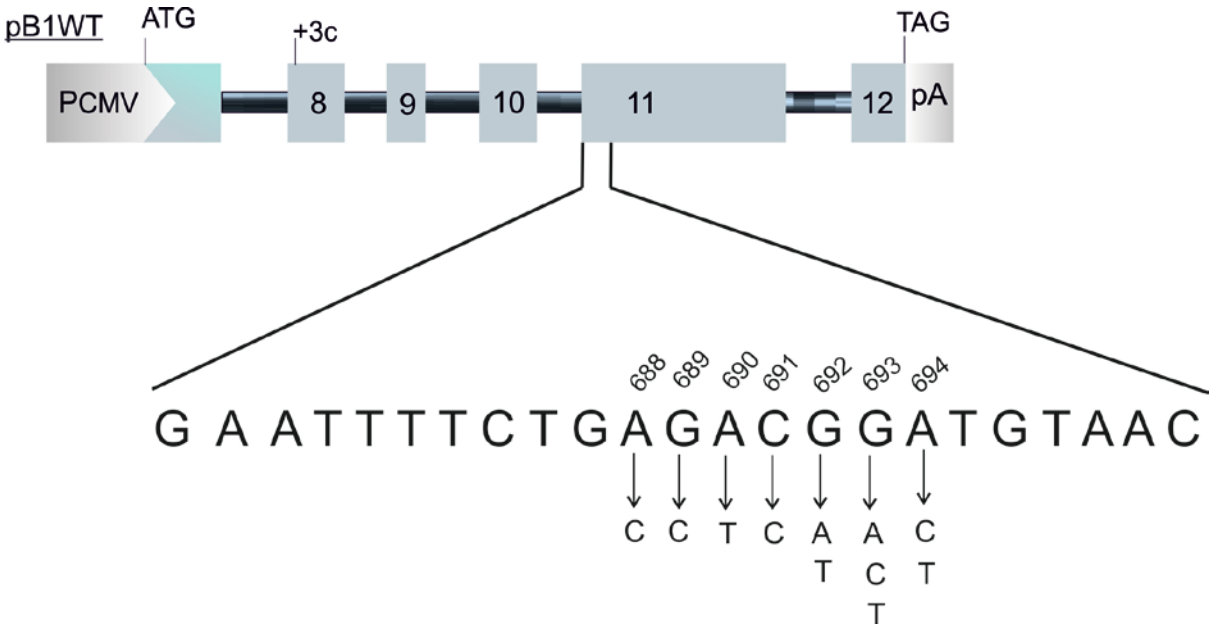


Figure 3.8: Schematic representation of the site directed mutagenesis.

The figure shows the mini-gene pB1WT and the single nucleotide changes from nucleotide c.688 to c.694 made to create mutant mini-genes.

3.3.1 Site directed and deletion analysis.

Single nucleotide changes were inserted in the pB1WT mini-gene. These were introduced from nucleotide c.688 to c.694 which included the predicted binding sites for SRSF1, SRSF5 and SRSF7. Generally, purines where changed to pyrimidine, and pyrimidine to purines.

Using the Human Splicing Finder, as already shown for the variant c.693G>A, each nucleotide change was predicted to alter the binding site for more than one SR proteins.

Following transfection of all the mutated mini-genes around the region c.688 to c.694 in MCF7 cells, the mRNA was analysed through RT-PCR and PCR amplification. The relative quantity of each isoform is presented as a percentage of the total transcript levels (Figure 3.9 shows a schematic representation).

The results of mutant mini-genes transfection are shown in Figure 3.10. The proportion of each of the three isoforms, expressed as a percentage of the total was calculated and the results show that all the variants, except for 689 A>C, show a relative increase in Δ11 isoform and decrease in Δ11q compared with WT mini-gene (Figure 3.10). The degree of change in isoforms depended on the nucleotide change involved.

These results therefore delineate a region between nucleotides c.688 and c.694 in exon 11 important for BRCA1 alternative splicing regulation.

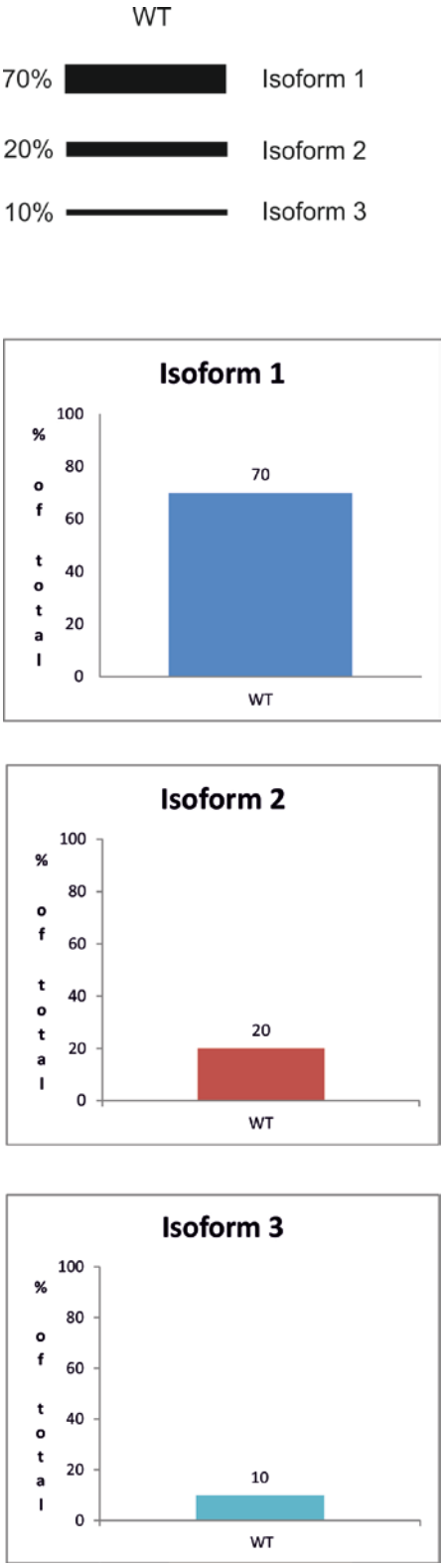


Figure 3.9: Schematic representation of the isoforms percept of the total transcript levels.

The figure shows a schematic representation of a mini-gene splicing isoforms (isoform 1, isoform 2, and isoform 3). The histograms show the percentage of isoform 1, 2 and 3 which is calculated against the total expression of the three isoforms.

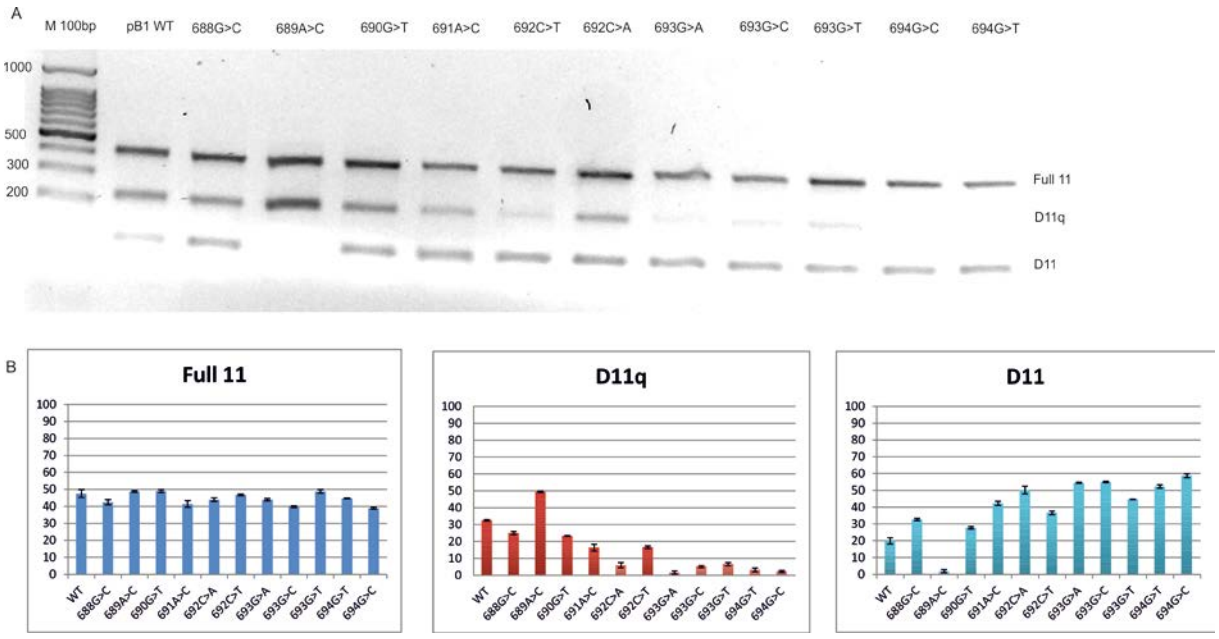


Figure 3.10: Mini-gene analysis by site directed mutagenesis of *BRCA1* exon11.

A) Single point mutation analysis of the exon 11 putative regulatory region. The constructs were transfected in the MCF7 cells, and the splicing pattern analysed by RT-PCR. Any change from position 688 to position 694 shows altered splicing process. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11. M is the 100 bp DNA ladder marker and the corresponding molecular weight for each band is shown on the left.
B) The histograms show the percentage (%) of Full 11, Δ11q and Δ11 isoform respectively, calculated against the total expression of the three isoforms. The intensity of each band (in A) was calculated using image-J (Methods 2.5.6).

In addition I have investigated if deletion of the region around the synonymous variant can affect the process of splicing.

Figure 3.11 shows that the mini-gene carrying deletion of the region between nucleotides c.688 and c.694 (lane 3) shows an increase of the Δ11 isoforms and decrease of the Δ11q isoforms compared with the pB1 WT mini-gene (lane 1). This result is comparable with the mini-gene carrying the synonymous variation c.693G>A (lane 2). A further mini-gene was created with binding site that matched one for the common SR protein, SRSF1, between position c.688 to c.694 (schematic representation in Figure 3.12). This mini-gene was termed pB1 SRSF1. In this mini-gene, the replacement of the wild-type GAGACGG sequence with the GAAGAAG sequence produced only the full length and Δ11q isoforms (Figure 3.11 lanes 4). This suggests that the replacement of a perfect binding site for SRSF1 may promote inclusion of exon 11. Therefore SRSF1 is a strong candidate for binding the region of interest and interfering with the splicing products. However, bioinformatic analysis predicted that the synonymous variant c.693 G> A could destroy binding sites for other splicing regulatory proteins, indicating that it is possible to have multiple composite exon regulatory elements distributed over the entire region of interest.

Site directed mutagenesis between the codon positions c.688 to c.694 has shown an alteration of splicing (Figure 3.10). In particular, most of the nucleotide changes caused a decrease of the Δ11q isoform and an increase of the Δ11 isoform. However in position c.689 the nucleotide change A>C caused an increase of the Δ11q and a decrease of the Δ11 isoform (Figure 3.10). For this reason a double mutant has been created in position c.689 (where the adenosine was substituted with a cytosine c.689 A>C) and in position c.693 G>A (the patient mutation) to verify whether c.689A>C could compensate for the aberrant splicing caused by c.693 G>A. Indeed the mini-gene with the double mutation (Figure 3.13 lane 2) was able to rescue the original mutant and recreate the wild type splicing pattern (Figure 3.13 lane 1).

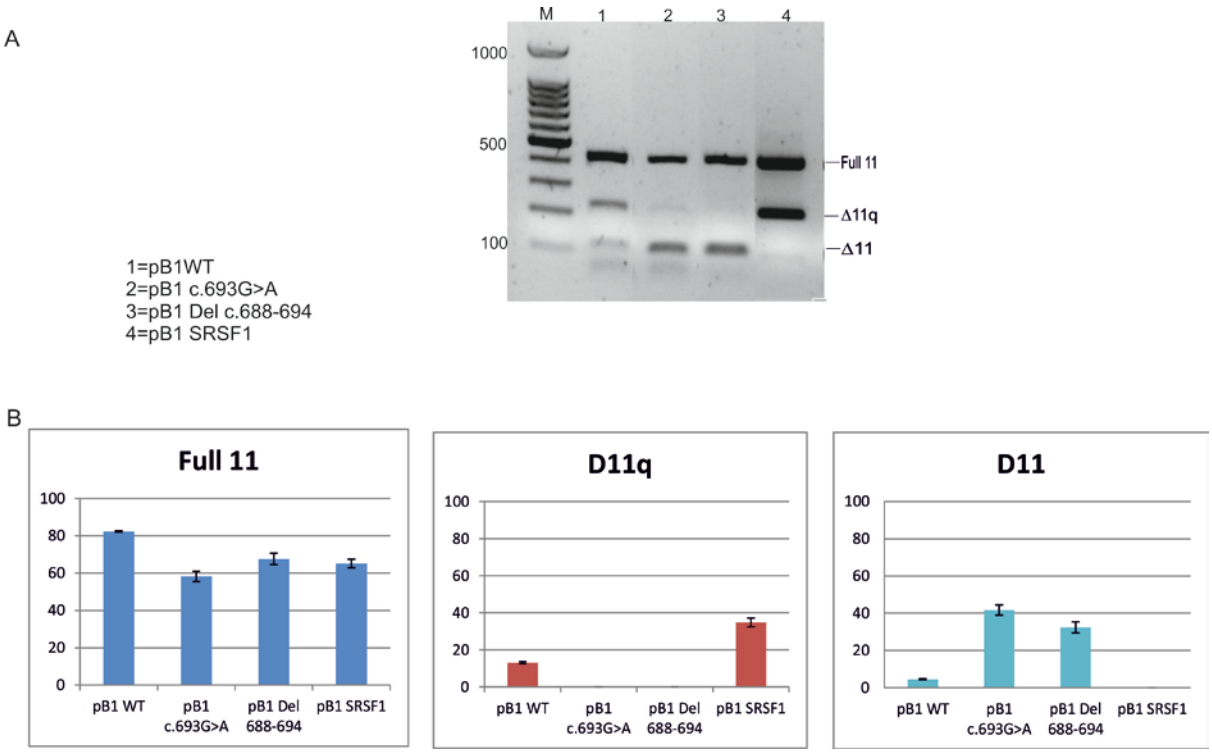


Figure 3.11: Splicing products of the mini-gene deletion c.688-c.694 and the mini-gene pB1SRSF1 transfection in the breast cancer cell line MCF7.

(**A**) The figure shows the splicing product for the pB1WT mini-gene (lane 1) the mini-gene carrying the mutation c.693 (lane 2) the pB1 mini-gene with the deletion of the region between c.688 to c.694 (lane 3) and the mini-gene pB1SRSF1 (lane 4). On the right of the gel is reported the isoform corresponding to each of the three bands: Full length (Full 11) the Δ11q and the Δ1. M is the 100bp DNA ladder marker and the corresponding molecular weight for each band is shown on the left. All the mini-genes are transfected in Breast cancer cells lines (MCF7).
(**B**) The histograms show the percentage (%) of Full 11, Δ11q and Δ11 isoform respectively, calculated against the total expression of the three isoforms. The intensity of each band (in A) was calculated using image-J (Methods 2.5.6).

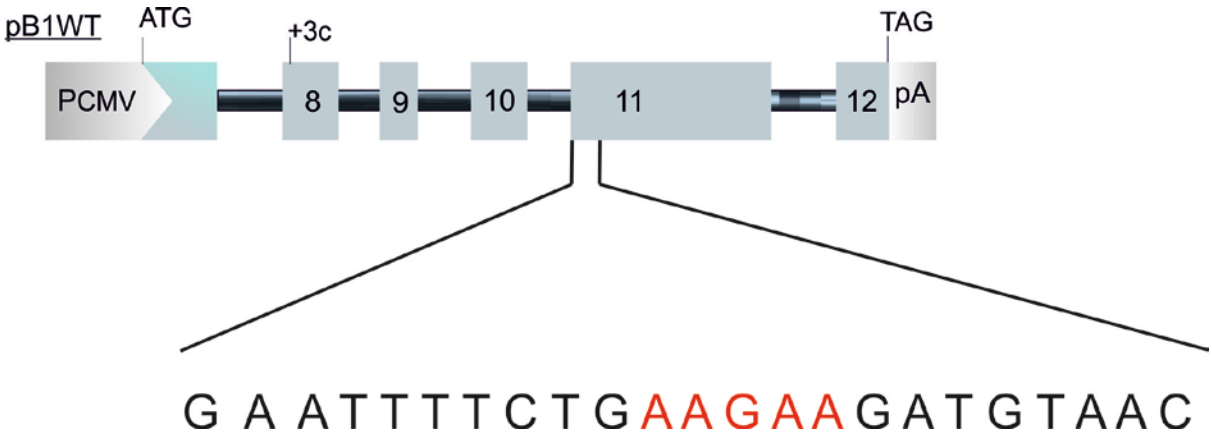


Figure 3.12: Schematic representation of the mini-gene with the binding site for SRSF1

The figure shows the mini-gene pB1WT with a sequence AAGAA (represented in red) which has been replaced in the pB1SRSF1 mini-gene.

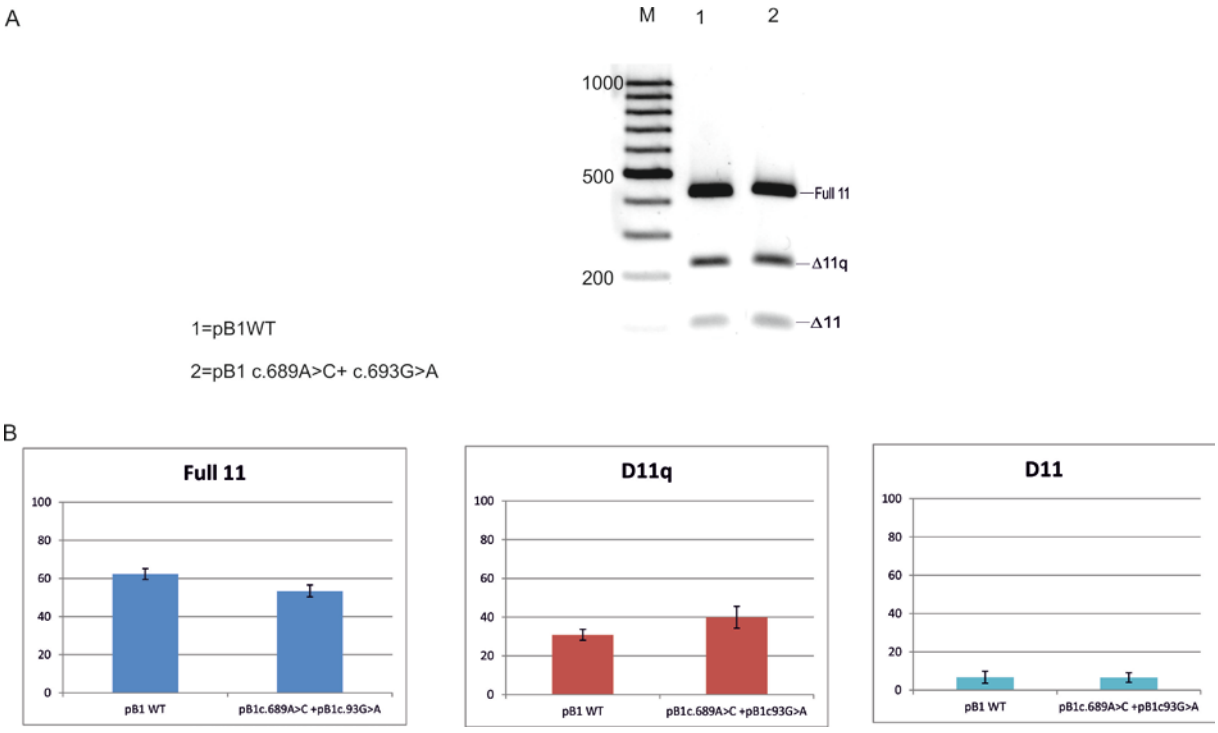


Figure 3.13: Splicing products of the double mutant mini-gene transfection in the breast cancer cell line MCF7.

(**A**) The figure shows the splicing product for the pB1WT mini-gene (lane 1) and the mini-gene with the double mutation (pB1c.689A>C+c.693G>A) (lane 2). All the mini-genes are transfected in Breast cancer cells lines (MCF7). On the right of the gel is reported the isoform corresponding to each of the three bands: Full length (Full 11) the Δ11 and the Δ11q. M is the 100bp DNA ladder marker and the corresponding molecular weight for each band is shown on the left. The RTPCR products are shown on a 1.5% agarose gel.
(**B**) The histograms show the percentage (%) of Full 11, Δ11q and Δ11 isoform respectively, calculated against the total expression of the three isoforms. The intensity of each band (in A) was calculated using image-J (Methods 2.5.6).

3.4 Conclusion.

Bioinformatic analysis using Human Splicing Finder strongly suggested that the sequence variant c.693G>A affects splicing due to the disruption of binding sites for SR proteins. This was confirmed by analysis of the RNA from the blood of the patient carrying the variant. In order to study unclassified variants when the RNA from the patient is not available or to investigate the splicing mechanism a mini-gene splicing assay can been used.

In this thesis I designed a mini-gene for a splicing assay that allows a study of unclassified variants in the *BRCA1* gene using a large region, from exon 8 to exon 12, with intronic flanking regions. I found that the synonymous variant c.693G>A affects the splicing process of the mini-gene, matching the splicing seen *in vivo* in the patient blood sample. I also demonstrate that it was possible to create a large mini-gene for the study of unclassified sequence variants.

Given the validity of the pB1 mini-gene I proceeded with the mechanistic studies which enabled the identification of a splicing regulatory region in exon 11. Human splicing finder predicted more than one splicing factor binding to the region around the synonymous variant c.693G>A, strongly suggesting the presence of an ESE in this region. Different mini-genes carrying single nucleotide changes from codon positions c.688 to c.694 suggested that there are composite regulatory elements as different nucleotide changes alter the inclusion or the exclusion of exon 11. The presence of these regulatory elements in this region makes exon 11 an interesting and complicated exon for analysis. Previous examples of exons that act in the same way are exons 9 and 12 of the *CFTR* gene. Site directed mutagenesis in *CFTR* exon 9 and exon 12 showed overlapping enhancer and silencer function; indicating the presence of composite regulatory elements (called CERES) (Pagani et al., 2003a; Pagani et al., 2003b).
However only one nucleotide change (c.689A>C) showed a decrease of the Δ11 isoform suggesting the creation of a new enhancer or the disruption of a silencer

interfering with the recognition of the 3’ splice site. For this reason an experiment of a double mutation was undertaken to understand if the mutation in c.689A>C could compensate for the presence of the synonymous variant c.693G>A.

The double mutant at codon positions c.688 and c.693 demonstrated that the substitution c.689A>C can restore inclusion of exon 11 caused by the synonymous substitution c.693G>A. This result is important because it suggests that the effect of the variant c.693G>A could be corrected by recruitment of enhancer protein or by masking the binding sites for a silencer protein. Identification of these proteins would be important to understand the mechanism underlying the altered splicing associated with c.693C>A.

Chapter 4

Results

Identification of binding proteins.

4 Introduction.

Data presented in chapter 3 demonstrated that the synonymous sequence variant c.693 G>A in a patient with a strong family history of breast cancer, is involved in the alteration of the splicing process. In particular *in silico* prediction using The Human Splicing Finder tools suggested that c.693G>A disrupts the binding sites for different SR proteins including SRSF5 (SRp40), SRSF1 (SF2/ASF) and SRSF7 (9G8) and creates a binding site for hnRNPA1. Results of site directed mutagenesis also suggested that c.693 G>A disrupts binding sites for enhancer proteins.

Binding of splicing regulatory proteins to RNA can be demonstrated using various techniques which are able to compare binding between a wild type sequence and mutant sequences. The techniques used are Electrophoretic Mobility Super-shift Assay (Super shift-EMSA), UV cross linking and Pull down analysis.

EMSA is a low cost technique to analyse the interactions between RNA and proteins. It is gel electrophoresis under native conditions where the interaction between the RNA and a recombinant protein is detected using radioactively labelled RNA. The protein is separated based on size, charge and conformation. In the Super shift-EMSA, an antibody is added that is specific for the RNA-binding protein. If the protein is binding the RNA, then the antibody against this specific protein can bind to the complex and a shift in mobility is detected in the gel. With this technique prior knowledge is required as to the likely protein binding the sequence of interest in order to choose the appropriate antibody.

UV cross-linking assay usually uses ^{32}P radiolabeled RNA substrate which is incubated with proteins (E.g. HeLa nuclear extract) to allow complex formation. The complex is irradiated with UV light to cross-link the RNA to its binding proteins and the remaining free RNA is digested with RNaseA. The cross-linked complex is analysed by SDS-PAGE and autoradiography. An advantage of this

assay is that proteins that bind to the RNA, either known or unknown can be identified on the basis of the molecular weight of the proteins.
In a pull down assay, RNA is bound on to agarose beads. The beads are pulled down by centrifugation together with RNA and its binding proteins. These complexes are denatured and the proteins are separated on SDS-PAGE acrylamide gel and visualized by Coomassie staining. The advantage of this technique compared with the other methods is the possibility of analysing the interaction between RNA and proteins without using radioactive RNA; in addition it is possible to identify unknown proteins binding the RNA of interest through mass spectrometry and to confirm binding by western blot analysis.

4.1 Identification of nuclear proteins binding to the splicing regulatory element of BRCA1 exon 11 by pull down analysis.

In order to verify that the enhancer proteins identified by the i*n silico* analysis, specifically interact with the regulatory sequence identified (Chapter 3); I performed a pull down analysis (Methods 2.6.1).
The binding of proteins from a HeLa cells nuclear extract to a synthetic RNA oligonucleotide spanning a region of 21 nucleotides around position c.693 (wild type oligo), was compared with binding to the same synthetic RNA oligo carrying the variant c.693 G>A (c.693 G>A oligo). The sequences of the RNA oligos are shown in Figure 4.1

Two bands were detected in the WT lane that were not seen in the c.693G>A lane (Figure 4.1). Each band was isolated and analysed by mass spectrometry analysis (Methods 2.5.3) to identify the proteins. For each band, a corresponding slice of gel (at the same molecular weight) was also isolated from the negative control line and analysed as control (Figure 4.2). As reported in Table 4.1 mass spectrometry

analysis on the nine isolated bands revealed differential binding of TDP43, SRSF9 (SRp30c) and hnRNPA1.

Differential binding of TDP43 was detected between wild type and mutant (c.693G>A) oligos. 22 peptides corresponding to TDP43 were identified in the wild type sample and only 8 peptides in the synonymous sample with a final protein score of 626 and 343 respectively. 13 hRNPA1 peptides were identified in the wild type sample and 23 peptides in the synonymous sample with a final score of 410 and 841 respectively Table 4.1.
In addition mass spectrometry analysis identified binding of SRSF9 in the WT lane (14 peptides) with a final protein score of 398 but not to the mutant oligo.

Identification of proteins by mass spectrometry analysis is considered significant if more than one matching peptide is recognized and if the protein score calculated is greater than 30. The results therefore indicate that TDP43 and hnRNPA1 are binding both the wild type and the mutated oligo, and SRSF9 is binding the WT RNA oligo only.

However, the score for TDP43 is much lower for the c. 693 G>A sample compared to WT. Also the score for hnRNPA1 is much lower in the WT sample compared to c.693 G>A. In order to determine the relative binding of these proteins to the RNA oligos WT and c.693 G>A, I performed a western blot analysis comparing with binding to hnRNP L. An RNA oligo with the sequence CACACACA which is able to bind hnRNP L has been added to the WT and the synonymous c.693G>A RNA oligo samples during the pull down assay in order to estimate the amount of protein pulled down and allow normalisation.

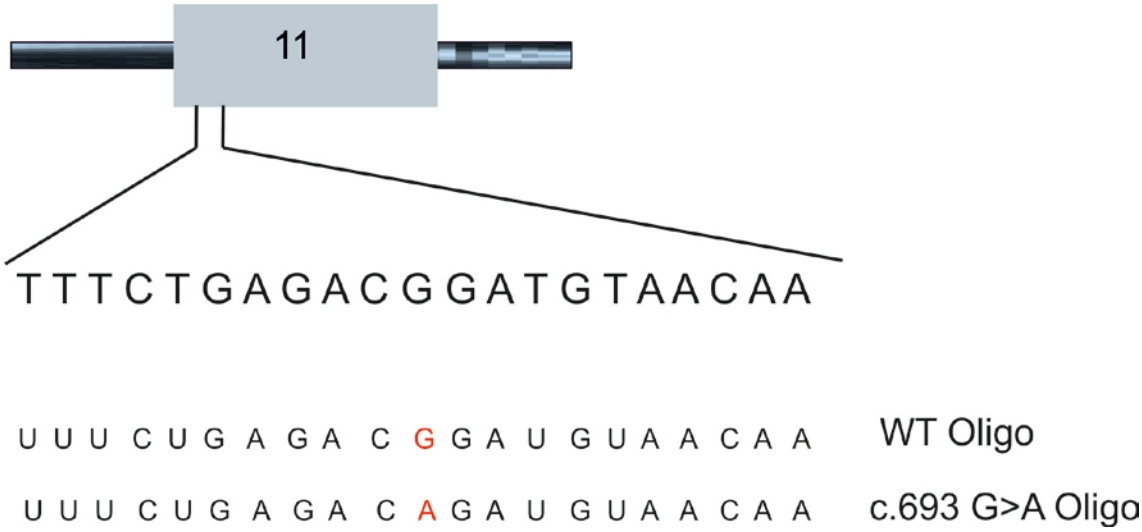


Figure 4.1: Schematic representation of the RNA oligo.

Figure shows the exon 11 and the two RNA oligonucleotides used for pull down assay. The WT oligo and the c.693 G>A oligo carrying the variant (showed in red).

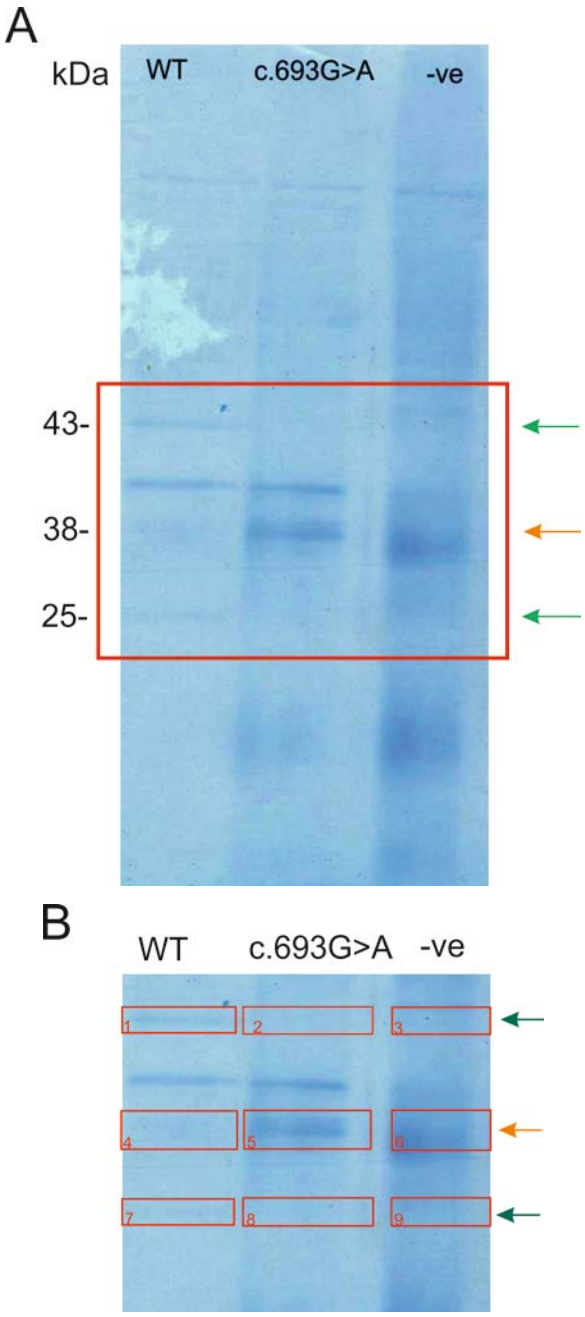


Figure 4.2: Pull down binding assay showing bands with different intensity.

Coomassie blue staining of a 12 % polyacrylamide gel.
The lanes WT and c.693G>A contain proteins purified from pull down assay. The lane –ve shows a negative control of beads only, with no RNA.
(A) Pull down binding assay. The red box indicates the selected region of interest where differences in bands intensity were detected. Two bands (indicated by green arrows) are visible in the WT lane but not in the c.693G>A lane. One band

(indicated by the orange arrow) is visible in the c.693G>A lane but not in the WT lane.
(B) Reproduction of the selected region from the Pull down binding assay shown in panel A. Red boxes (1-9) indicate the nine bands of interest isolated from the gel and analysed by mass spectrometry.

A

Band	Lane	Protein visible	Protein identified
1	WT	yes	TDP43
2	c.693G>A	no	TDP43
3	-ve	no	-
4	WT	no	hnRNPA1
5	c.693G>A	yes	hnRNPA1
6	-ve	no	-
7	WT	yes	SRSF9 (SRp30c)
8	c.693G>A	no	-
9	-ve	no	-

B

Lane	Protein	Number of Hits	Score
WT	TDP43	22	626
c.693G>A	TDP43	8	343
WT	hnRNPA1	13	410
c.693G>A	hnRNPA1	23	841
WT	SRSF9 Srp30c	14	398
c.693G>A	SRSF9	0	0

Table 4.1: Identification of the proteins in the wild type and c.693 G>A variant samples

(**A**) Table summarising the results from mass spectrometry analysis. For each sample (band), the gel fragment number (referring to Figure 4.2) and the proteins identified by mass spectrometry in the sample are indicated.
(**B**) Table summarising the mass spectrometry results for the proteins identified in sample 1, 5 and 7. For each protein, the number of matching peptides (number of hits) and the protein score are indicated.

4.2 Identification of regulatory proteins binding RNA by western blot analysis.

In order to verify TDP43 binding, I performed a pull down analysis (Method 2.10.1) followed by western blot (Method 2.5.4) using the specific polyclonal antibody anti- TDP43.

The results shown in Figure 4.3 confirmed that the WT RNA oligo and the RNA oligo carrying the synonymous c.693 G>A variant both have the ability to bind TDP43. The TDP43 antibody detected the presence of TDP43 in both the WT and the c.693G>A lane (Figure 4.3 A), although the intensity of the TDP43 band relative to the loading control hnRNP L is higher in the WT lane (Figure 4.3 B).

Mass spectrometry analysis also identified binding of hnRNPA1 to the synonymous c.693G>A variant. In order to verify binding of hnRNPA1 a pull down analysis followed by western blot has been undertaken using a specific polyclonal antibody to hnRNPA1. As shown in Figure 4.4 the antibody against hnRNPA1 detects hnRNPA1 binding to the wild type RNA oligonucleotide, as well as to the oligonucleotide with the synonymous variant c.693G>A, although the intensity of the hnRNPA1 band relative to the loading control hnRNP L is higher in the c.693G>A lane.

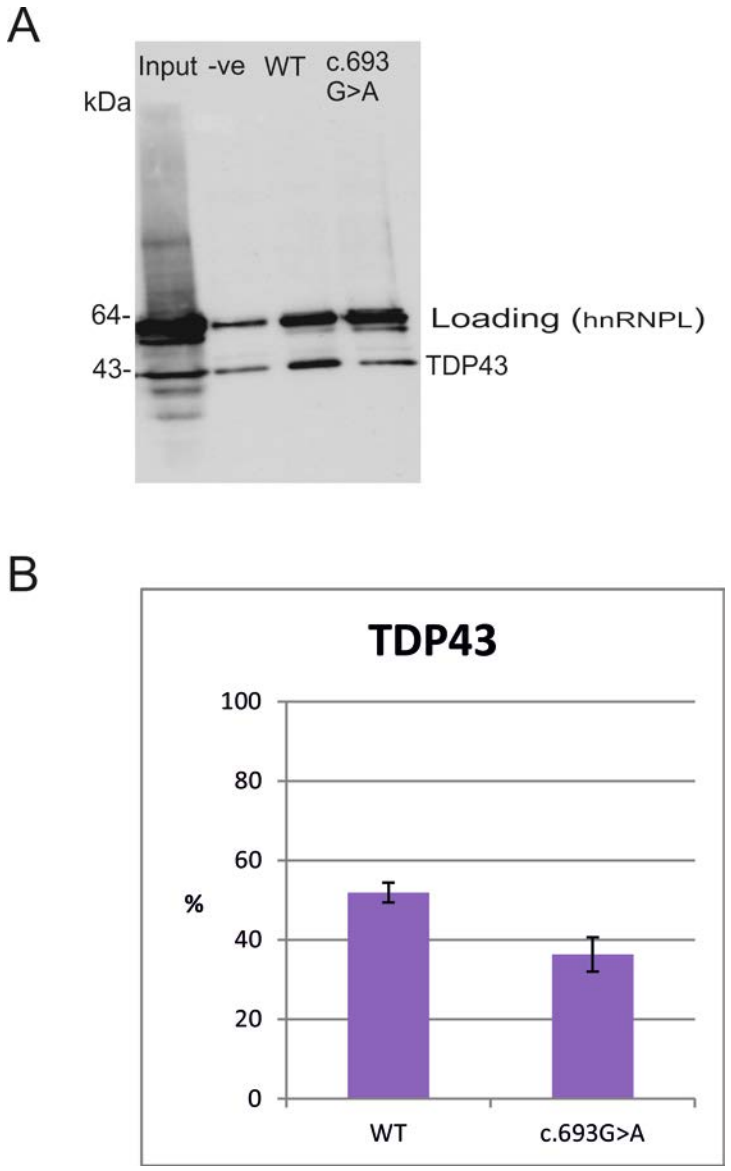


Figure 4.3: Western blot of pull down analysis of BRCA1 oligo WT and synonymous c.693 G>A variant to determine the presence of TDP43.

(**A**) Western blot of pull down samples using specific antibodies against TDP43 and hnRNP L. The first lane (input) contains 1/10 of input nuclear extract. The lane –ve shows a negative control of beads only, with no RNA. WT and c.693G>A lanes contain proteins purified from pull down assays.
(**B**) The histogram shows the percentage (%) of intensity of the TDP43 band calculated against the loading (intensity of the hnRNP L band) using Image J (Methods 2.5.6). The percentages and standard deviation (error bars) are calculated from two biological replicates.

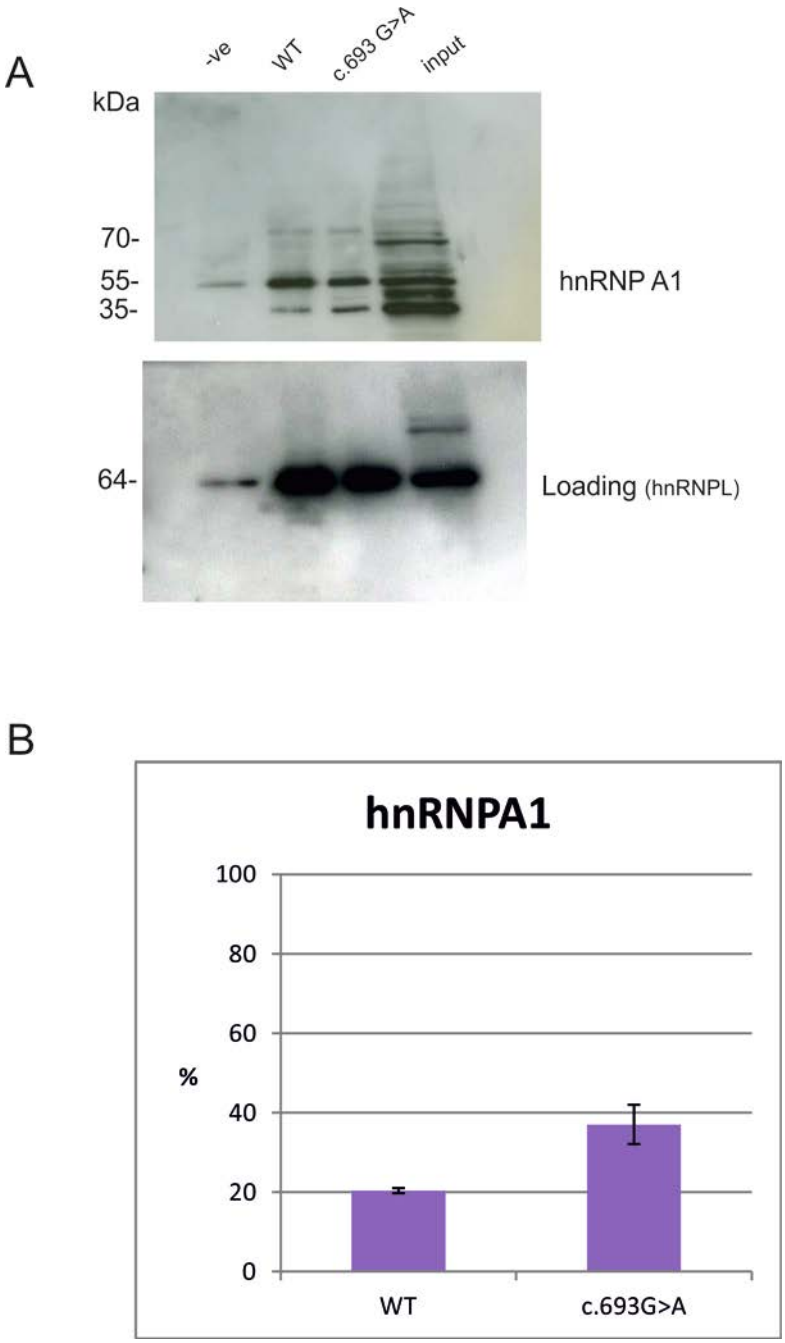


Figure 4.4: Western blot of pull down analysis of BRCA1 oligo WT and synonymous c.693 G>A variant to determine the presence of hnRNPA1.

(**A**) Western blot of pull down samples using specific antibodies against hnRNPA1. The first lane (-ve) is the negative control of beads, with no RNA. WT and c.693G>A lanes contain proteins purified from pull down assays. Input contains 1/10 of input nuclear extract.
(**B**) The histogram shows the % of intensity of the hnRNPA1 band calculated against the loading (intensity of the hnRNP L band) using Image J (Methods

2.5.6). The percentages and standard deviation (error bars) are calculated from two biological replicates.

The bioinformatic approach initially suggested that the synonymous c.693G>A variant destroyed binding sites for SRSF1, SRSF5 and SRSF7 (Table 3.1 and Table 3.2). However, it has not been possible to demonstrate binding of these proteins to the RNA oligos with pull down analysis and Coomassie staining. As these low molecular weight proteins can be easily identified by combining a western blot analysis with the pull down, I used this approach for further confirmation. (Methods 2.5.4). Antibodies to SRSF1, SRSF7 and 1H4 (detects phosphorylated SR proteins including SRSF5) were available and used to detect binding to WT and mutant RNA oligos.

The western blot result using anti SRSF1 antibody shows that SRSF1 is able to bind the wild type RNA sequence, but not the RNA sequence with the synonymous variant c.693 G>A (Figure 4.5).

The pull down analysis followed by western blot using the 1H4 antibody, demonstrated that SRSF6 and SRSF9 are able to bind the RNA WT sequence only (Figure 4.6); while binding of SRSF5 was not detected.

These results suggest that SRSF1 SRSF6 and SRSF9 are the splicing factors binding the regulatory region in exon 11. The synonymous variant c.693G>A disrupts this binding capacity which in turn may cause the altered splicing observed in the patient.

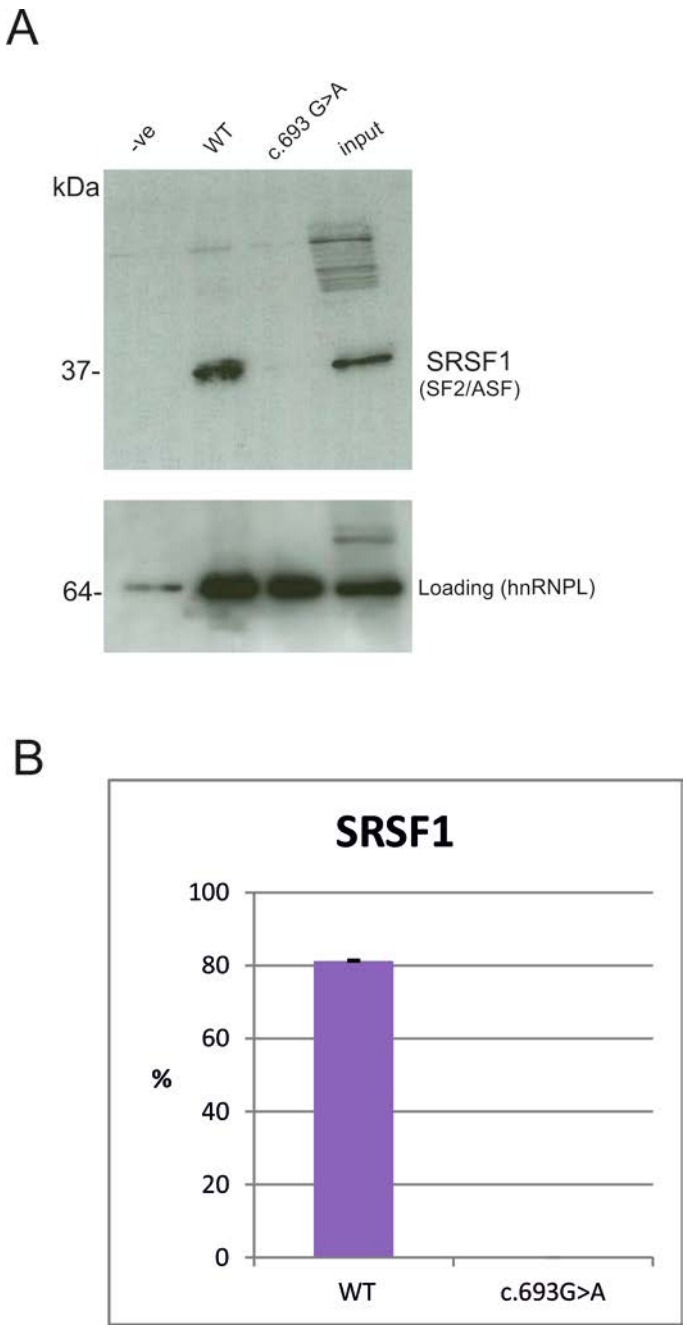


Figure 4.5: Western blot of pull down analysis of BRCA1 oligo WT and c.693 G>A variant to determine the presence of SRSF1.

(**A**)Western blot of pull down samples using SRSF1 antibody (upper panel) and hnRNP L (lower panel). hnRNP L represents the loading control. The –ve lane represents the control sample from beads only with no RNA. WT and c.693 G>A lanes contain proteins purified from pull down assays. The input lane contains 1/20 of input HeLa nuclear extract.

(**B**) The histogram shows the % of intensity of the SRSF1 band calculated against the loading (intensity of the hnRNP L band) using Image J (Methods 2.5.6). The percentages and standard deviation (error bars) are calculated from two biological replicates.

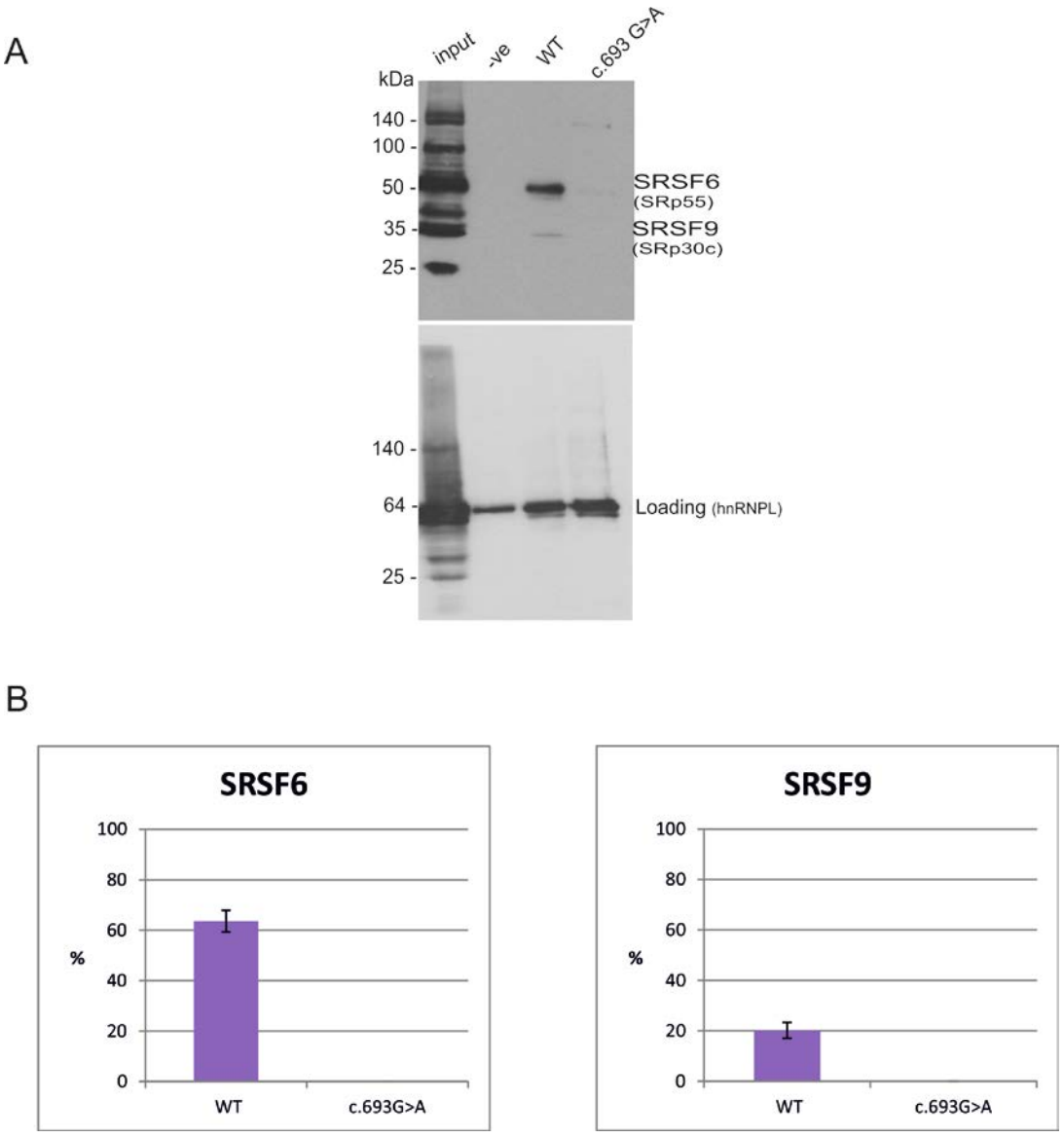


Figure 4.6: Western blot of pull down analysis of BRCA1 WT and c.693 G>A variant to determine the presence of SR proteins.

(**A**)Western blot of pull down samples using 1H4 antibody (upper panel). This antibody is able to detect several SR proteins (input lane) but only two bands corresponding to SRSF6 and SRSF9 are visible in the WT lane. The band detected with the hnRNP L antibody (lower panel) represents the loading control. The input lane contains 1/10 of input HeLa nuclear extract. The –ve lane represents the control sample from beads only with no RNA. WT and c.693 G>A lanes contain proteins purified from pull down assays.
(**B**) The histograms show the % of intensity of the SRSF6 and SRSF9 bands calculated against the loading (intensity of the hnRNP L band) using Image J (Methods 2.5.6). The percentages and standard deviation (error bars) are calculated from two biological replicates.

The Human Splicing Finder predicted that in the presence of the synonymous variant c.693G>A the binding site for SRSF7 is destroyed. Using an antibody against SRSF7 it was not possible to detect any binding either in the wild type oligo or the RNA oligo with variant c.693G>A (Figure 4.7).

It has been shown in the literature that the binding site of SRSF7 to the RNA may require binding of accessory SR proteins and in particular SFRS10. Therefore, using The Human Splicing Finder, I checked for potential binding sites for SFRS10 near the predicted binding sequences for SRSF7 in BRCA1 exon 11. The Figure 4.8 shows that binding sequences for SFRS10 are predicted to be downstream of the binding sequences for SRSF7.

The RNA oligonucleotide that has been used for the pull down analysis covered a region of 21 nucleotides that are not enough to cover the binding site of both SRSF7 and SFRS10. For this reason another RNA oligonucleotide was then used, spanning a longer region of RNA (a schematic representation of the RNA oligo is shown in Figure 4.9).

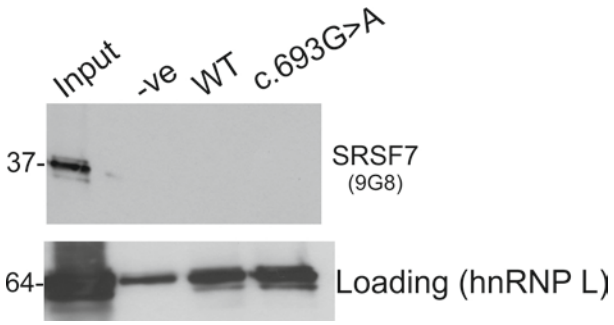


Figure 4.7: Western blot of pull down analysis of BRCA1 WT and c.693 G>A variant to detect SRSF7.

Western blot of 12 % polyacrylamide gel from pull down samples using RNA short oligonucleotide WT or carrying the variation c.693G>A. The –ve lane represents a control sample from beads only with no RNA. WT and c.693 G>A lanes contain proteins purified from pull down assays. The input lane contains 1/10 of input HeLa nuclear extract. The band detected with the hnRNP L antibody (lower panel) represents the loading control.

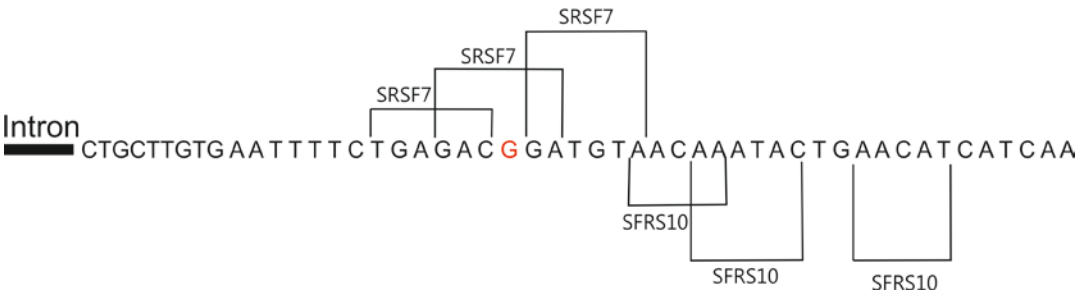


Figure 4.8: Schematic representation of the binding sequences for SRSF7 and SFRS10.

The figure shows the beginning of exon 11 with flanking introns. In red is the synonymous variation c.693G>A. The square lines indicate the predicted binding sites for SRSF7 and for SFSR10 along exon 11.

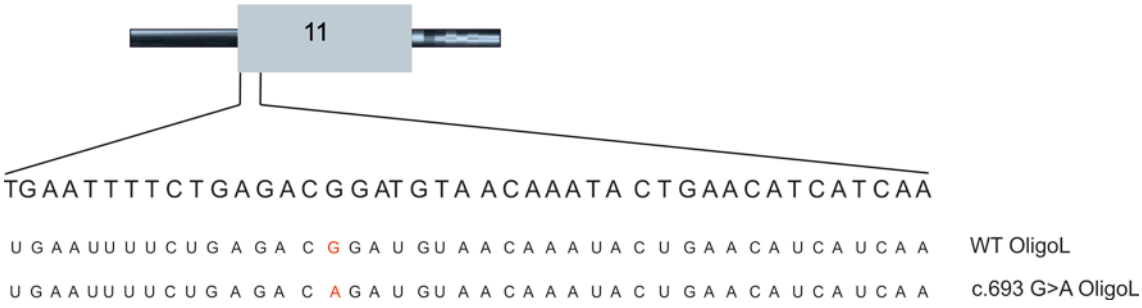


Figure 4.9: Schematic representation of the synthetic RNA long oligos.

The figure shows exon 11 with flanking introns and the two RNA Long oligonucleotides used for the pull down analysis and the western blot analysis. The WT Oligo L and the c.693G>A oligo L carrying the variant. In red is the synonymous variation c.693G>A.

The pull down analysis using the long RNA oligo for the RNA WT and the RNA carrying the c.693G>A variant was performed. Western blot analysis has been done using a polyclonal antibody against SRSF7 which demonstrated that the antibody is not able to detect the presence of the SRSF7 protein both in the wild type and the c.693G>A lane (Figure 4.10).

On the other hand the SFRS10 antibody is able to detect the presence of the SFRS10 protein in both the wild type and the c.693G>A lane (Figure 4.11).

In conclusion, western blot analysis demonstrated that SRSF1, SRSF9 and SRSF6 are able to bind wild type sequence but are not able to bind the sequence with the variant c.693G>A. The western blot analysis also demonstrated, that TDP43 and hRNPA1 are able to bind wild type and the c.693G>A variant with different affinities and that SRSF10 is able to bind a region downstream position c.693 both in presence or absence of the c.693G>A variant.

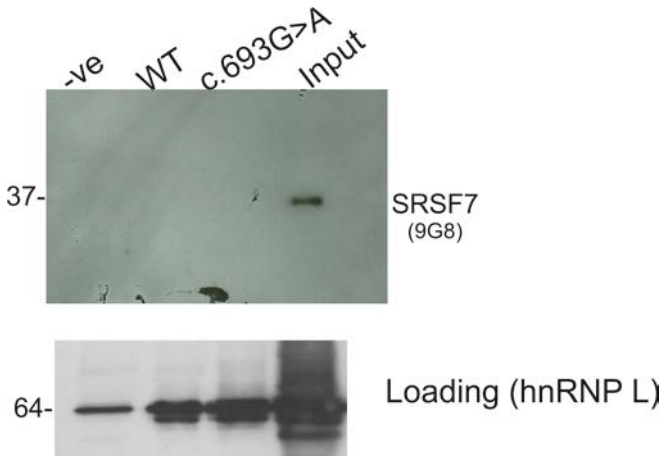


Figure 4.10: Western blot of pull down analysis of BRCA1 WT and c.693 G>A variant to detect SRSF7.

Western blot of 12 % polyacrylamide gel from pull down samples using RNA long oligonucleotide WT or carrying the variation c.693G>A. The –ve lane represents a control sample from beads only with no RNA. WT and c.693 G>A lanes contain proteins purified from pull down assays. The input lane contains 1/10 of input HeLa nuclear extract. The band detected with the hnRNP L antibody (lower panel) represents the loading control.

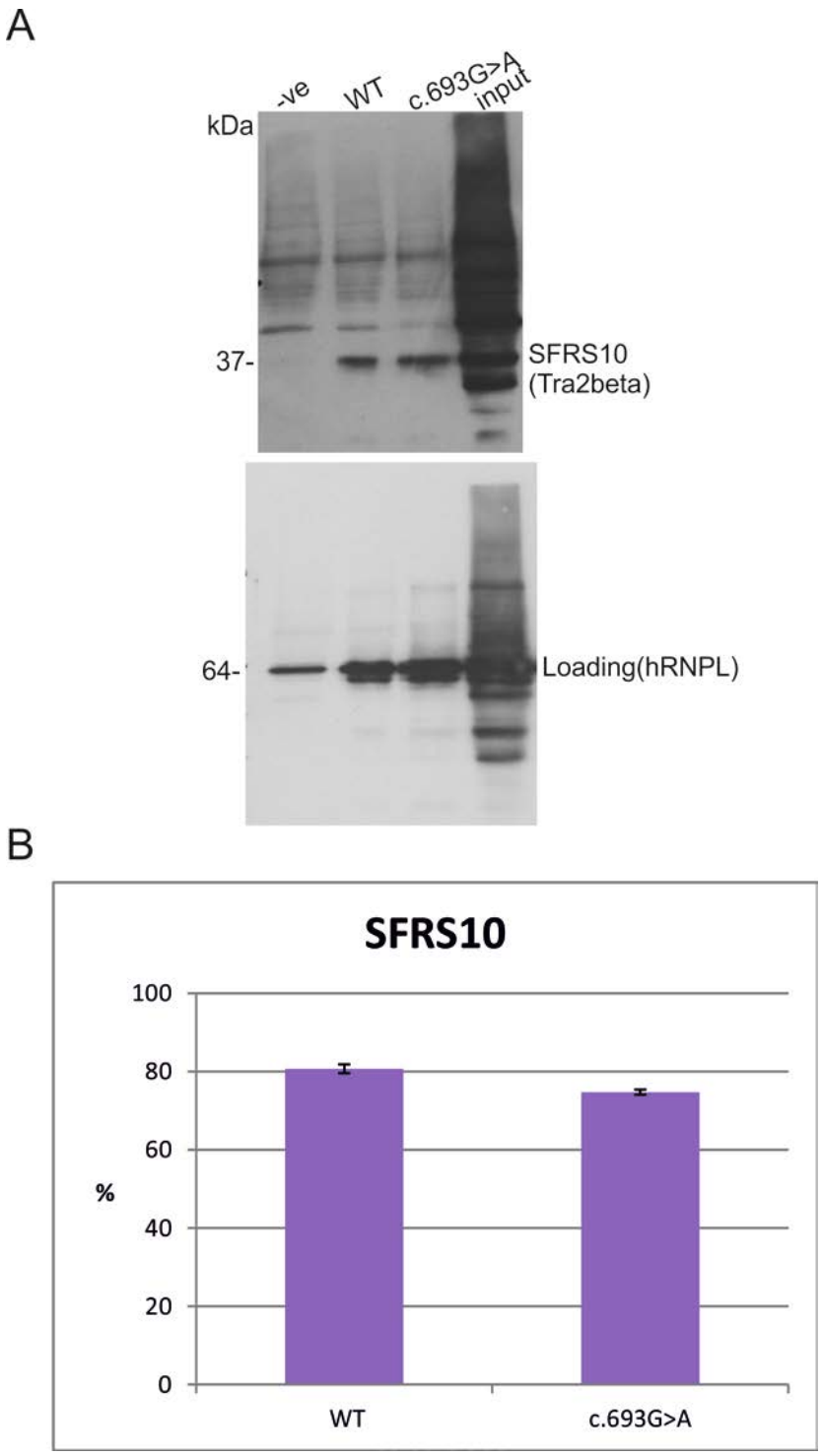


Figure 4.11: Western blot of pull down analysis of BRCA1 WT and c.693 G>A variant to detect SFRS10.

(**A**)Western blot of 12 % polyacrylamide gel from pull analysis, the antibody used are the SFRS10 antibody (upper panel) and hnRNP L antibody (lower panel).

hnRNP L represents the loading control. The –ve lane represents at the control sample from beads only with no RNA. WT and c.693 G>A lanes contain proteins purified from pull down assays. The input lane contains 1/10 of input HeLa nuclear extract.
(**B**) The histogram shows the % of intensity of the SFRS10 band calculated against the loading (intensity of the hnRNP L band) using Image J (Methods 2.5.6). The percentages and standard deviation (error bars) are calculated from two biological replicates.

4.3 Exploring the functional role of SRSF1, SRSF6, and SRSF9 in the regulation of BRCA1 exon 11 splicing.

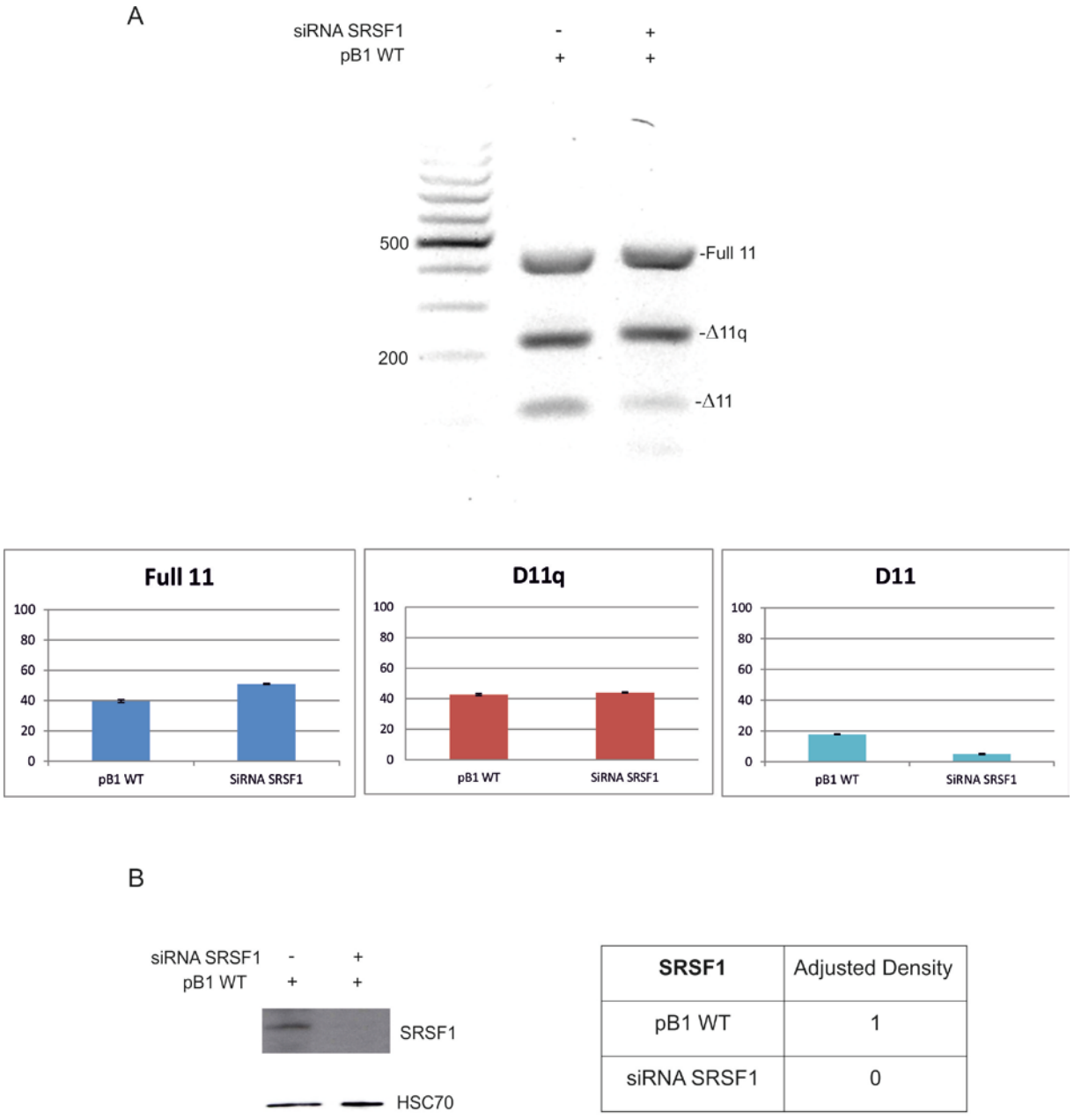
The sequence variant c.693G>A has been shown to cause skipping of exon 11 and decreased production of the Δ11q isoform. (Chapter 3). In addition *in vitro* analysis showed that binding sites for SRSF1, SRSF6 and SRSF9 are disrupted in the presence of the variant. Therefore it is possible that all these proteins contribute to the inclusion of exon 11. This hypothesis predicts that depletion of these proteins should cause skipping of exon 11 and/or reduction of exon 11q isoform in the pB1WT mini-gene but not in the pB1 mini-gene with the variant c.693G>A because the proteins SRSF1, SRSF5 and SRSF9 have been previously shown not to bind the RNA oligo containing this variant (Results 4.2).

In this section, experiments of silencing and overexpression of SR proteins in MCF7 cell lines were used to confirm a functional role of these proteins in the regulation of BRCA1 exon 11. Silencing (siRNA) was performed as described in Methods 2.7.3.

I observed that following knockdown of SRSF1 the pB1 WT mini-gene shows an increase of 10% of the Full 11 isoform and a decrease of 10% of the Δ11 isoform compared with pB1 transfected alone (without knockdown of SRSF1) as is shown in Figure 4.12.

Western blot analysis against HSC70 was performed in order to control that the protein levels loaded on the gel were similar as measured by nanodrop (Figure 4.12B).

The knocking down of SRSF1 was also done in association with the pB1 c.693G>A mini-gene. As expected there was no particular difference between the splicing outcome of pB1 c.693G>A alone and pB1 c.693G>A following knock-down of SRSF1 (Figure 4.13).

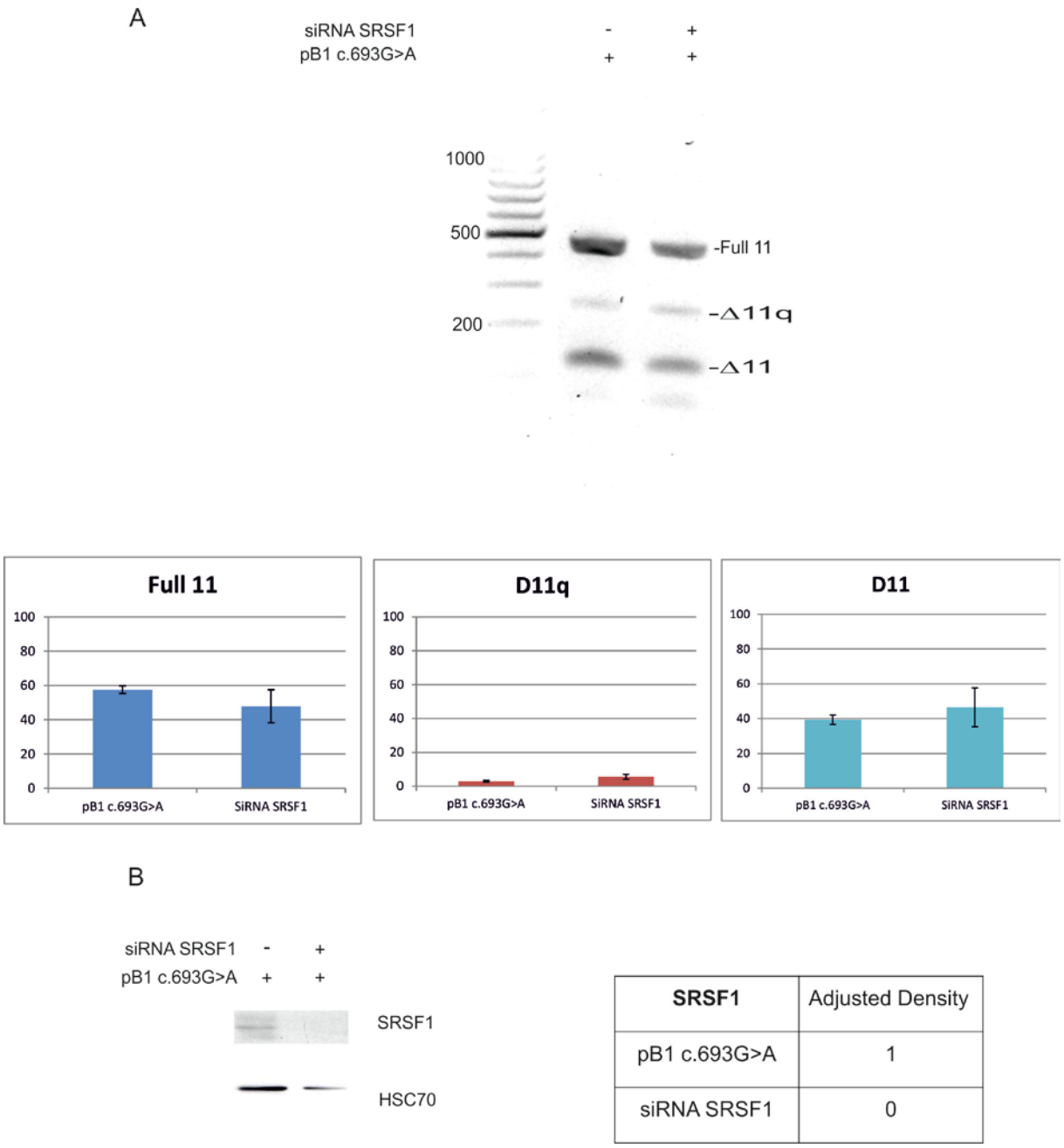


SRSF1	Adjusted Density
pB1 WT	1
siRNA SRSF1	0

Figure 4.12: Silencing of SRSF1 and transfection of the pB1 WT mini-gene in breast cancer cell lines (MCF7).

(**A**) pB1 WT mRNA processing in SRSF1 siRNA treated cells. The bottom panel shows a histogram where the percentages (%) of Full 11, Δ11q and Δ11 isoform respectively are calculated against the total expression of the three isoforms. The percentages and standard deviation (error bars) are calculated from two biological replicates. The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is reported the isoform

corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.
(B) Western blot using antibody antiSRSF1 performed following protein extraction from siRNA treated and untreated MCF7 cells shows evidence that the knock-down of SRSF1 satisfactorily worked. The Relative Densities for the sample band (+ siRNA SRSF1) was normalized relative to the standard in lane scramble (- siRNA SRSF1). The "Adjusted Density" for each sample lane was calculated by dividing the sample relative density by the loading-control (HSC70) relative density (Methods 2.5.6).



SRSF1	Adjusted Density
pB1 c.693G>A	1
siRNA SRSF1	0

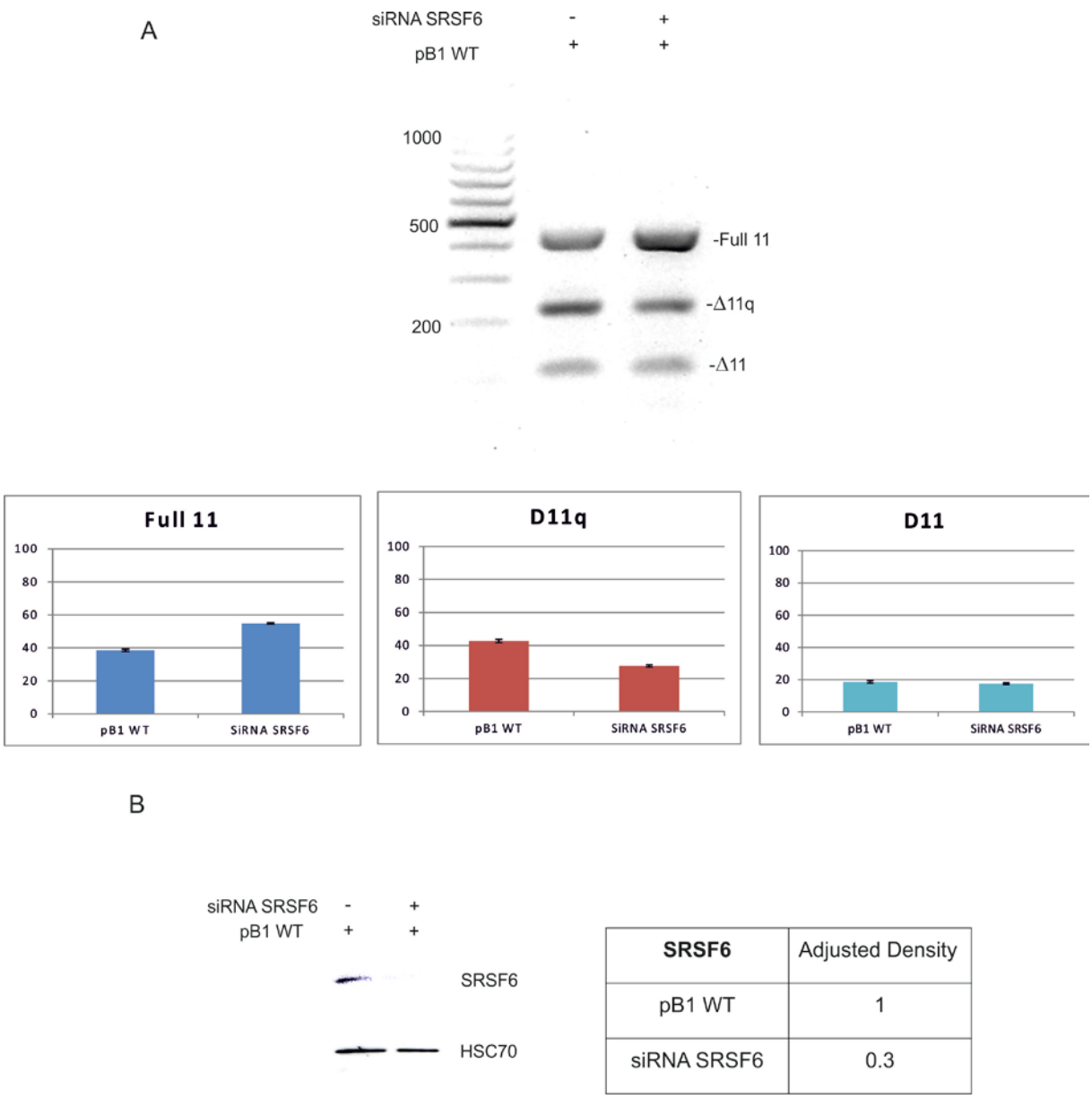
Figure 4.13: Silencing of SRSF1 and transfection of the pB1 c.693G>A mini-gene in breast cancer cell lines (MCF7).

(**A**) pB1 c.693G>A mRNA processing in SRSF1 siRNA treated cells. The bottom panel shows a histogram where the percentages (%) of Full 11, Δ11q and Δ11 isoform respectively are calculated against the total expression of the three isoforms. The percentages and standard deviation (error bars) are calculated from two biological replicates. The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is

reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.
(B) Western blot using antibody antiSRSF1 performed following protein extraction from siRNA treated and untreated MCF7 cells shows evidence that the knock-down of SRSF1 satisfactorily worked. The Relative Densities for the sample band (+ siRNA SRSF1) was normalized relative to the standard in lane scramble (- siRNA SRSF1). The "Adjusted Density" for each sample lane was calculated by dividing the sample relative density by the loading-control (HSC70) relative density (Methods 2.5.6).

Knock-down of the SR proteins SRSF6 and SRSF9 was also undertaken. Figure 4.14 shows knock-down of SRSF6 with transfection of pB1 WT which showed an increase of the Full 11 isoform of 10% and a decrease of the Δ11q isoform. Knock-down of SRSF6 with the pB1 c.693G>A did not cause any change of the splicing pattern compared with the pB1 c.693G>A alone (Figure 4.15).

Figure 4.16 shows knock-down of SRSF9 with transfection of pB1 WT which showed a small increase of the Full 11 isoform and a decrease of the Δ11q and Δ11 isoforms. The knock-down experiment using SRSF9 siRNA did not show a difference in the splicing pattern of the pB1 c.693G>A (Figure 4.17) mini-gene. Following knock down of SRSF6 or SRSF9, the reduction of Δ11q isoform with the pB1 c.693G>A mini-gene (Figure 4.15, Figure 4.17) is like an artefact of the reduction in the amount of mini-gene RNA. This reduction may be caused by a secondary effect of the knock down in the transfection efficiency.

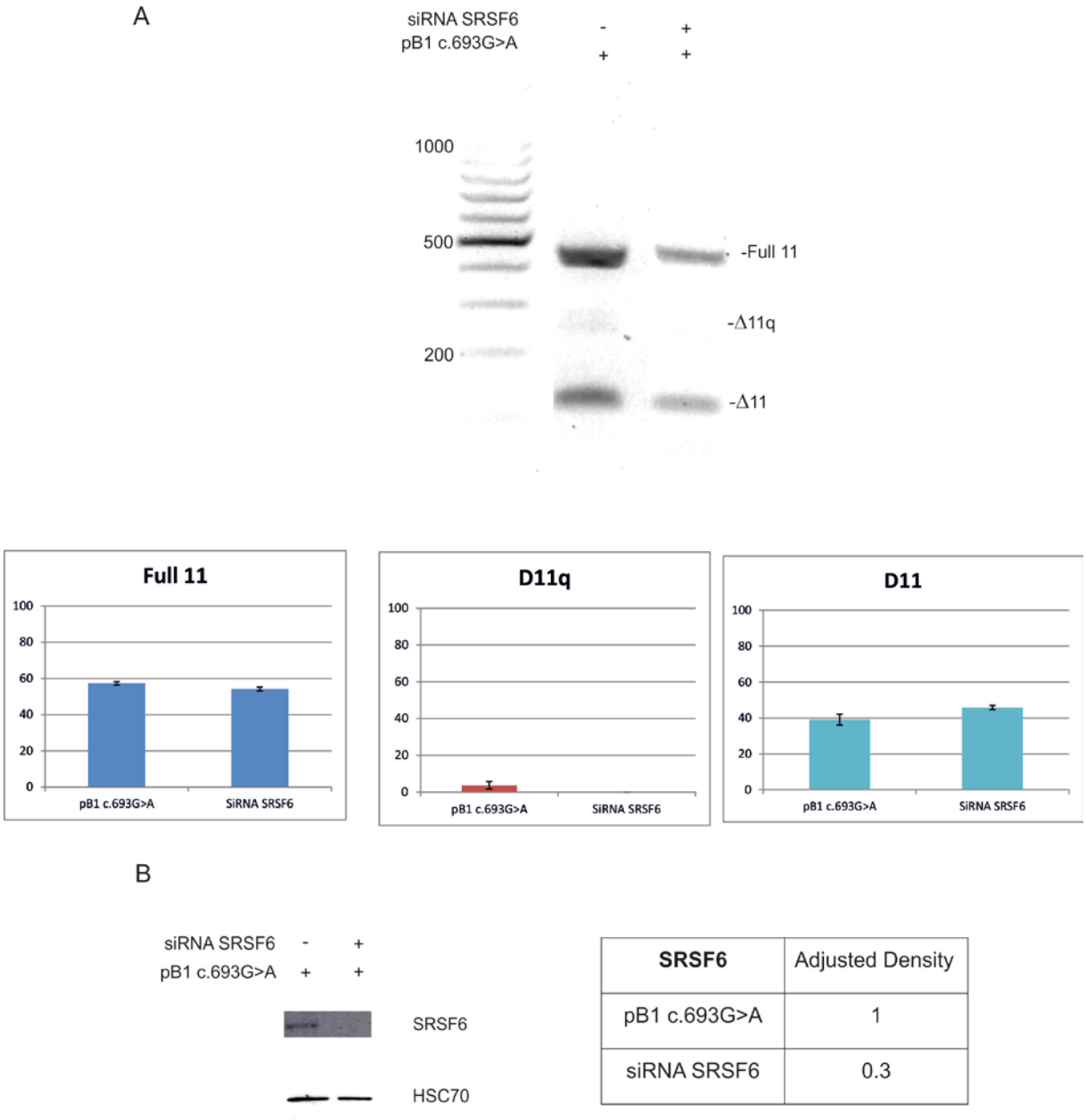


SRSF6	Adjusted Density
pB1 WT	1
siRNA SRSF6	0.3

Figure 4.14: Analysis of the silencing of SRSF6 co-transfection of the pB1 WT mini-gene in MCF7 cell line.

(**A**) pB1 WT mRNA processing in siRNA of SRSF6 treated cells. The bottom panel shows a histogram where the percentages (%) of Full 11, Δ11q and Δ11 isoform respectively are calculated against the total expression of the three isoforms. The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.

(B) Western blot using antibody antiSRSF6 performed following protein extraction from siRNA treated and untreated MCF7 cells showed evidence that the knock-down of SRSF6 satisfactorily worked. The Relative Densities for the sample band (+ siRNA SRSF6) was normalized relative to the standard in lane scramble (- siRNA SRSF6). The “Adjusted Density” for each sample lane was calculated by dividing the sample relative density by the loading-control (HSC70) relative density (Methods 2.5.6).

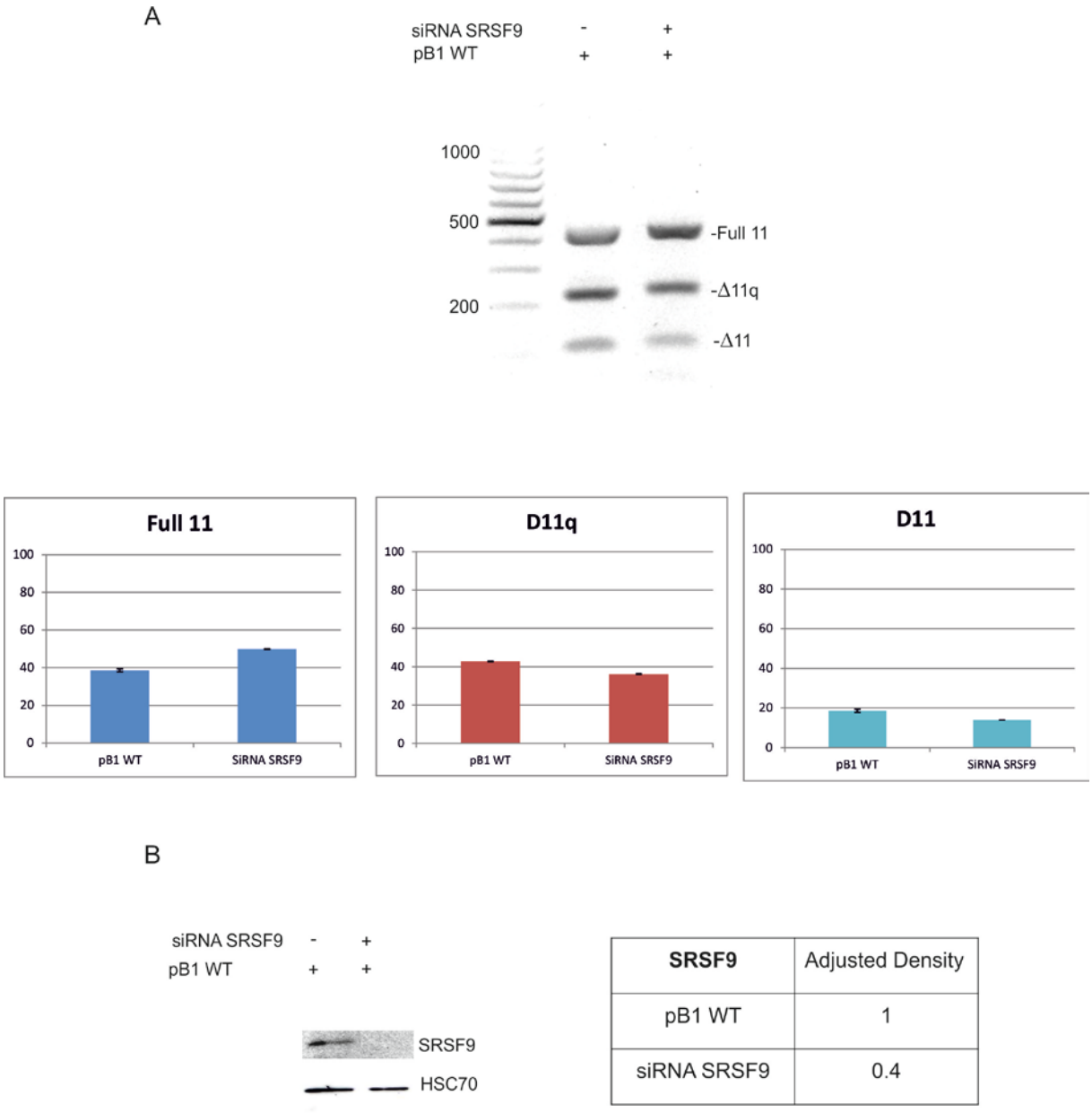


SRSF6	Adjusted Density
pB1 c.693G>A	1
siRNA SRSF6	0.3

Figure 4.15: Silencing of SRSF6 co-transfected with pB1 c.693G>A mini-gene in breast cancer cell line.

(**A**) pB1 c.693G>A mRNA processing in cells treated with siRNA against SRSF6 (SRp55). The bottom panel shows a histogram where the percentages (%) of Full 11, Δ11q and Δ11 isoform respectively are calculated against the total expression of the three isoforms. The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.

(B) Western blot using antibody antiSRSF6 performed following protein extraction from siRNA treated and untreated MCF7 cells showed evidence that knock-down of SRSF6 satisfactorily worked. The Relative Densities for the sample band (+siRNA SRSF6) was normalized relative to the standard in lane scramble (- siRNA SRSF6). The “Adjusted Density” for each sample lane was calculated by dividing the sample relative density by the loading-control (HSC70) relative density (Methods 2.5.6).

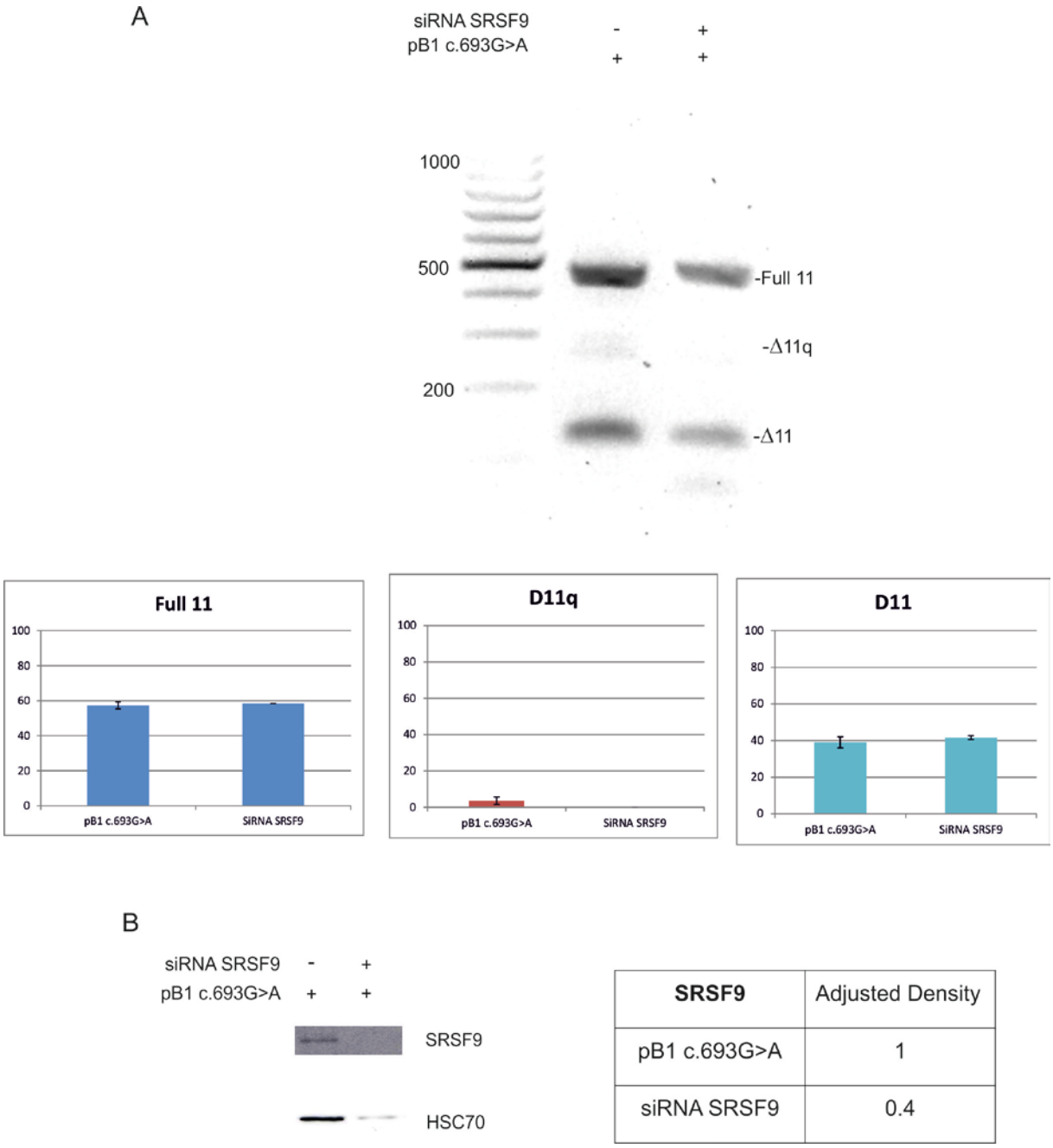


SRSF9	Adjusted Density
pB1 WT	1
siRNA SRSF9	0.4

Figure 4.16: Silencing of SRSF9 and transfection of the pB1 WT mini-gene.

(**A**) pB1 WT mRNA processing following siRNA of SRSF9 treated cells. The bottom panel shows a histogram where the percentages (%) of Full 11, Δ11q and Δ11 isoform respectively are calculated against the total expression of the three isoforms. The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.
(B) Western blot using antibody antiSRSF9 performed following protein extraction from siRNA treated and untreated MCF7 cells showed evidence that

knock-down of SRSF9 satisfactorily worked. The protein amount loaded was checked using an anti-HSC70 antibody. The Relative Densities for the sample band (+siRNA SRSF9) was normalized relative to the standard in lane scramble (-siRNA SRSF9). The “Adjusted Density” for each sample lane was calculated by dividing the sample relative density by the loading-control (HSC70) relative density (Methods 2.5.6).



SRSF9	Adjusted Density
pB1 c.693G>A	1
siRNA SRSF9	0.4

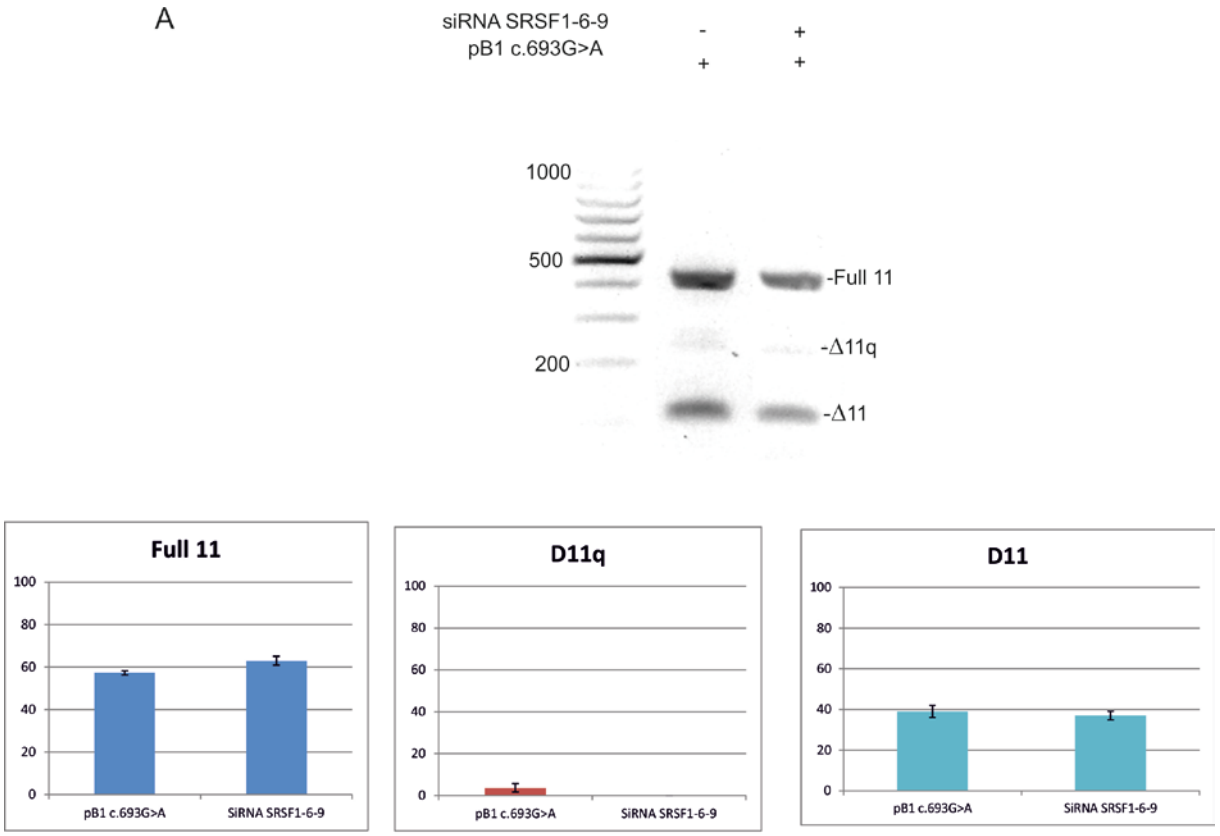
Figure 4.17: Silencing of SRSF9 and co-transfection of the pB1 c.693G>A mini-gene in breast cancer cell line.

(**A**) pB1 c.693G>A mRNA processing following siRNA of SRSF9. The bottom panel shows a histogram where the percentages (%) of Full 11, Δ11q and Δ11 isoform respectively are calculated against the total expression of the three isoforms. The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is reported the isoform

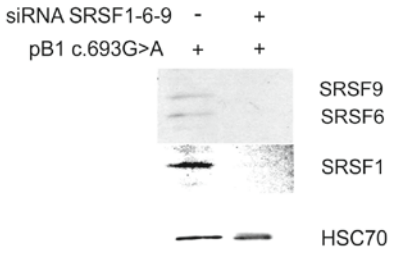
corresponding to each of the three bands: Full length 11 (Full 11), the Δ11q and the Δ11.
(B) Western blot using antibody antiSRSF9 performed following protein extraction from siRNA treated and untreated MCF7 cells shows evidence that knock-down of SRSF9 satisfactorily worked. The Relative Densities for the sample band (+siRNA SRSF9) was normalized relative to the standard in lane scramble (-siRNA SRSF9). The "Adjusted Density" for each sample lane was calculated by dividing the sample relative density by the loading-control (HSC70) relative density (Methods 2.5.6).

If SRSF1 binds to the enhancer regulatory sequence and is depleted following a knock-down experiment, it is possible that an alternative SR protein (e.g. SRSF6 and SRSF9) that also binds the regulatory sequence may compensate. Considering the possibility that all three SR proteins (SRSF1, SRSF6 and SRF9) could bind the regulatory sequence identified in chapter 3, I performed a siRNA against all three proteins at the same time.
Simultaneous knock-down of the three SR proteins did not show any effect on splicing pattern (Figure 4.18) of the pB1 c.693G>A mini-gene. However triple knock-down affected the pB1 WT mini-gene causing an increase of the full 11 isoform and a decrease of the Δ11q and Δ11 isoforms (Figure 4.19)

Silencing of SR proteins (proteins that usually behave as enhancers of splicing) unexpectedly caused a reduced recognition of exon 11 (increase of Δ11 and decrease of Δ11q isoforms).



B

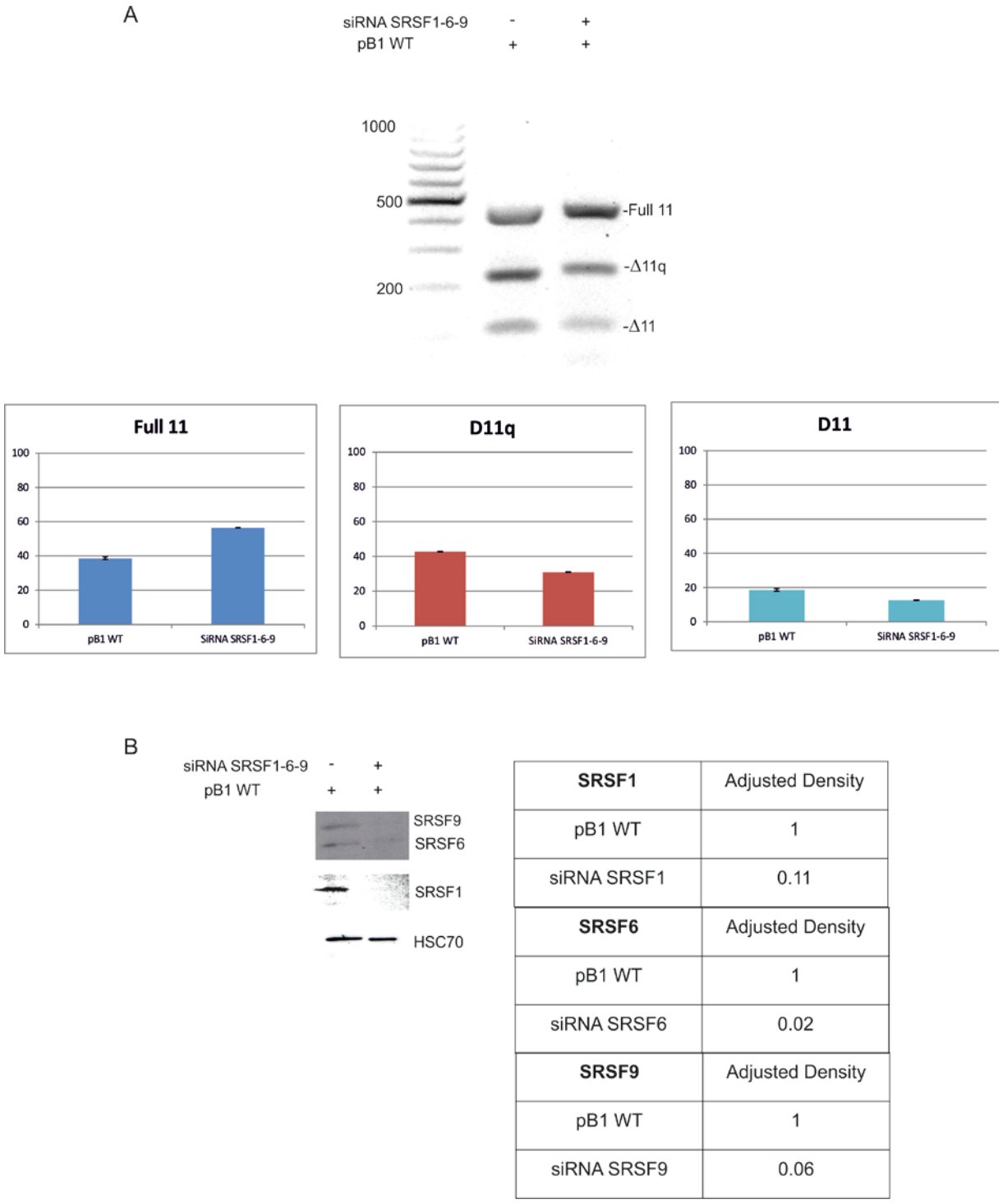


SRSF1	Adjusted Density
pB1 c.693G>A	1
siRNA SRSF1	0.11
SRSF6	Adjusted Density
pB1 c.693G>A	1
siRNA SRSF6	0.02
SRSF9	Adjusted Density
pB1 c.693G>A	1
siRNA SRSF9	0.06

Figure 4.18: Silencing of SRSF1, SRSF6 and SRSF9 and transfection of the pB1 c.693G>A mini-gene.

(**A**) pB1 c.693G>A mRNA processing following siRNA of SRSF1 SRSF9, SRSF6. The bottom panel shows a histogram where the percentages (%) of Full

11, Δ11q and Δ11 isoform respectively are calculated against the total expression of the three isoforms. The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.
(B) Western blot using antibody anti1H4 performed following protein extraction from siRNA treated and untreated MCF7 cells shows evidence that the knock-down of SRSF1, SRSF6 and SRSF9 satisfactorily worked. The protein amount loaded was checked using an anti-HSC70 antibody. The Relative Densities for the sample bands (+siRNA SRSF1-6-9) were normalized relative to the standard in lane scramble (-siRNA SRSF1-6-9). The "Adjusted Density" was calculated by dividing the sample relative density by the loading-control (HSC70) relative density (Methods 2.5.6).



SRSF1	Adjusted Density
pB1 WT	1
siRNA SRSF1	0.11
SRSF6	Adjusted Density
pB1 WT	1
siRNA SRSF6	0.02
SRSF9	Adjusted Density
pB1 WT	1
siRNA SRSF9	0.06

Figure 4.19: Silencing of SRSF1, SRSF6 and SRSF9 and transfection of the pB1 WT mini-gene in breast cancer cell line.

(A) pB1 WT mRNA processing in SRSF1, SRSF6 and SRSF9 siRNA treated cells. The bottom panel shows a histogram where the percentages (%) of Full 11, Δ11q and Δ11 isoform respectively are calculated against the total expression of

the three isoforms. The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.
(B) Western blot using antibody anti 1H4 performed following protein extraction from siRNA treated and untreated MCF7 cells. The Relative Densities for the sample band (+ siRNA SRSF1-6-9) were normalized relative to the standard in lane scramble (- siRNA SRSF1-6-9). The "Adjusted Density" was calculated by dividing the sample relative density by the loading-control (HSC70) relative density (Methods 2.5.6).

If the hypothesis that the sequence variant destroyed the binding of enhancer proteins SRSF1 SRSF6 and SRSF9 is correct then it predicts that a cellular increase of these proteins should lead to an increase of BRCA1 exon 11 splicing. To investigate this I performed co-transfection experiments with the pB1 WT mini-gene and expression vectors encoding SRSF1 or SRSF6.

The overexpression of SRSF6 co-transfected with the pB1 WT mini-gene showed a decrease of the Δ11 isoform and increase of the Full 11 isoform compared with the pB1 WT mini-gene alone (Figure 4.20). In contrast overexpressing SRSF6 with pB1 c.693G>A showed an increase of the Δ11q isoform and a decrease of the Δ11 isoform, shown in Figure 4.20.
Overexpression of SRSF1 showed a relative increase of full 11 isoform in pB1 WT mini-gene as expected. The Δ11q and the Δ11 isoform decreased (Figure 4.21).
If the pB1 c.693G>A mini-gene alters the splicing pattern because of complete disruption of SRSF1 binding it is expected that overexpression of SRSF1 would not have any effect on the splicing pattern in the mutant mini-gene pB1 c.693G>A. However, overexpression of SRSF1 co-transfected with the pB1 c.693G>A mini-gene showed a decrease of the Δ11 isoform and increase of the full relative to the total (Figure 4.21).

It is possible that the synonymous mutation c.693G>A does not completely destroy the binding site for SRSF1 or SRSF6 or/and these proteins may also binding another region in the pB1 mini-gene (probably in BRCA1 exon 11 or flanking introns).

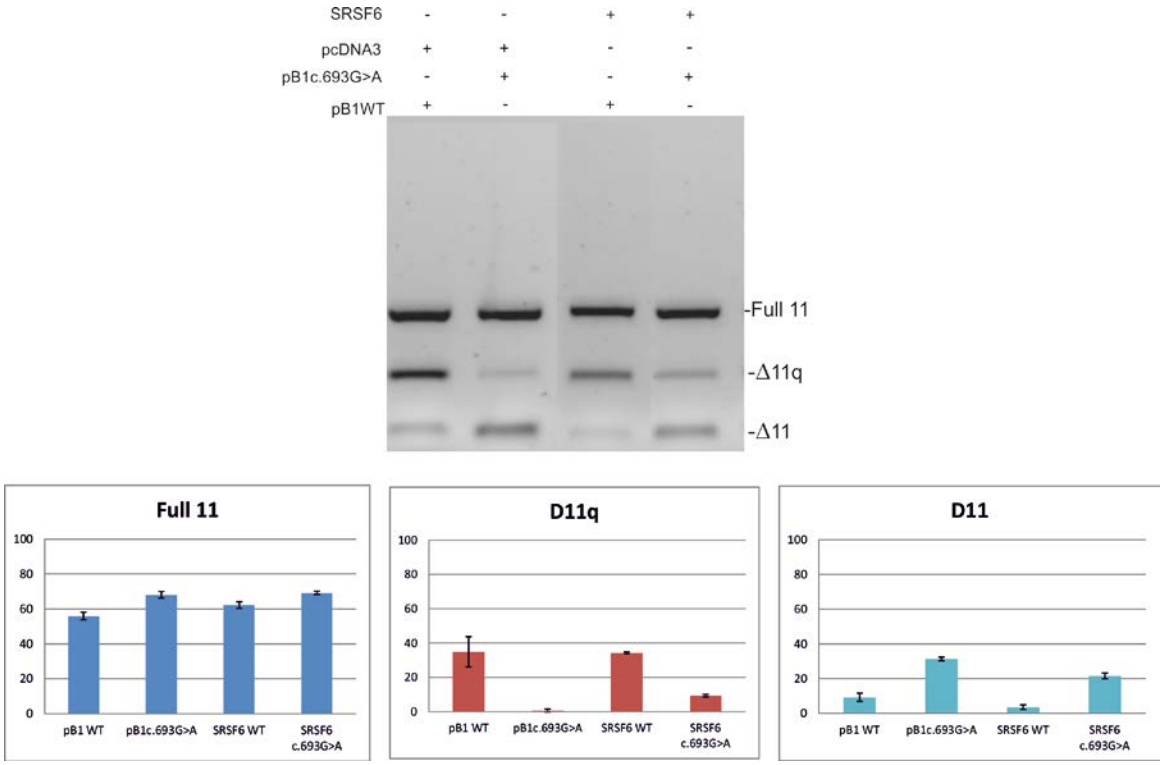


Figure 4.20: Overexpression of SRSF6 co-transfected with the pB1 WT or pB1 c.693G>A mini-gene.

RT-PCR of pB1 WT or pB1c.693G>A co-transfected with a plasmid expressing SRSF6 or the empty vector pCDNA3. The PCR products were separated on 1.5 % agarose gel. The histograms show the percentage (%) of Full 11, Δ11q and Δ11 isoform respectively, calculated against the total expression of the three isoforms. The intensity of each band was calculated using image-J (Methods 2.5.6). On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.

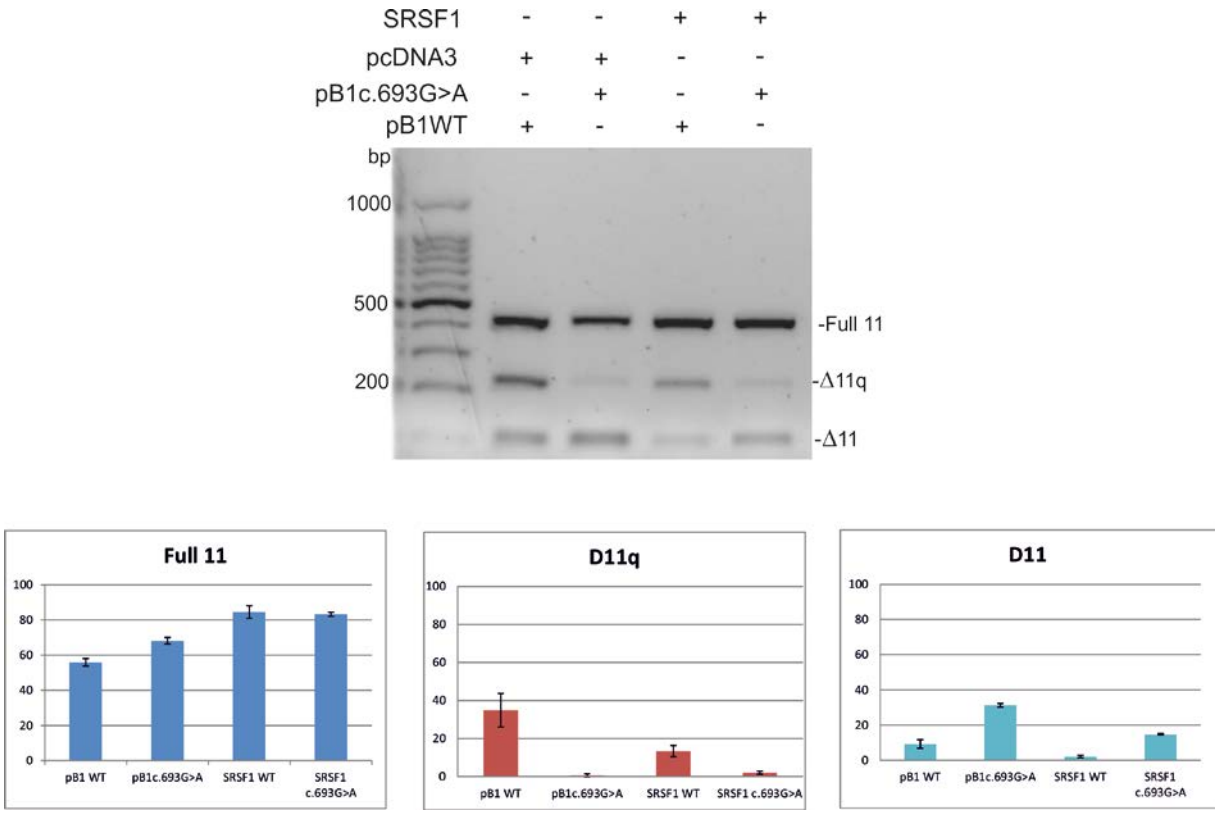


Figure 4.21: Overexpression of SRSF1 co-transfected with the pB1 WT or pB1 c.693G>A mini-gene.

RT-PCR of pB1 WT or pB1c.693G>A co-transfected with a plasmid expressing SRSF1 or the empty vector pCDNA3. The PCR product was separated on 1.5 % agarose gel. The histograms show the percentage (%) of Full 11, Δ11q and Δ11 isoform respectively, calculated against the total expression of the three isoforms. The intensity of each band was calculated using image-J (Methods 2.5.6). The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.

4.4 Conclusion.

My results presented here demonstrate *in vitro* binding sites for different SR proteins in the wild type RNA oligo which are disrupted in the presence of the variant c.693G>A. In particular pull down analysis and western blot analysis demonstrated that SRSF1 (Figure 4.5), SRSF6 (Figure 4.6), and SRSF9 (Figure 4.6), are able to bind wild type RNA sequence but not the mutant sequence. These results partially confirmed the bioinformatic analysis which predicted SRSF1, but not SRSF6, and SRSF9, binding (Table 3.1and Table 3.2).
These results demonstrated that the bioinformatic analysis is not always precise when predicting a binding site; consequently it is important to do *in vitro* analysis.
Pull down analysis showed the presence of other splicing factors: TDP43 and hnRNPA1 (Figure 4.2). Mass spectrometry analysis demonstrated that TDP43 and hnRNPA1 are binding both the wild type RNA oligo and synonymous oligo, and these results were confirmed using specific antibodies for hRNPA1 and TDP43. Taken together these results suggests that TDP43 and hRNPA1 are not the splicing factors that are able to determine the exclusion of exon 11 in presence of the synonymous variation c. 693G>A.
An additional siRNA analysis has been done to demonstrate that SRSF1, SRSF6 and SRSF9 are able to bind the wild type sequence. The knock-down of these SR proteins did not confirm the *in vitro* analysis. Usually if these SR protein are able to bind the wild type sequence, knock-down experiment should increase skipping of exon 11 and decrease the Δ11q isoform as demonstrated for the variant c.693G>A. The fact that there are unexpected results with the knock-down analysis, which are not consistent with the *in vitro* result, could be because several splicing factors can bind the region surrounding c.693 masking the effect of SRSF1, SRSF6 and SRSF9 down regulation. Another explanation may be the large size of exon 11 which could bind those SR proteins in more than one region and binding of the SR proteins may have different silencer or enhancer effects depending on position. Finally it is also possible that knock down of these

proteins could cause a secondary effect on the expression of other splicing regulatory proteins and these in turn are affecting the splicing pattern of exon 11.

In support of this last hypothesis is the fact that over expression of SR proteins is unexpectedly having an effect on mini-gene splicing which is similar of that obtained following depletion of SR proteins (relative increase of full isoform and decrease of Δ11 isoform for both WT and c.693 G>A mini-genes (Figure 4.20 and Figure 4.21).

The overexpression results confirm that SRSF6 and SRSF9 are enhancer of exon 11 splicing. In addition this experiment suggests that besides the splicing regulatory region identified in this thesis they are also binding additional not yet identified regulatory regions of BRCA1 exon 11 splicing. In fact their over-expression was able to affect exon 11 splicing also in the presence of the c.693G>A variant.

Chapter 5

Results

Correction of exon 11 skipping in *BRCA1*.

5 Introduction.

The synonymous sequence variant c.693G>A has been described in the previous chapters to cause skipping of exon 11 in *BRCA1* gene, and to affect binding sites for splicing regulatory proteins. In particular, sites for splicing factors SRSF1, SRSF6 and SRSF9 were shown *in vitro* to be disrupted in the presence of the sequence change. In this chapter reversal of the altered splicing is attempted using bifunctional oligonucleotides.

Different oligonucleotides for splicing correction exist (Introduction 1.4.2) that are modified with different chemistries for instance 2'-*O*-MePs (Skordis et al., 2003a) 2'-*O*-Me (Williams et al., 2009) 2'-*O*-MOE (Hua et al., 2010) and PNA (Cartegni and Krainer, 2003).

Bifunctional oligonucleotides have a tail that recruits SR proteins and can therefore stabilize exon inclusion (Introduction 1.4.2).

It has been previously demonstrated by Skordis et al. that it is possible to modify *SMN2* splicing using bifunctional oligonucleotides and that this strategy could be applicable to any mutation affecting exonic splicing enhancers (ESE). The bifunctional oligonucleotides used by Skordis are complementary to the target exon and present non complementary regions named "tail", which consist of a sequence that mimics an ESE region, and thus recruit splicing factors. One of this oligonucleotide with tail is able to induce the inclusion of *SMN2* exon 7 (Skordis et al., 2003a).

Bifunctional oligonucleotides with a tail for SRSF1 were designed in this thesis using a 2'O-MePS modification in the annealing domain. This modification has been shown to give more stability and more specificity in duplex formation (Owen et al., 2011b).

In these thesis three different bifunctional oligonucleotides have been designed: Syn Oligo (Figure 5.1A), Del1 oligo (Figure 5.1B) and oligo Del7a (Figure 5.1C).

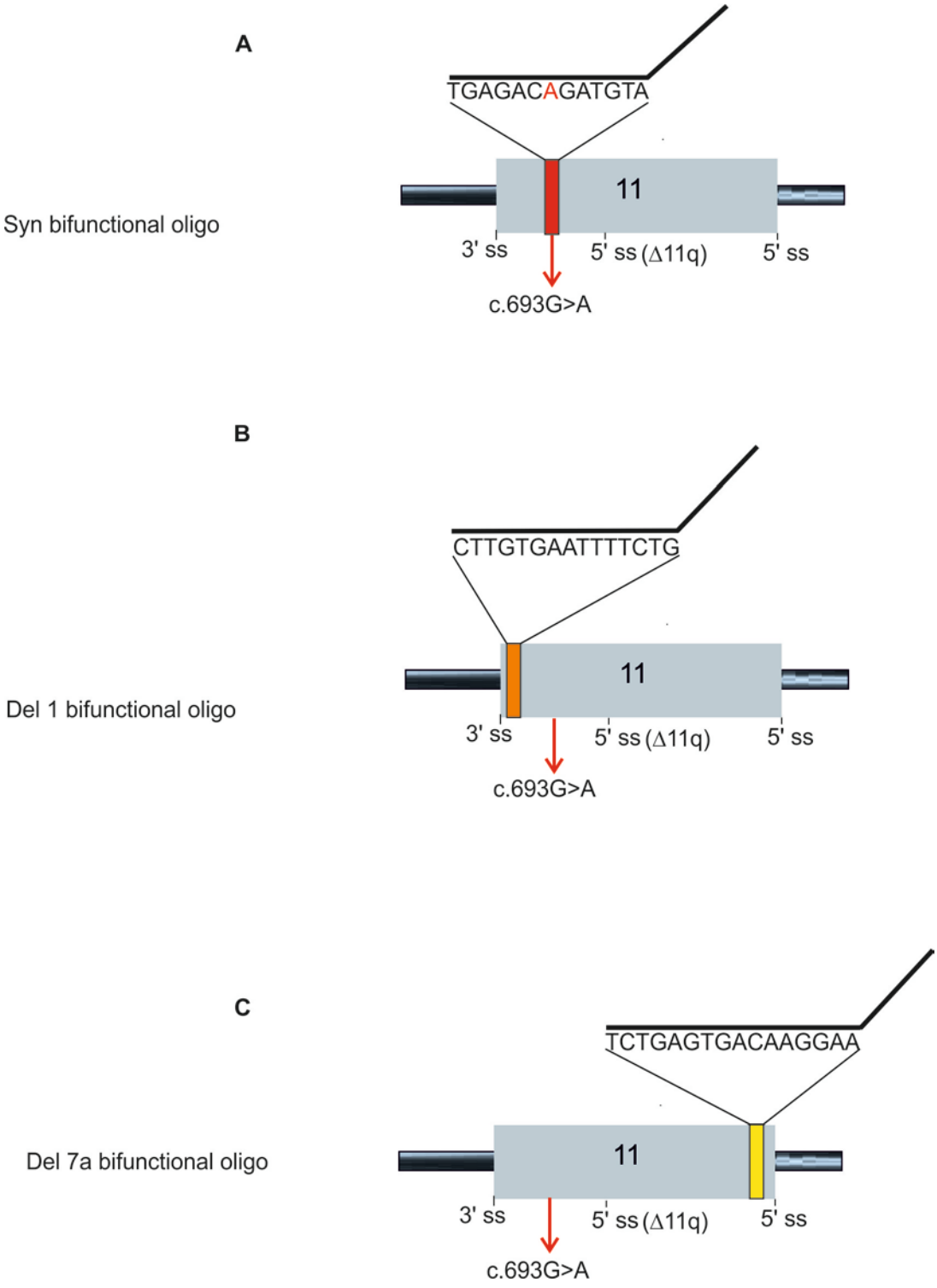


Figure 5.1: Schematic representation of bifunctional oligonucleotides.

The figure shows the three bifunctional oligonucleotides used. All the three bifunctional oligonucleotides present a tail for SRSF1 (oblique line). Schematic

representations of exon 11 (grey box) with flanking introns (black thick lines) are reported. The three splice sites of exon 11 are indicated (3'ss, 5'ss (Δ11q), 5'ss). The red arrow shows the position of the c.693G>A variant.
(**A**) Syn bifunctional oligo designed to bind the sequence with the synonymous variant c.693G>A (red box).
(**B**) Del 1 bifunctional oligo designed to bind a neutral region close to the 3' splice site (orange box).
(**C**) Del 7a bifunctional oligo designed to bind a neutral region (yellow box) situated close to the 5' splice site at the end of exon 11.

5.1 Analysis of the Effect of Bifunctional Oligonucleotides on Splicing using the mini-gene constructs.

The first bifunctional oligonucleotide used in this thesis was the Syn oligo. This bifunctional oligonucleotide was designed to be complementary to a region of BRCA1 exon 11 with the nucleotide change in position c.693G>A and presents an additional non complementary region (AGGAGGACGGAGGACGGAGGACA) that represents the tail for the recruitment of the splicing regulatory protein SRSF1. The purpose of this bifunctional oligonucleotide was to specifically bind the exon sequence of interest, recruit SRSF1 through the enhancer sequence of the tail and increase the possibility of the exon to be recognised by spliceosome factors.

In order to verify that the effect is due to the presence of the tail two different controls have been used (Method 2.6.2, Table 2.22); a Scrambled oligo (which is a bifunctional oligonucleotide with a tail for SRSF1 but not able to bind exon 11 nor match any sequence in the human genome) and a 'no tail' oligonucleotide (which has the same complementary region of the bifunctional oligonucleotide without the specific tail for SRSF1).
The Syn oligo co-transfected with the pB1 WT mini-gene was used as an additional control to show that the oligonucleotide had no particular effect in the absence of the c.693 variant (Appendix C, Figure 9.1 A).

The Figure 5.2 and Figure 5.3 show co-transfection of the Syn oligo (250uM and 400uM respectively) and the pB1 c.693G>A mini-gene. The Syn oligo decreases the use of the exon 11 3' splice site. The effect is neither relative decrease of both full nor Δ11q isoforms.

In conclusion the syn oligo bifunctional oligonucleotide, with the SRSF1 recruiting tail, is not able to stimulate inclusion of exon 11 with the synonymous variant c.693G>A.

A possible explanation for this could be that the oligonucleotide is not close enough to the 3' to enable the SR protein (SRSF1) binding the tail to recognise the splice site and activate the splicing machinery. The increase in exon 11 skipping could be explained by the covering of additional splicing enhancer sequences with the antisense complementary region of the syn oligo. However this hypothesis is not confirmed because the antisense oligonucleotide without the tail (syn NT) has no effect on exon 11 splicing (Figure 5.2 and Figure 5.3). Therefore it is possible that SRSF1 binding to the bifunctional oligo designed in this thesis is acting as a negative regulator of exon 11 inclusion.

.

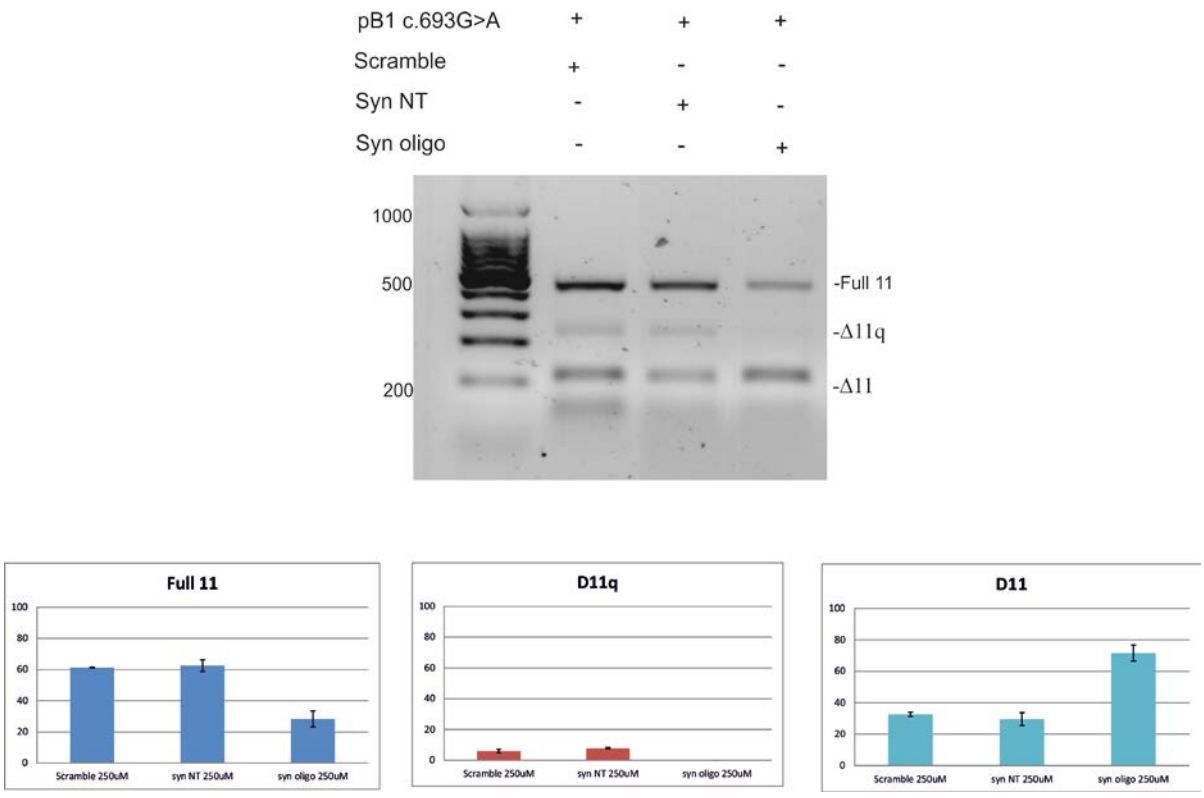


Figure 5.2: Effect of 250 uM syn oligo bifunctional oligonucleotide on c.693G>A BRCA1 exon 11.

The figure shows RT-PCR products from total RNA extracted after co-transfection of Syn oligo and pB1 c.693G>A mini-gene. Syn NT is the bifunctional oligonucleotide without the tail and the scramble oligo represents the control. The PCR product is separated on a 1.5% agarose gel. 250uM of Syn oligo has been used. The marker is indicated on the left of the gel and the corresponding size is reported next to each band. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11. The histogram shows the percentage of the Full isoform the D11q isoform and the D11 isoform respectively, and has been calculated against the total expression of the three isoforms (Methods 2.5.6).

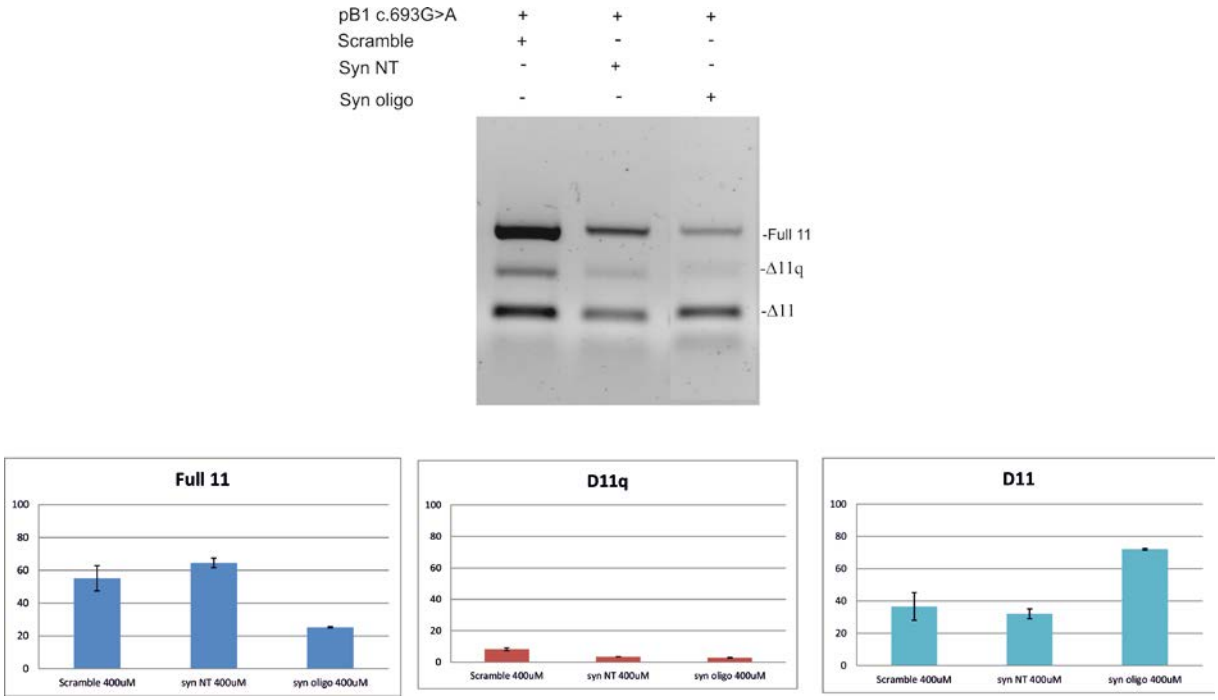


Figure 5.3: Effect of 400 uM syn oligo bifunctional oligonucleotide on c.693G>A BRCA1 exon 11.

The figure shows RT-PCR products from total RNA extracted after co-transfection of Syn oligo and pB1 c.693G>A mini-gene. Syn NT is the bifunctional oligonucleotide without the tail and the scramble oligo represents the control. 400uM of Syn oligo has been used. The PCR product is separated on a 1.5% agarose gel. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11. The histograms show the percentage of the Full isoform the D11q isoform and the D11 isoform respectively, and it has been calculated against the total expression of the three isoforms (Methods 2.5.6).

5.2 Analysis of the effect of the bifunctional oligonucleotide Del 1.

A second bifunctional oligonucleotide was designed to bind a neutral region situated inside exon 11 of the *BRCA1* gene which has been demonstrated previously in our laboratory by Raponi et al. to not change the splicing process (Raponi et al., 2012). This neutral region is situated 3 nucleotides downstream the 3'splice site of exon 11 (Raponi et al., 2012).
The bifunctional oligonucleotide which binds the neutral region is designed to be closer to the 3' splice site of exon 11 and carries an additional non complementary tail (AGGAGGACGGAGGACGGAGGACA) for the recruitment of SRSF1. This bifunctional oligonucleotide was named Del1 (Figure 5.1B).

The bifunctional oligonucleotide Del1 was co-transfected with the pB1 c.693G>A mini-gene in MCF7 cell lines (Methods 2.7.3). Figure 5.4 shows that the Del1 oligonucleotide with the mini-gene carrying the mutation c.693G>A increased the Δ11q isoform and decreased the Full 11 isoform compared to the scramble oligo with the mini-gene c.693G>A (Figure 5.4). A control oligonucleotide without the tail (Del1 NT) was used to verify that the effect of the bifunctional oligonucleotide with the tail was due to the presence of the tail. This oligo did not change the splicing pattern of the mini-gene (Figure 5.4).
Co-transfection of the bifunctional oligonucleotide Del1 with the pB1 WT mini-gene did not show an effect on the splicing pattern (Appendices C, Figure 9.2 B). This suggests that Del 1 oligo is a good candidate for correction of splicing mutations in heterozygous patients as it should not affect splicing of the WT allele.
In conclusion the bifunctional oligonucleotide designed to target the neutral region at the beginning of exon 11 appears to have a positive effect on the recognition of the 5' splice site of the exon 11q isoform. In order to properly correct the splicing of c.693G>A it would be also necessary to increase the

recognition of the 5’splice site at the end of exon 11. In this way it would be possible to increase the production of the Full 11 isoform.

For this reason another bifunctional oligonucleotide has been designed to use in association with the bifunctional oligonucleotide Del1 in order to increase recognition of the 5’ splice site at the end of exon 11.

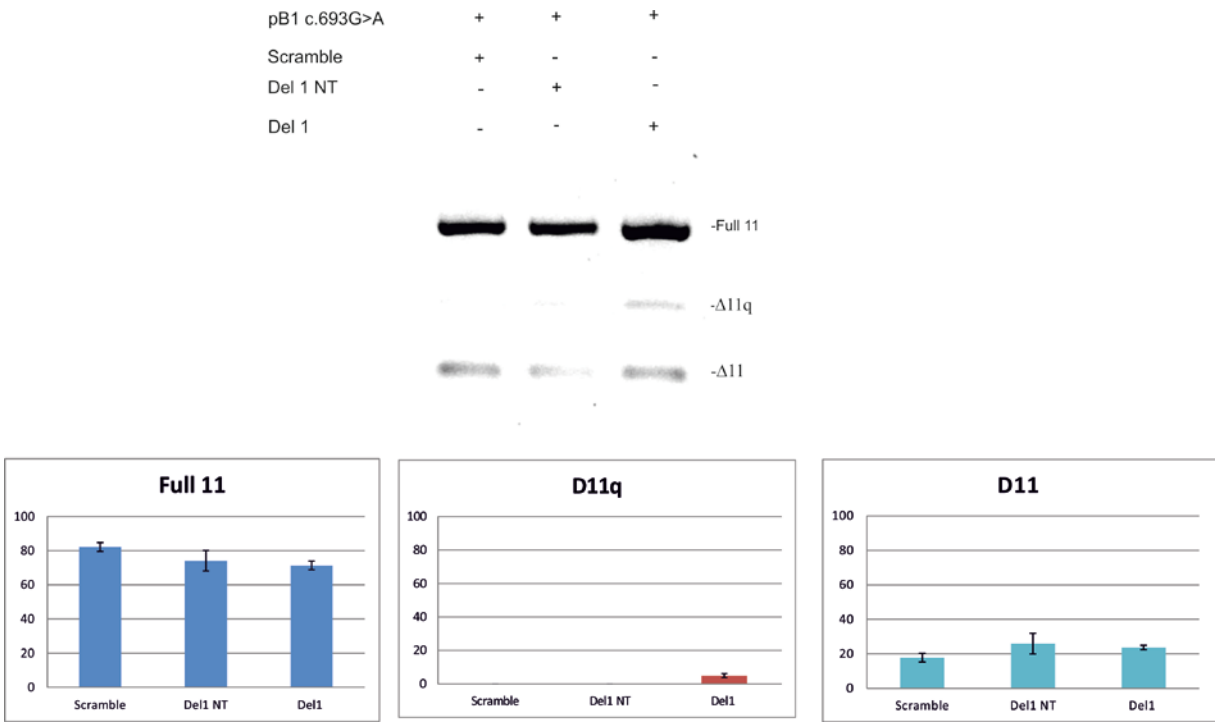


Figure 5.4: *In vivo* correction of c.693G>A BRCA1 exon 11 using 400 uM of Del 1 bifunctional oligonucleotide.

The figure shows RT-PCR products from total RNA extracted after co-transfection of Del 1 oligo and pB1 c.693G>A mini-gene. Del 1 NT is the bifunctional oligonucleotide without the tail and the scramble oligo represents the control. 400uM of each oligo has been used. The PCR product is separated on a 1.5% agarose gel. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11. The histograms show the percentage of the Full isoform the D11q isoform and the D11 isoform respectively, and it has been calculated against the total expression of the three isoforms (Methods 2.5.6).

5.3 Analysis of the effect of the bifunctional oligonucleotide Del 7.

Taking into consideration that the exon 11 of *BRCA1* gene is a large exon (3.4kb) it is probable that the use of one bifunctional oligonucleotide is not enough to allow inclusion of the entire exon.

For this reason a third bifunctional oligonucleotide, Del 7a, with a tail for SRSF1 was designed (Figure 5.1C) in order to improve the recognition of the 5'splice site at the end of exon 11. This oligonucleotide binds a neutral region situated at the end of exon 11. Following deletion of this region in the pB1 WT mini-gene the splicing pattern does not change (personal communication from Raponi).

Transfection of this oligo alone was expected to increase the Full 11 isoform. In addition the transfection of this oligo together with Del 1 oligo (shown in Figure 5.4 to increase the Δ11q isoform) was expected to recreate the WT splicing pattern in c.693G>A mini-gene.

Experiments using a single co-transfection of the bifunctional oligonucleotide Del7 with the pB1 c.693G>A mini-gene has been undertaken in MCF7 cell lines. Figure 5.5, shows that the bifunctional oligonucleotides Del 7 has a inhibitory effect on exon 11 5'splice site because it induces a relative increase of Δ11 isoform and decrease of Full 11 isoform. As an additional control co-transfection of the bifunctional oligonucleotide Del1 with the pB1 WT mini-gene did not show an effect on the splicing pattern (Appendix C, Figure 9.3C).

The bifunctional oligonucleotide Del7 has also been co-transfected with the bifunctional oligonucleotide Del 1 and the mini-gene c.693 G>A. In this case it was possible to see an increase of the splicing isoform Δ11q (Figure 5.5) comparable with that observed in Figure 5.4 with transfection of Del1 oligo alone. Therefore the Del 7 oligo is not a good candidate for correction of aberrant splicing of BRCA1 exon11.

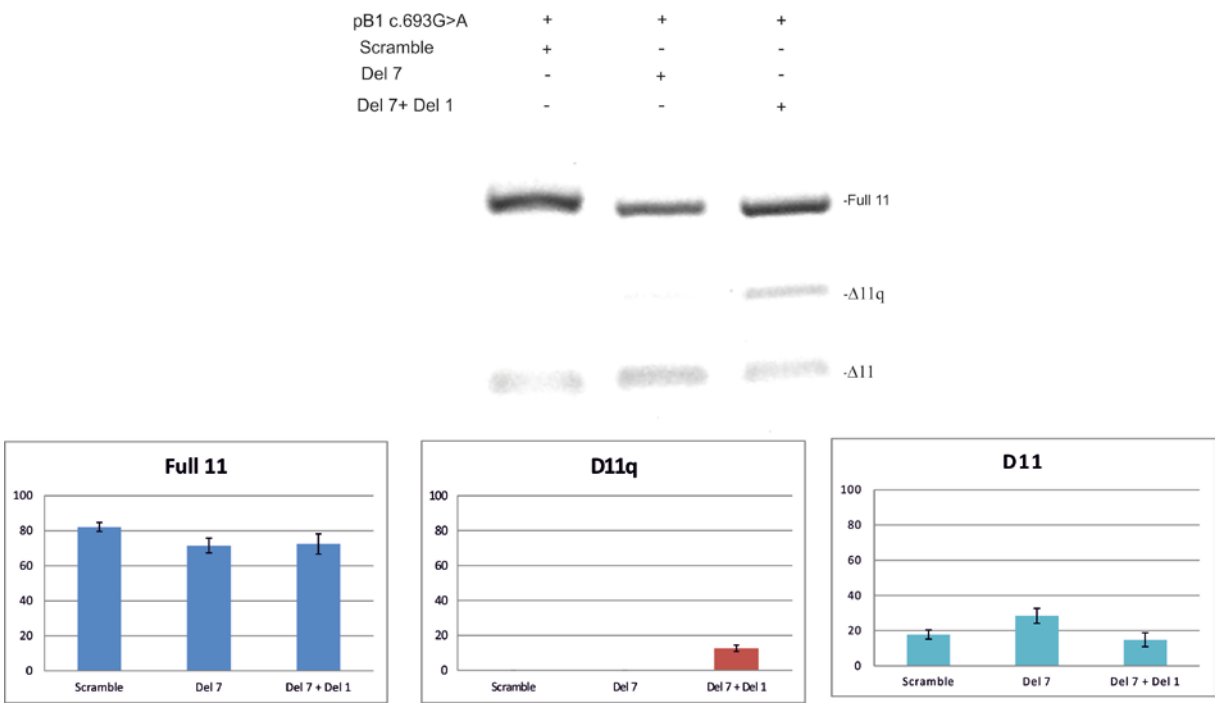


Figure 5.5: pB1 c.693G>A mini-gene co-transfected with Del 7 bifunctional oligonucleotide.

The figure shows RT-PCR products from total RNA extracted after co-transfection of Del7 oligo and pB1 c.693G>A mini-gene. The scramble oligo represents the control. The PCR product is separated on a 1.5% agarose gel. On the right of the gel is reported the isoform corresponding to each of the 3 bands: Full length 11 (Full 11), the Δ11q and the Δ11.The histograms show the percentage of the Full isoform the D11q isoform and the D11 isoform respectively, and have been calculated against the total expression of the three isoforms (Methods 2.5.6).

5.4 Conclusion.

The potential of correcting single point mutations using antisense oligonucleotide technology is a growing field offering a new therapeutic approach.

The c.693G>A mutation was shown to disrupt an enhancer sequence in BRCA1 exon 11, and therefore a bifunctional splicing that actives splicing could have been a solution to enhance inclusion of exon 11.

However, the Syn oligo bifunctional oligonucleotide designed to bind mutant exon 11 and recruit the enhancer protein SRSF1 was not successful (Figure 5.2 and Figure 5.3). This oligonucleotide was designed to match the synonymous variant c.693G>A. Additional bifunctional oligonucleotides (Del1 and Del7) able to bind neutral regions at the beginning or at the end of exon 11 have been employed to attempt correction of the splicing defect caused with the synonymous variant c.693 G>A. Because these oligos are designed to bind a neutral region in exon 11 if successful it would have been possible to employ them as therapeutic oligos for patients that have splicing mutations anywhere in exon 11 and not just patients with the c.693G>A nucleotide variant. However, co-transfection of the Del1 or Del7a oligo with pB1 c.693G>A mini-gene did not recreate the wild type splicing pattern (Figure 5.4 and Figure 5.5).

Targeting the Del 1 bifunctional oligonucleotide close to the 3' splice site would possibly have enhanced the inclusion of the exon. This bifunctional oligonucleotide improved the recognition of the 5' splice site of the Δ11q isoform but not the recognition of exon 11 terminal 5' splice site. This could have probably happened because of the distance of the exon 11 terminal 5' splice site and the region to which oligo Del 1 was designed to bind.

In conclusion each of the bifunctional oligonucleotides were unsuccessful in facilitating facilitate inclusion of the entire exon 11 in the presence of the variant; this suggests that the implementation of modified oligonucleotides to correct splicing of large exons is not easy and it may require a deeper knowledge of how splicing in these large exons is regulated.

.

Chapter 6

Discussion.

6 Discussion.

6.1 Identification of BRCA1 exon 11 splicing mutation.

The Breast Cancer Gene (*BRCA1)* is one of two major genes associated with an increased risk of breast cancer, although mutations in this gene are responsible for less than 5% of the risk of developing breast cancer, and for less than 20% of all familial cases.

In the *BRCA1* gene different splicing mutations have been identified, however at the time of writing this thesis no splicing mutations have been detected in BRCA1 exon 11.

An RNA sample was available from a breast cancer patient with a strong family history of breast cancer who was found to have a synonymous sequence variant in exon 11. I was able to analyse the RNA and identified for the first time that the presence of this unclassified variant in exon 11 (c.693G>A) affects splicing of exon 11 (Figure 3.2).

Before we were able to publish this Dias Brandao et al. reported that c.693G>A affected exon 11 (Brandao et al., 2011). However in this latter article the unclassified variant was not considered to be a pathogenic variant did not satisfy two necessary criteria: monoallelic expression of the full length transcript and that the unclassified variant creates an aberrant transcript containing a premature termination codon. For that reason Brandao et al classified the variant c.693G>A as unclear pathogenicity (Brandao et al., 2011).

However, it is still possible that this variant could give cancer predisposition. In fact, other studies already suggest that splicing isoforms around BRCA1 exon 11 are implicated in cancer. These isoforms are naturally expressed; and proportions change in breast cancer cell lines and during the cell cycle (Orban and Olah,

2003). This suggests maintaining correct isoform proportions are important in preventing cell transformation.

Exon 11 has three functional splice sites shown previously in Figure 1.15 (Introduction) The alternative use of these splice sites determines the production of the three BRCA1 isoforms Full 11, Δ11q, and Δ11 (Figure 6.1).

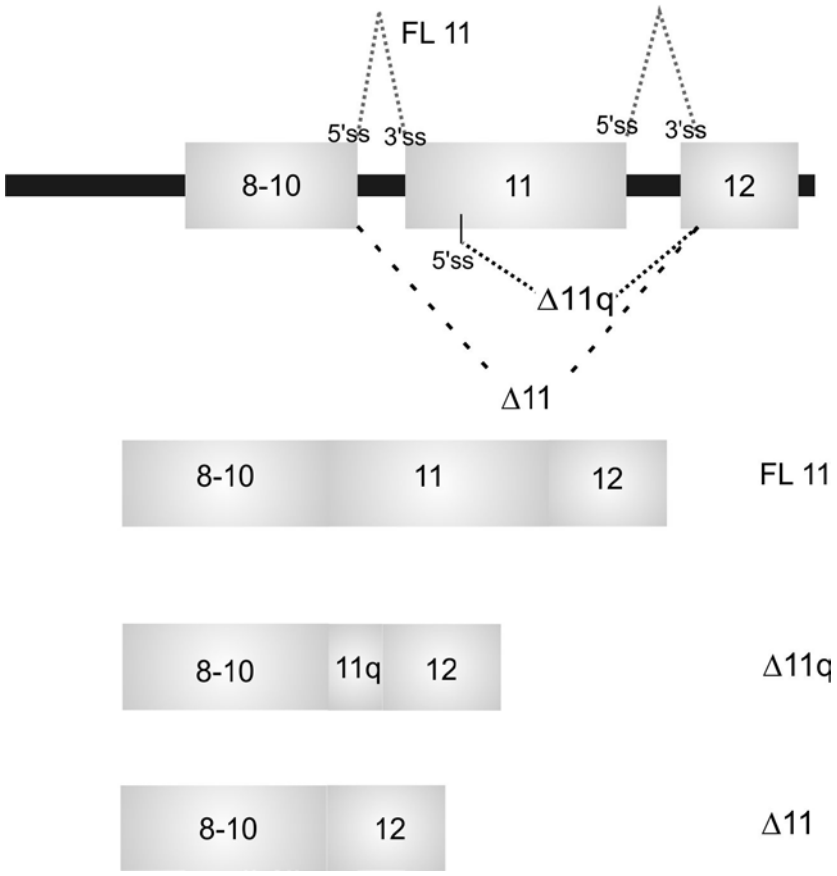


Figure 6.1: Schematic representation of the splice sites of exon 11.

The grey boxes represent exons; the black lines represent introns and the dotted lines represent the splicing outcome according with the use of exon 11 splice sites. The splice sites are represented with the abbreviation ss.

The Δ11 isoform results from skipping of exon 11 has been implicated in various aspects of cell proliferation and apoptosis (Bachelier et al., 2002; Cao et al., 2003; Huber et al., 2001; Kim et al., 2006; Tammaro et al., 2012).

Consequently current literature in association with the result shown in chapter 3 strongly suggests that the synonymous variant c.693G>A can induce predisposition to cancer. In this thesis the variant was demonstrated shown to cause an increase of Δ11 and decrease of Δ11q and full 11 isoforms altering the balance of the BRCA1 isoforms tumorigenic and apoptotic function.

6.2 The pB1 mini-gene as a useful tool for splicing and mechanistic investigation.

A mini-gene for splicing assay has been constructed and used in this study. The mini-gene used to analyse the unclassified variant c.693G>A that affected our patient contains exon 8 to exon 12 of the *BRCA1* gene and shows skipping of exon 11 with a decrease in the Δ11q isoform compared to pB1 WT mini-gene confirming that the sequence variant c.693G>A alters the splicing process of exon 11 (Chapter 3, Figure 3.6).

The result obtained using the mini-gene correlates well with that observed in the patient RNA and confirmed that this assay is a useful tool for testing the effects of unclassified variants on splicing (Chapter 3, Figure 3.6).

Before 2008 nobody had created a mini-gene containing exon 11, the biggest exon of the *BRCA1* gene, at around 3.4kb. In 2008 Olga Anczukow and co-workers designed and created a mini-gene spanning exons 10 to exon 12 to study 108 unclassified variants identified in BRCA1 exon 11 in a French genetic diagnostic centre (Anczukow et al., 2008). However, only one variant, c.3719 G>T was found to affect splicing by reducing intron 11 splicing. This information was not confirmed in a patient RNA.

In another recent study from our laboratory, the pB1 mini-gene constructed in this thesis was used to demonstrate that splicing regulatory sequences are present at the beginning of exon 11 and that these sequences are important for the production of the three isoforms (Raponi et al., 2012).

6.3 Detection of a composite splicing regulatory region.

Data presented in Chapter 3 showed that systematic site directed mutagenesis experiments using the pB1 mini-gene system suggest a complex regulatory region surrounding the c.693 G>A variant (Chapter 3, Figure 3.10).

In fact the 11 single nucleotide substitutions between codon position c.688 to c.694 showed different effects on the level of inclusion of exon 11. The nucleotide change in position c.689 showed a decrease of the Δ11 isoform and an increase of Δ11q isoform. The additional nucleotide substitutions created around position c.693 caused skipping of exon 11 and part of exon 11 to various degrees potentially inactivating ESEs and/or creating a new ESS.

These differences around this region suggest that the sequence may represent a Composite Regulatory Element of Splicing (CERES) with overlapping enhancer and silencer functions as found previously in *CFTR* exon 9 and 12. The CERES element identified in *CFTR* exon 12 appears to be strongly dependent on its context for its function and can also be extended to the flanking regions. The identification of a similar situation for exon 9 in CFTR indicates a common splicing regulatory role of these elements (Pagani et al., 2003a; Pagani et al., 2003b).

It is also interesting to note that it is possible to recreate a pB1WT sequence pattern in presence of a double mutant at position c.689A>C and c.693G>A. In fact, in the presence of c.689A>C, the c.693G>A does not affect splicing anymore (Chapter 3 Figure 3.13). This indicates that eventual compensatory nucleotide variants in the population could mask the effect of c.693G>A and prevent the alteration of exon 11 splicing.

6.4 In *vitro* identification of splicing proteins interacting with the regulatory sequence detected in *BRCA1*.

The Human Splicing Finder program predicted that the sequence surrounding the c.693 position of BRCA1 exon 11 consists of a high affinity binding sequence for SRSF1, SRSF5, and SRSF7 (Chapter 3, Table 3.1and Table 3.2).
Importantly the index of prediction is reduced for these splicing factors, if the synonymous variant c.693G>A is introduced which indicates that loss of SRSF1, SRSF5, and SRSF7 binding may cause exon 11 skipping.
However pull down experiments using wild type RNA oligo and an RNA oligo carrying the synonymous variant c.693G>A failed to show disruption of all binding predicted by Human Splicing Finder. Instead the pull down experiments using Coomassie staining detection demonstrated that the wild type sequence binds TDP43, whereas the sequence incorporating the synonymous change binds hnRNPA1. Mass spectrometry sequencing confirmed the presence of these two proteins and a summary listing the isolated proteins is given in Table 4.1 (Chapter 4). In chapter 4, I found that TDP43 and hnRNPA1 were able to bind with different affinities to both sequences (Figure 4.3 and Figure 4.4).

TDP43 (trans-activation response DNA-binding protein) is like many others heterogeneous ribonucleoproteins (hnRNP) protein family members (Ou et al., 1995) which was first described as a DNA binding protein .(Strong et al., 2007) Subsequently it was also shown that TDP43 is able to bind RNA, through the RRM1 motif (Buratti and Baralle, 2001; Kuo et al., 2009; Shodai et al., 2013; Winton et al., 2008).
TDP43 is involved in the regulation of splicing, translation and transcription. It has been shown that TDP43 has a preferred target sequence that is represented by the dinucleotide repeats (UG)n or (TG)n (Ayala et al., 2005; Buratti and Baralle, 2001).

Concerning the role of TDP43 in splicing regulation, two main mechanisms of action for this protein have been described. Evidence suggests that TDP43 acts as a negative splicing regulator binding to intronic UG repeats element and causes skipping of the *CFTR* gene exon 9 and of ApoAIII exon 3 (Buratti et al., 2001; Mercado et al., 2005). The role of TDP43 has been studied in the *CFTR* gene, where TDP43 has the ability to inhibit the inclusion of exon 9 through association with the UG repeat situated near the 3'splice site and creates a non-functional protein (Buratti et al., 2001).

A second mechanism of TDP43 splicing regulation has also been described, that involves binding of TDP43 to GU repeats near weak 5'splice sites and thus modulating exon inclusion (Passoni et al., 2012).

The results reported in this thesis were unexpected as they suggest binding of TDP43 within an exonic region lacking UG repeats. A recent study reported a new TDP43 binding consensus sequence to be a pyrimidine rich motif and UG rich sequence of a variable length (Sephton et al., 2011; Tollervey et al., 2011).

A small number of U repeats are present in the synthetic RNA oligonucleotide used in the pulldown assay (Chapter 4, Figure 4.1). Therefore it could be speculated that the ability of TDP43 to regulate splicing by binding pyrimidine rich motifs could be reduced to a small number of U repeats.

However as there is potential binding of TDP43 in both the wild type sequence and in the variant sequence, albeit with different intensity, it is difficult to assess whether this is an important factor in the regulation of BRCA1 exon 11 splicing.

hnRNP A1 belongs to an important class of RNA binding proteins thought to inhibit the use of proximal splice sites. This activity is opposite to that of SR proteins and as an added complication, the SR protein SRSF1 can compete with hnRNPA1 for the binding of exonic regulatory elements, (Zerbe et al., 2004; Zhu et al., 2001), also described for the insulin receptor gene (Talukdar et al., 2011) and Gs α gene (Pollard et al., 2002).

The ratio of SRSF1 and hnRNP A1 expression in normal breast compared to breast cancer tissues has not been reported, however different distributions of

hnRNP A1 and SRSF1 in different tissues have been described. In particular Hanamura et al suggested that a change in the ratio of expression of these two proteins could effect the alternative splicing of many genes (Hanamura et al., 1998).

Therefore, the different proportions of BRCA1 exon 11 splicing isoforms observed in different cellular contexts could be due to a difference in the ratio of hnRNPA1 and SRSF1 expression (Hanamura et al., 1998).

Here pull down assay suggested that the binding of SRSF1 is lost with the variant c.693G>A, while hnRNPA1 is still able to bind. Therefore it is possible that putative competition between SRSF1 and hnRNPA1 is lost with the variant, where hnRNPA1 can now bind undisturbed and reduce the use of the exon 11 3'splice site.

Besides SRSF1, the pull down also showed binding in the wild type sequence of SRSF9 and SRSF6, which is disrupted in the presence of the synonymous change c.693G>A (Chapter4 Figure 4.6).

This result supports the hypothesis that inclusion of exon 11 is due to the presence of different SR proteins, and the possible presence of a composite regulatory element (see section Discussion 6.2).

SRSF1 belongs to the SR protein family. These are enhancer proteins of splicing as they usually bind exonic RNA sequences and assist the splicing machinery to recognize splice sites. However in some cases it has been shown that SRSF1 promotes exon skipping, but its mode of action in this negative way is still unclear (Clery et al., 2013).

SRSF1 is over-expressed in cancer and acts as an oncogene by effecting splicing of several genes involved in cell transformation (Shimoni-Sebag et al., 2013).

There are several mutations reported to alter the binding of SRSF1 and cause aberrant splicing and disease. For instance, in spinal muscle atrophy a variant in exon 7 of the *SMN* gene causes exon 7 exclusion , caused by deletion of an exon

splicing enhancer (ESE) associated with binding of SRSF1(Koed Doktor et al., 2011; Singh et al., 2004).
The fact that several splicing mutations are found to cause aberrant splicing by disrupting the SRSF1 binding site suggests a fundamental role of this regulatory protein in maintaining splicing fidelity.
Since SRSF1 may control the splicing of several genes, maintenance of balanced expression is important. This is also an explanation as to why overexpression of SRSF1 causes cell transformation.

The sequence variant c.693 G>A appears to disrupt binding of SRSF1 and cause increased production of the Δ11 isoform which is involved in cell transformation (Chapter 4, Figure 4.5).
SRSF1 has previously been shown to be important for the constitutive inclusion of BRCA1 exon 6, which was demonstrated by depletion of SRSF1 in Breast cancer cells line (Raponi et al., 2011).
In the present study depletion of SRSF1 did not induce exon 11 exclusion in the pB1 WT mini-gene (Chapter 4, Figure 4.5), indicating that SRSF1 may be not essential for BRCA1 exon 11 splicing.
Knock down experiments demonstrated that silencing SRSF1 causes a decrease in the amount of Δ11 isoform and an increase of full 11 isoform in the pB1 WT mini-gene but has no effect on pB1 c.693 G>A mini-gene.
A possible explanation for this could be that unusually SRSF1 has an inhibitory effect on exon 11 which seems to be dependent on the regulatory region around position c.693.

Another SR protein that has been involved in this region is SRSF6. Pull down experiments followed by western blot demonstrated that SRSF6 is able to bind the wild type sequence and that binding is disrupted by the variant c.693G>A (Chapter 4, Figure 4.6).

SRSF6 is an SR protein that has been implicated in the regulation of the apoptotic gene *BIM* (Jiang et al., 2010), and in the regulation of the epidermal growth factors, inducing the angiogenic isoform VEGF (Carter et al., 2011).
Using knock down experiments, SRSF6 has previously been shown to regulate the inclusion of Tau exon 10, where mis-regulation of alternative splicing of Tau exon 10 causes neurodegenerative disease (Yin et al., 2012).
In this thesis SRSF6 knock down resulted in a relative decrease of the Δ11q isoform towards an increase of the Full 11 isoform, suggesting that when SRSF6 is not present the 5' splice site of the Δ11q may not be used. No effect was observed on splicing using the pB1 c.693G>A mutant mini-gene following depletion of SRSF6. This suggests that SRSF6 acts as an enhancer of the Δ11q 5' splice site by binding the regulatory region surrounding position c.693.

I have shown that both SRSF1 and SRSF6 are able to bind the wild type sequence *in vitro,* in addition to that, western blot analysis showed that another SR protein (SRSF9) is able to bind the wild type sequence but not the synonymous variant sequence (Chapter 4, Figure 4.6).
SRSF9 (SRp30c) has most similarity with the splicing factors SRSF1 (SRp30a) and SRSF2 (SRp30b). SRSF9 has a direct contact with RNA through two RNA recognition motifs that are linked by a glycine and act as a repressor of the use of the 3' splice site (Simard and Chabot, 2002).
SRSF9 is also able to determine inclusion of exons. In particular has been demonstrated that SRSF9 can stimulate high levels of SMN2 (inducing inclusion of SMN2 exon 7) especially when associated with SFSR10 which had already been shown to interact with the AG-rich enhancer sequence located in the center of SMN exon 7. Therefore the association of SRSF9 and SFRS10 could compensate for the C/T exchange in SMN2 (Young et al., 2002).
Other experiments demonstrate a cooperative interaction between *trans*-acting SR proteins SFRS10, SRSF7 and SRSF9 in ESE-dependent gonotropin releasing hormone pre-mRNA splicing (Park et al., 2006).

In this thesis I showed that SRSF7 is not able to bind either the wild type or synonymous sequence variants c.693G>A, while SFRS10 binds both (Chapter 4, Figure 4.7, Figure 4.10, and Figure 4.11).
At the distal end of exon 11 there is a repeated binding site for SFRS10 (personal communication of Raponi) important for the recognition of the distal donor site and for the inclusion of exon 11. In a similar way binding of SFRS10 in the region next to nucleotide position c.693 may favour the recognition of exon 11q 5' splice site and therefore the production of the Δ11q isoform. However this hypothesis was not verified in this thesis, because SRSF10 binding is not affected by the variant c.693G>A.

In conclusion, several splicing regulatory proteins have the potential to bind the regulatory region identified in this thesis highlighting the presence of a CERES in BRCA1 exon 11. However the unpredictable effect of depletion and overexpression of those regulatory proteins shown in chapter 4 underlines the difficulty in identifying which splicing factors are responsible for exclusion/inclusion of exon 11 in the presence of the synonymous variant.

pB1 WT	Overexpression	Depletion
SRSF1	↑ Full 11 ↓ Δ11q ↓ Δ11	↑ Full 11 ↓ Δ11
SRSF6	↑ Full 11 ↓ Δ11	↑ Full 11 ↓ Δ11q

pB1 c.693G>A	Overexpression	Depletion
SRSF1	↑ Full 11 ↓ Δ11	
SRSF6	↑ Δ11q ↓ Δ11	

Figure 6.2: Effect of depletion or overexpression of SRSF1 or SRSF6 in pB1 WT and pB1 c.693 G>A mini-gene.

The effect in isoform increase (↑) or decrease (↓) is reported in correspondence of each SR protein (SRSF1, SRSF6) overexpression or depletion. The effect on pB1 WT mini-gene is reported at the top and the effect on pB1 c.693 G>A mini-gene is reported at the bottom.

.

6.5 Correction with bifunctional oligonucleotide.

Correction of aberrant splicing using oligonucleotide based technology is a new growing field in genetic therapy.

The c.693G>A variant was shown to disrupt an enhancer sequence in BRCA1 exon 11 and cause aberrant splicing. I tested the hypothesis that this aberrant splicing could be restored by using a bifunctional oligonucleotide with a tail for an SR protein able to recruit SRSF1 (syn oligo).

However, when I undertook this, I found no effect on the inclusion of exon 11 (Chapter 5, Figure 5.1). Conversely, this bifunctional oligonucleotide increased exclusion of the exon 11 suggesting an inhibitory effect on the recognition of the 3'splice site (Figure 6.3). This unexpected result may be due to the fact that the positioning of the oligonucleotide is important for exon inclusion.

In fact several studies indicate that the target positions of bifunctional oligonucleotide are crucial for proper function. A further Del1 oligonucleotide was therefore designed to target a neutral region situated in the beginning of exon 11, close to the 3'splice site (Chapter 5. Figure 5.1).

The Del1 oligonucleotide had a positive effect only on the inclusion of the Δ11q isoform (Chapter 5, Figure 5.4). Whilst it is not clear what causes inclusion of exon 11q; it appears that having a bifunctional oligonucleotide closer to the 3'splice site does not increase whole exon 11 inclusion but increases the delta 11q isoform. This could be due to the presence of the SR tail, which can have a positive effect on the weak 5' splice site of 11q (Figure 6.4).

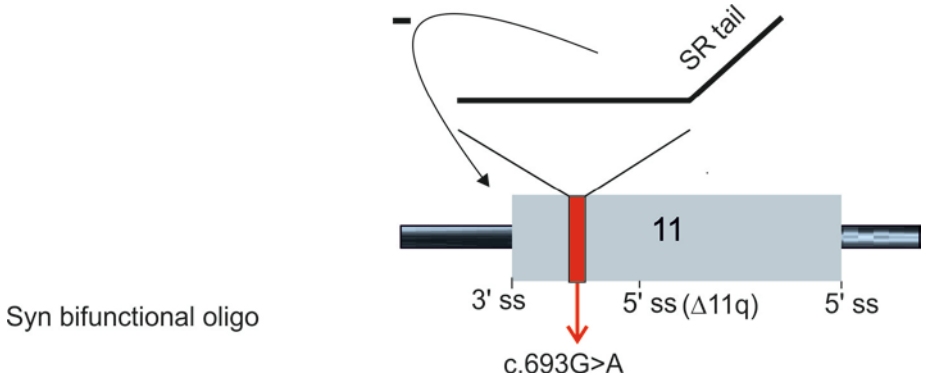


Figure 6.3: Model of action of the Syn bifunctional oligonucleotide.

Schematic representation of the bifunctional oligonucleotide created to bind the region with the synonymous variant (red box). The arrow indicates a negative effect on the recognition of the 3'splice site. The figure shows a schematic mechanism of action of the SR tail of the bifunctional oligo, suggesting that the SR tail has a negative effect on the 3' splice site of exon 11.

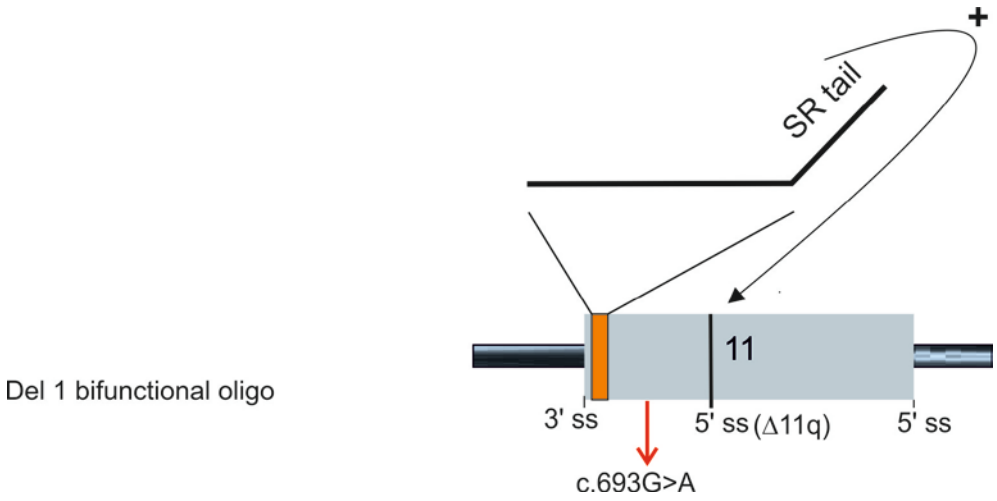


Figure 6.4: Model of action of the Del1 bifunctional oligonucleotide.

Schematic representation of the bifunctional oligonucleotide Del1 designed to bind a neutral region close to the 3' splice site (orange box). The arrow indicates a positive effect on the recognition of the 5'splice site (D11q). The figure shows a schematic mechanism of action of the SR tail of the bifunctional oligo, suggesting that the SR tail has a positive effect on the 5' splice site of the exon 11q, causing inclusion of Δ11q isoform.

.

A third bifunctional oligonucleotide (Del7a) at the end of exon 11 was investigated. This was expected to restore the inclusion of exon 11 in the presence of the variant c.693G>A, however the bifunctional oligonucleotide increased exon 11 skipping in the pB1 c.693 G>A mini-gene suggesting an inhibitory effect on the recognition of exon 11 5'splie site (Figure 6.5).

Taken together these results have shown that the syn oligo and the Del7a oligo cannot be used to correct aberrant splicing of the c.693 G>A variant. However the Del1 oligo increased the Δ11q isoform, suggesting that binding of an enhancer in that region is able to stimulate the recognition of the Δ11q 5' splice site and to partially rescue the inclusion of part of exon 11. Therefore Del1 oligo could be tested in the future to correct eventual splicing variants that cause a decrease of the D11q isoform.

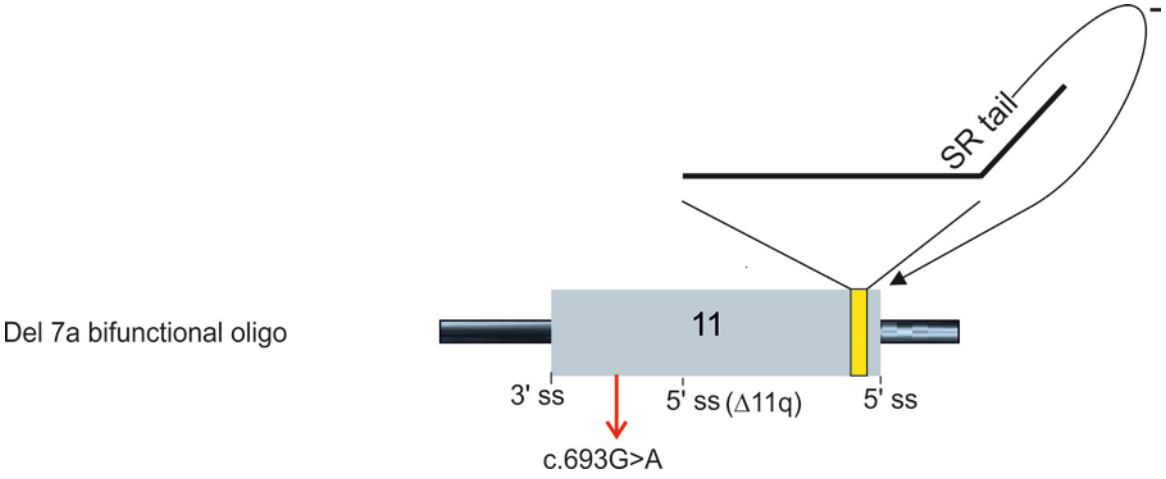


Figure 6.5: Model of action of the Del7a bifunctional oligonucleotide.

Schematic representation of the bifunctional oligonucleotide Del7a designed to bind a neutral region close to the 5' splice site (yellow box). The arrow indicates a negative effect on the recognition of the 5'splice site. The figure shows a schematic mechanism of action of the SR tail of the bifunctional oligo, suggesting that the SR tail has a negative effect on the 5' splice site of the exon 11, causing increase of Δ11 isoform.

6.6 Future work.

This study shows that the characterization of a nucleotide variant involved in a pre-mRNA processing defect should not be done without considering its natural context.
In recent years the complexity of regulatory elements that participate in splicing control has taken an enormous step forward with the discovery of potential connections between splicing, the specific structural features of pre-RNAs and many of the processes that participate in their life-cycle such as genome stability, RNA processing kinetics, transport and translation.

Following the construction of a mini-gene for BRCA exon 11 splicing assay and the discovery of composite regulatory elements that regulate processing of this exon two major lines of investigation could be undertaken:

1. Investigation of splicing regulation of large exons;
2. Restoration of BRCA1 exon inclusion.

1. **Investigate splicing regulation of large exons**.

The pB1 mini-gene will be useful to identify and study other unclassified variants situated around exon 11. It could be employed to better understand these unclassified variants in correlation with the alteration of the splicing process and thus identify the splicing mechanisms that regulate BRCA1 alternative splicing. This is of particular importance as it helps clarify the splicing regulation of unusually large exons. which are currently not fully understood.

In addition of using the pB1 mini-gene, these mechanisms could be investigated using RNA immunoprecipitation experiments from breast cancer cell lines and *in vitro* splicing assay. *In vivo* RNA immunoprecipitation is an antibody based technique used to map RNA and protein interactions through immunoprecipitation of a particular protein after UV crosslinking of cell lines. The RNA cross-linked to the protein can be detected through RT-PCR and quantitative PCR. This is particularly useful to demonstrate *in vivo* binding of a splicing factor to a specific region when this cannot be demonstrated with siRNA/overexpression experiments. Because of the long nucleotide sequence of BRCA1 exon 11 the same splicing factor could bind in several different regions. This factor could act as an enhancer or silencer depending on position, and therefore mask the effect of siRNA.

2. **Restoration of BRCA exon 11 inclusion.**

Regarding the restoration of c.693G>A BRCA1 exon 11 skipping, the use of bifunctional oligonucleotides failed to induce inclusion of the entire exon 11 in this study, however it was possible to recreate inclusion of the delta 11q isoform with one of the bifunctional oligonucleotides tested. Consequently this bifunctional oligonucleotide could be exploited to correct eventual mutations that are found to cause skipping of exon 11q only. In addition this bifunctional oligonucleotide could prove useful to manipulate the level of Δ11q isoform in cell lines; allowing functional study on the effect of Δ11q levels and functional studies

on the effect of Δ11q levels in the cell in response to DNA damage and chemotherapy.

Appendices

7 Appendices A.

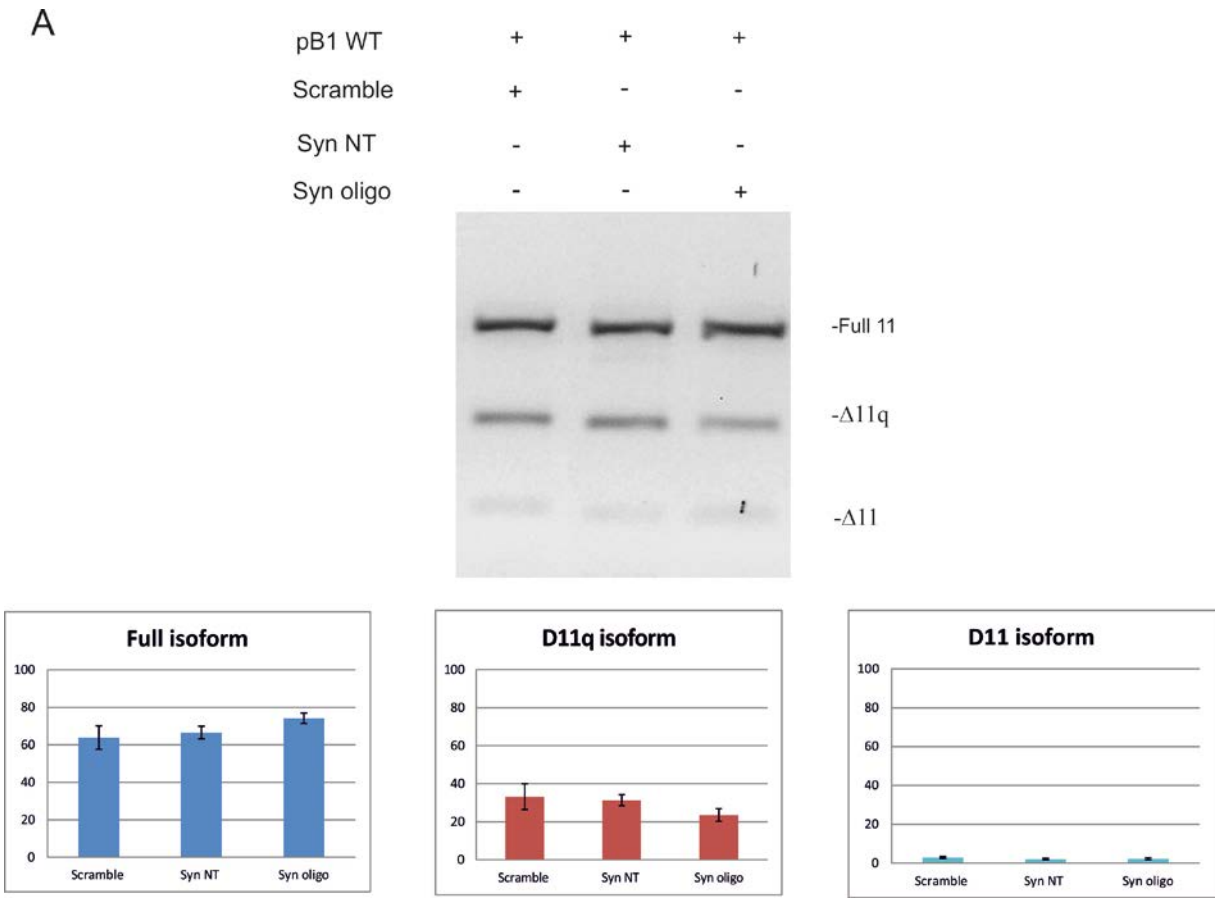


Figure 7.1:Co-transfection of bifunctional oligonucleotides on the syn variation and pB1 WT mini-gene.

Syn NT is the bifunctional oligonucleotide without the tail and the scramble oligo represents the control. The PCR product is separated on a 1.5% agarose gel. The histograms show the percentage of the full isoform the Δ11q isoform and the Δ11 isoform respectively, and it has been calculated against the total expression of the three isoforms (Methods 2.5.6).
(**A**) RT-PCR products from total RNA extracted after co-transfection of Syn oligo and pB1 WT mini-gene and histograms showing the percentage of the isoforms.

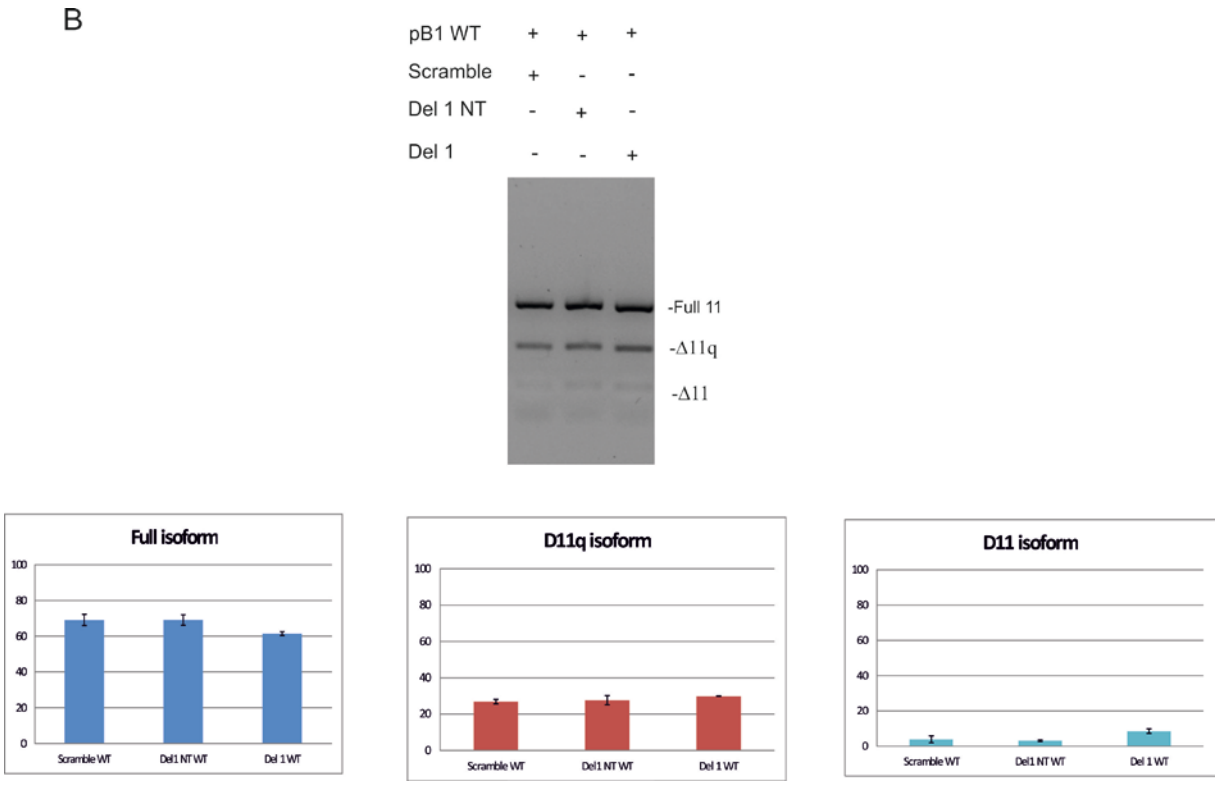


Figure 7.2: Co-transfection of bifunctional oligonucleotides Del1 and pB1 WT mini-gene.

Del1 NT is the bifunctional oligonucleotide without the tail and the scramble oligo represents the control. The PCR product is separated on a 1.5% agarose gel. The histograms show the percentage of the full isoform the Δ11q isoform and the Δ11 isoform respectively, and it has been calculated against the total expression of the three isoforms (Methods 2.5.6).
(**B**) RT-PCR products from total RNA extracted after co-transfection of Del1 oligo and pB1 WT mini-gene and histograms showing the percentage of the isoforms.

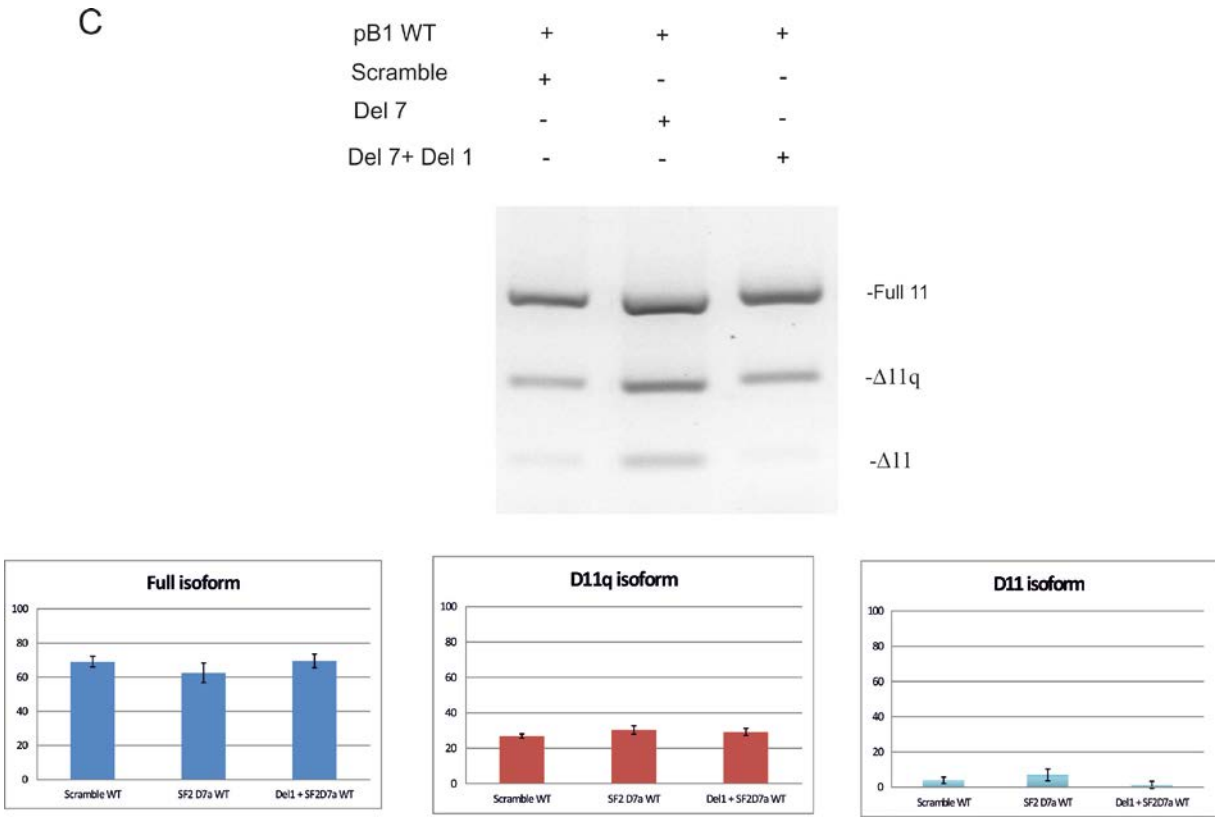


Figure 7.3: Co-transfection of bifunctional oligonucleotides Del7 and pB1 WT mini-gene.

The scramble oligo represents the control. The PCR product is separated on a 1.5% agarose gel. The histograms show the percentage of the full isoform the Δ11q isoform and the Δ11 isoform respectively, and it has been calculated against the total expression of the three isoforms (Methods 2.5.6).
(**C**) RT-PCR products from total RNA extracted after co-transfection of Del7 oligo and pB1 WT mini-gene and histograms showing the percentage of the isoforms

8 Appendices B.

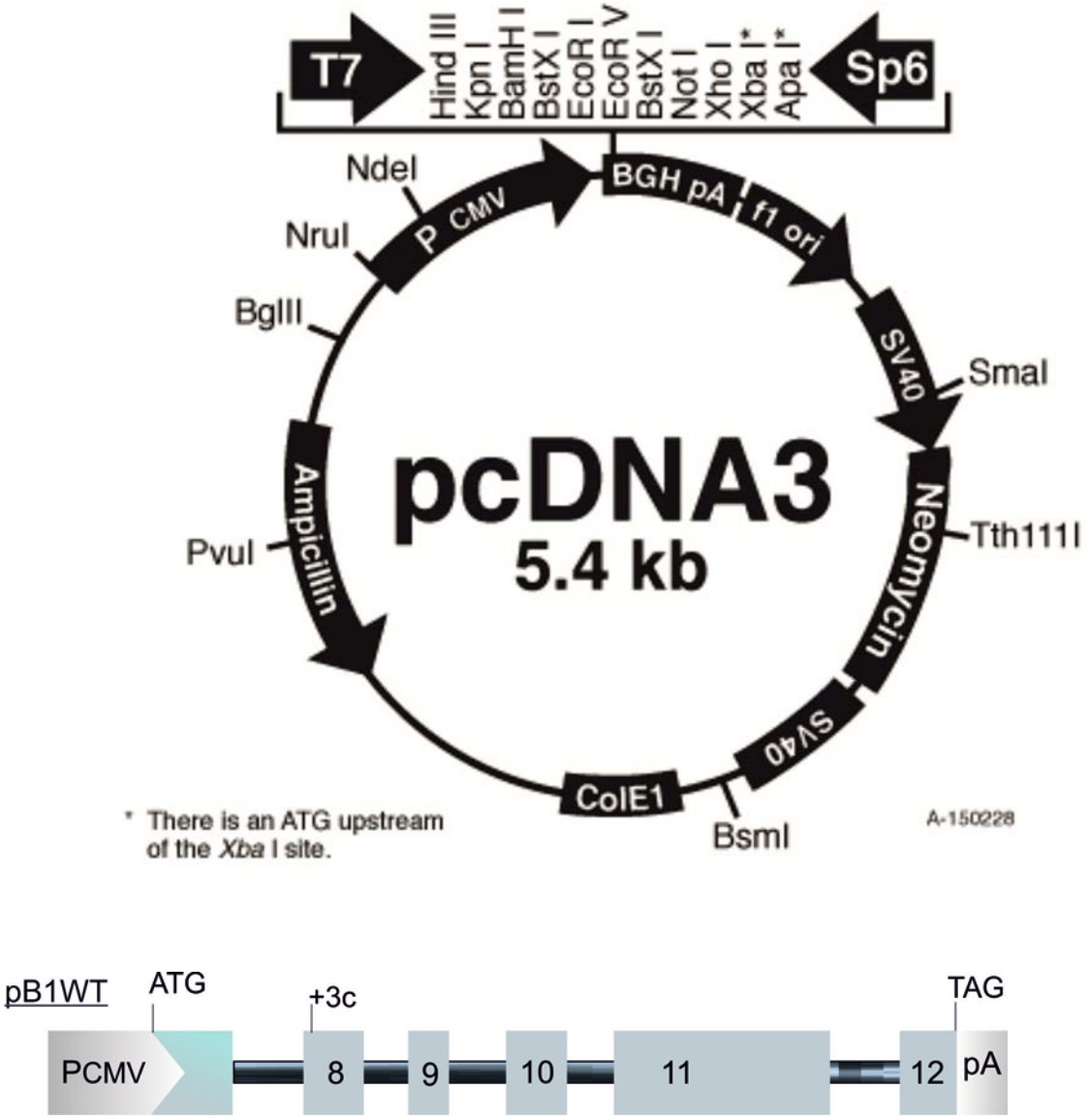


Figure 8.1: Vector pcDNA3.

La figure shows the pcDNA 3 vector used in this thesis (upper panel) and the schematic representations of the pB1 mini-gene (low panel).

ATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCCTGGGGTAAGGT
CGGCGCGCACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAGgtgaggctccc
tcccctgctccgacccgggctcctcgcccgcccggacccacaggccaccctca
accgtcctggccccggacccaaaccccacccctcactctgcttctccccgcag
GAcAACCAGTCTCAGTGTCCAACTCTCTAACCTTGGAACTGTGAGAACTCTGA
GGACAAAGCAGCGGATACAACCTCAAAAGACGTCTGTCTACATTGAATTGGgt
aagggtctcaggtttttaagtatttaataataattgctggattccttatctt
atagttttgccaaaaatcttggtcataatttgtatttgtggtaggcagctttg
ggaagtgaatttatgagccctatggtgagttataaaaaatgtaaaagacgca
gttcccaccttgaagaatcttactttaaaaagggagcaaaagaggccaggcat
ggtggctcacacctgtaatcCtcaaataatccacccatctcggcctcctcaag
tgctgggattacaggtgagagccactgtgcctggcgaagcccatgcctttaac
cacttctctgtattacatactagcttaactagcattgtacctgccacagtaga
tgctcagtaaatatttctagttgaatatctgtttttcaacaagtacatttttt
taacccttttaattaagaaaactttattgatttatttttgggggaaattt
tttagGATCTGATTCTTCTGAAGATACCGTTAATAAGGCAACTTATTGCAGgt
gagtcaaagagaacctttgtctatgaagctggtattttcctatttagttaata
ttaaggattgatgtttctctcttttaaaaatattttaacttttattttaggt
tcagggatgtatgtgcagtttgttatataggtaaacacacgacttgggatttg
gtgtatagatttttttcatcatccgggtactaagcataccccacagttttttg
tttgctttctttctgaatCcacccgcctcagcctcccaaagtgctggaattac
aggcttgagccaccgtgcccagcaaccatttcatttcaactagaagtttctaa
aggagagagcagctttcactaactaaataagattggtcagctttctgtaatcg
aaagagctaaaatgtttgatcttggtcatttgacagttctgcatacatgtaac
tagtgtttcttattaggactctgtcttttccctatagTGTGGGAGATCAAGAA
TTGTTACAAATCACCCCTCAAGGAACCAGGGATGAAATCAGTTTGGATTCTGC
AAAAAAGGgtaatggcaaagtttgccaacttaacaggcactgaaaagagagtg
ggtagatacagtactgtaattagattattctgaagaccatttgggacctttac
aacccacaaaatctcttggcagagttagagtatcattctctgtcaaatgtcgt
ggtatggtctgatagatttaaatggtactagactaatgtacctataataagac
cttctgtaactgattgttgccctttcgttttttttttgtttgtttgtttgtt
ttttttgagatggggtctcactctgttgcccaggctggagtgcagtgatgca
atcttggctcactgcaacctccacctccaaggctcaagctattattaagtaga
cttaatggccagtaatctttagagtacatcagaaccagttttctgatggccaa
tctgcttttaattcactcttagacgttagagaaataggtgtggtttctgcata
gggaaaattctgaaattaaaaatttaatggatcctaagtggaaataatctagg
taaataggaattaaatgaaagagtatgagctacatcttcagtatacttggtag
tttatgaggttagtttctctaatatagccagttggttgatttccacctccaag
gtgtatgaagtatgtatttttttaatgacaattcagtttttgagtaccttgtt
atttttgtatattttcagCTGCTTGTGAATTTTCTGAGACGGATGTAACAAAT
ACTGAACATCATCAACCCAGTAATAATGATTTGAACACCACTGAGAAGCGTGC
AGCTGAGAGGCATCCAGAAAAGTATCAGGGTAGTTCTGTTTCAAACTTGCATG
TGGAGCCATGTGGCACAAATACTCATGCCAGCTCATTACAGCATGAGAACAGC
AGTTTATTACTCACTAAAGACAGAATGAATGTAGAAAAGGCTGAATTCTGTAA

TAAAAGCAAACAGCCTGGCTTAGCAAGGAGCCAACATAACAGATGGGCTGGAA
GTAAGGAAACATGTAATGATAGGCGGACTCCCAGCACAGAAAAAAAGGTAGAT
CTGAATGCTGATCCCCTGTGTGAGAGAAAAGAATGGAATAAGCAGAAACTGCC
ATGCTCAGAGAATCCTAGAGATACTGAAGATGTTCCTTGGATAACACTAAATA
GCAGCATTCAGAAAGTTAATGAGTGGTTTTCCAGAAGTGATGAACTGTTAGGT
TCTGATGACTCACATGATGGGGAGTCTGAATCAAATGCCAAAGTAGCTGATGT
ATTGGACGTTCTAAATGAGGTAGATGAATATTCTGGTTCTTCAGAGAAAATAG
ACTTACTGGCCAGTGATCCTCATGAGGCTTTAATATGTAAAAGTGAAAGAGTT
CACTCCAAATCAGTAGAGAGTAATATTGAAGACAAAATATTTGGGAAAACCTA
TCGGAAGAAGGCAAGCCTCCCCAACTTAAGCCATGTAACTGAAAATCTAATTA
TAGGAGCATTTGTTACTGAGCCACAGATAATACAAGAGCGTCCCCTCACAAAT
AAATTAAAGCGTAAAAGGAGACCTACATCAGGCCTTCATCCTGAGGATTTTAT
CAAGAAAGCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATAAATCAGGGAA
CTAACCAAACGGAGCAGAATGGTCAAGTGATGAATATTACTAATAGTGGTCAT
GAGAATAAAACAAAAGGTGATTCTATTCAGAATGAGAAAAATCCTAACCCAAT
AGAATCACTCGAAAAAGAATCTGCTTTCAAAACGAAAGCTGAACCTATAAGCA
GCAGTATAAGCAATATGGAACTCGAATTAAATATCCACAATTCAAAAGCACCT
AAAAAGAATAGGCTGAGGAGGAAGTCTTCTACCAGGCATATTCATGCGCTTGA
ACTAGTAGTCAGTAGAAATCTAAGCCCACCTAATTGTACTGAATTGCAAATTG
ATAGTTGTTCTAGCAGTGAAGAGATAAAGAAAAAAAAGTACAACCAAATGCCA
GTCAGGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACCTGCAACTGG
AGCCAAGAAGAGTAACAAGCCAAATGAACAGACAAGTAAAAGACATGACAGCG
ATACTTTCCCAGAGCTGAAGTTAACAAATGCACCTGGTTCTTTTACTAAGTGT
TCAAATACCAGTGAACTTAAAGAATTTGTCAATCCTAGCCTTCCAAGAGAAGA
AAAAGAAGAGAAACTAGAAACAGTTAAAGTGTCTAATAATGCTGAAGACCCCA
AAGATCTCATGTTAAGTGGAGAAAGGGTTTTGCAAACTGAAAGATCTGTAGAG
AGTAGCAGTATTTCATTGGTACCTGGTACTGATTATGGCACTCAGGAAAGTAT
CTCGTTACTGGAAGTTAGCACTCTAGGGAAGGCAAAAACAGAACCAAATAAAT
GTGTGAGTCAGTGTGCAGCATTTGAAAACCCCAAGGGACTAATTCATGGTTGT
TCCAAAGATAATAGAAATGACACAGAAGGCTTTAAGTATCCATTGGGACATGA
AGTTAACCACAGTCGGGAAACAAGCATAGAAATGGAAGAAAGTGAACTTGATG
CTCAGTATTTGCAGAATACATTCAAGGTTTCAAAGCGCCAGTCATTTGCTCCG
TTTTCAAATCCAGGAAATGCAGAAGAGGAATGTGCAACATTCTCTGCCCACTC
TGGGTCCTTAAAGAAACAAAGTCCAAAAGTCACTTTTGAATGTGAACAAAAGG
AAGAAAATCAAGGAAAGAATGAGTCTAATATCAAGCCTGTACAGACAGTTAAT
ATCACTGCAGGCTTTCCTGTGGTTGGTCAGAAAGATAAGCCAGTTGATAATGC
CAAATGTAGTATCAAAGGAGGCTCTAGGTTTTGTCTATCATCTCAGTTCAGAG
GCAACGAAACTGGACTCATTACTCCAAATAAACATGGACTTTTACAAAACCCA
TATCGTATACCACCACTTTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAA
GAAAAATCTGCTAGAGGAAAACTTTGAGGAACATTCAATGTCACCTGAAAGAG
AAATGGGAAATGAGAACATTCCAAGTACAGTGAGCACAATTAGCCGTAATAAC
ATTAGAGAAAATGTTTTTAAAGAAGCCAGCTCAAGCAATATTAATGAAGTAGG
TTCCAGTACTAATGAAGTGGGCTCCAGTATTAATGAAATAGGTTCCAGTGATG
AAAACATTCAAGCAGAACTAGGTAGAAACAGAGGGCCAAAATTGAATGCTATG

CTTAGATTAGGGGTTTTGCAACCTGAGGTCTATAAACAAAGTCTTCCTGGAAG
TAATTGTAAGCATCCTGAAATAAAAAAGCAAGAATATGAAGAAGTAGTTCAGA
CTGTTAATACAGATTTCTCTCCATATCTGATTTCAGATAACTTAGAACAGCCT
ATGGGAAGTAGTCATGCATCTCAGGTTTGTTCTGAGACACCTGATGACCTGTT
AGATGATGGTGAAATAAAGGAAGATACTAGTTTTGCTGAAAATGACATTAAGG
AAAGTTCTGCTGTTTTTAGCAAAAGCGTCCAGAAAGGAGAGCTTAGCAGGAGT
CCTAGCCCTTTCACCCATACACATTTGGCTCAGGGTTACCGAAGAGGGGCCAA
GAAATTAGAGTCCTCAGAAGAGAACTTATCTAGTGAGGATGAAGAGCTTCCCT
GCTTCCAACACTTGTTATTTGGTAAAGTAAACAATATACCTTCTCAGTCTACT
AGGCATAGCACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGGAGAATTT
ATTATCATTGAAGAATAGCTTAAATGACTGCAGTAACCAGGTAATATTGGCAA
AGGCATCTCAGGAACATCACCTTAGTGAGGAAACAAAATGTTCTGCTAGCTTG
TTTTCTTCACAGTGCAGTGAATTGGAAGACTTGACTGCAAATACAAACACCCA
GGATCCTTTCTTGATTGGTTCTTCCAAACAAATGAGGCATCAGTCTGAAAGCC
AGGGAGTTGGTCTGAGTGACAAGGAATTGGTTTCAGATGATGAAGAAAGAGGA
ACGGGCTTGGAAGAAAATAATCAAGAAGAGCAAAGCATGGATTCAAACTTAGg
tattggaaccaggtttttgtgtttgccccagtctatttatagaagtgagctaa
atgtttatgcttttggggagcacattttacaaatttccaagtatagttaaagg
aactgcttcttaaacttgaaacatgttcctcctaaggtgcttttcatagaaaa
aagtccttcacacagctaggacgtcatctttgactgaatgagctttaacatcc
taattactggtggacttacttctggtttcattttataaaagcaaatccaggtg
tcccaaagcaaggaatttaatcattttgtgtgacatgaaagtaaatccagtcc
tgccaatgagaagaaaaagacacagcaagttgcagcgtttatagtctgctttt
acatctgaacctctgtttttgttatttaagGTGAAGCAGCATCTGGGTGTGAG
AGTGAAACAAGCGTCTCTGAAGACTGCTCAGGGCTATCCTCTCAGAGTGACAT
TTTAACCAtctAGa

Figure 8.2: Sequence of the pB1 mini-gene.

The figure shows the sequence of the pB1 mini-gene. The introns are represented in low letters and the exon sequences from exon 8 to exon 12 are highlight and they are in capital letter.

References.

9 References.

Aartsma-Rus, A. (2010). Antisense-mediated modulation of splicing: therapeutic implications for Duchenne muscular dystrophy. RNA biology *7*, 453-461.

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

Anczukow, O., Buisson, M., Salles, M.J., Triboulet, S., Longy, M., Lidereau, R., Sinilnikova, O.M., and Mazoyer, S. (2008). Unclassified variants identified in BRCA1 exon 11: Consequences on splicing. Genes, chromosomes & cancer *47*, 418-426.

Aprelikova, O.N., Fang, B.S., Meissner, E.G., Cotter, S., Campbell, M., Kuthiala, A., Bessho, M., Jensen, R.A., and Liu, E.T. (1999). BRCA1-associated growth arrest is RB-dependent. Proceedings of the National Academy of Sciences *96*, 11866-11871.

Ars, E., Serra, E., García, J., Kruyer, H., Gaona, A., Lázaro, C., and Estivill, X. (2000). Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. Human molecular genetics *9*, 237-247.

Ayala, Y.M., Pantano, S., D'Ambrogio, A., Buratti, E., Brindisi, A., Marchetti, C., Romano, M., and Baralle, F.E. (2005). Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. Journal of molecular biology *348*, 575-588.

Bachelier, R., Vincent, A., Mathevet, P., Magdinier, F., Lenoir, G.M., and Frappart, L. (2002). Retroviral transduction of splice variant Brca1-Delta11 or mutant Brca1-W1777Stop causes mouse epithelial mammary atypical duct hyperplasia. Virchows Archiv : an international journal of pathology *440*, 261-266.

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

Behrens, S.E., and Luhrmann, R. (1991). Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. Genes Dev *5*, 1439-1452.

Behrens, S.E., Tyc, K., Kastner, B., Reichelt, J., and Luhrmann, R. (1993). Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. Mol Cell Biol *13*, 307-319.

Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol *14*, 336-342.

Berget, S.M. (1995). Exon recognition in vertebrate splicing. The Journal of biological chemistry *270*, 2411-2414.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proceedings of the National Academy of Sciences of the United States of America *74*, 3171-3175.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell *89*, 781-787.

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. RNA (New York, NY *16*, 2384-2403.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Luhrmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846-850.

Best, A., Dagliesh, C., Ehrmann, I., Kheirollahi-Kouhestani, M., Tyson-Capper, A., and Elliott, D.J. (2013). Expression of Tra2β in Cancer Cells as a Potential Contributory Factor to Neoplasia and Metastasis. International Journal of Cell Biology *2013*, 9.

Birney, E., Kumar, S., and Krainer, A.R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic acids research *21*, 5803-5816.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem *72*, 291-336.

Blaustein, M., Pelisch, F., and Srebrow, A. (2007). Signals, pathways and splicing regulation. The International Journal of Biochemistry & Cell Biology *39*, 2031-2048.

Boehringer, D., Makarov, E.M., Sander, B., Makarova, O.V., Kastner, B., Luhrmann, R., and Stark, H. (2004). Three-dimensional structure of a pre-catalytic human spliceosomal complex B. Nat Struct Mol Biol *11*, 463-468.

Boichard, A., Venet, L., Naas, T., Boutron, A., Chevret, L., de Baulny, H.O., De Lonlay, P., Legrand, A., Nordman, P., and Brivet, M. (2008). Two silent

substitutions in the PDHA1 gene cause exon 5 skipping by disruption of a putative exonic splicing enhancer. Molecular genetics and metabolism *93*, 323-330.

Boise, L.H., González-García, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nuñez, G., and Thompson, C.B. (1993). bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. Cell *74*, 597-608.

Boulton, S.J. (2006). Cellular functions of the BRCA tumour-suppressor proteins. Biochem Soc Trans *34*, 633-645.

Bourgeois, C.F., Lejeune, F., and Stevenin, J. (2004). Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. Prog Nucleic Acid Res Mol Biol *78*, 37-88.

Brandao, R.D., van Roozendaal, K., Tserpelis, D., Gomez Garcia, E., and Blok, M.J. (2011). Characterisation of unclassified variants in the BRCA1/2 genes with a putative effect on splicing. Breast cancer research and treatment *129*, 971-982.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proceedings of the National Academy of Sciences of the United States of America *75*, 4853-4857.

Buee, L., Bussiere, T., Buee-Scherrer, V., Delacourte, A., and Hof, P.R. (2000). Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. Brain research Brain research reviews *33*, 95-130.

Buratti, E., and Baralle, F.E. (2001). Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. The Journal of biological chemistry *276*, 36337-36343.

Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y.M., and Baralle, F.E. (2004). hnRNP H binding at the 5´ splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSH β genes. Nucleic acids research *32*, 4224-4236.

Buratti, E., Dhir, A., Lewandowska, M.A., and Baralle, F.E. (2007). RNA structure is a key regulatory element in pathological ATM and CFTR pseudoexon inclusion events. Nucleic acids research *35*, 4369-4383.

Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M., and Baralle, F.E. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. EMBO J *20*, 1774-1784.

Burge, C., and Sharp, P. (1999). Splicing of precursor to mRNAs by the spliceosome. In The RNA World, R.F. Gesteland., T.R. Cech., and J.F. Atkins., eds. (New York: Cold Spring Harbor Laboratory Press), pp. 525-560.

Caceres, J.F., Screaton, G.R., and Krainer, A.R. (1998). A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. Genes Dev *12*, 55-66.

Cao, L., Li, W., Kim, S., Brodie, S.G., and Deng, C.X. (2003). Senescence, aging, and malignant transformation mediated by p53 in mice lacking the Brca1 full-length isoform. Genes Dev *17*, 201-213.

Caputi, M., and Zahler, A.M. (2002). SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. Embo J *21*, 845-855.

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.

Cartegni, L., Hastings, M.L., Calarco, J.A., de Stanchina, E., and Krainer, A.R. (2006). Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. American journal of human genetics *78*, 63-77.

Cartegni, L., and Krainer, A.R. (2002). Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nature genetics *30*, 377-384.

Cartegni, L., and Krainer, A.R. (2003). Correction of disease-associated exon skipping by synthetic exon-specific activators. Nature structural biology *10*, 120-125.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic acids research *31*, 3568-3571.

Carter, J.G., Cherry, J., Williams, K., Turner, S., Bates, D.O., and Churchill, A.J. (2011). Splicing factor polymorphisms, the control of VEGF isoforms and association with angiogenic eye disease. Current eye research *36*, 328-335.

Chakarova, C.F., Hims, M.M., Bolz, H., Abu-Safieh, L., Patel, R.J., Papaioannou, M.G., Inglehearn, C.F., Keen, T.J., Willis, C., Moore, A.T., *et al.* (2002). Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. Human molecular genetics *11*, 87-92.

Chao, H., Mansfield, S.G., Bartel, R.C., Hiriyanna, S., Mitchell, L.G., Garcia-Blanco, M.A., and Walsh, C.E. (2003). Phenotype correction of hemophilia A mice by spliceosome-mediated RNA trans-splicing. Nature medicine *9*, 1015-1019.

Cho, S., Hoang, A., Sinha, R., Zhong, X.-Y., Fu, X.-D., Krainer, A.R., and Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. Proceedings of the National Academy of Sciences.

Clery, A., Sinha, R., Anczukow, O., Corrionero, A., Moursy, A., Daubner, G.M., Valcarcel, J., Krainer, A.R., and Allain, F.H. (2013). Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. Proceedings of the National Academy of Sciences of the United States of America *110*, E2802-2811.

Corrionero, A., Raker, V.A., Izquierdo, J.M., and Valcarcel, J. (2011). Strict 3' splice site sequence requirements for U2 snRNP recruitment after U2AF binding underlie a genetic defect leading to autoimmune disease. RNA (New York, NY *17*, 401-411.

Coulter, L.R., Landree, M.A., and Cooper, T.A. (1997). Identification of a new class of exonic splicing enhancers by in vivo selection. Mol Cell Biol *17*, 2143-2150.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Crispino, J.D., Blencowe, B.J., and Sharp, P.A. (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. Science *265*, 1866-1869.

Danglot, G., Régnler, V., Fauvet, D., Vassal, G., Kujas, M., and Bernheim, A. (1995). Neurofibromatosis 1 (NF1) mRNAs expressed in the central nervous system are differentially spliced in the 5´ part of the gene. Human molecular genetics *4*, 915-920.

David, C.J., and Manley, J.L. (2008). The search for alternative splicing regulators: new approaches offer a path to a splicing code. Genes Dev *22*, 279-285.

Deng, C.-X. (2006a). BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. Nucleic acids research *34*, 1416-1426.

Deng, C.-X., and Wang, R.-H. (2003). Roles of BRCA1 in DNA damage repair: a link between development and cancer. Human molecular genetics *12*, R113-R123.

Deng, C.X. (2006b). BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. Nucleic acids research *34*, 1416-1426.

Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., and Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research *37*, e67.

Dias, N., and Stein, C.A. (2002). Antisense oligonucleotides: basic concepts and mechanisms. Molecular cancer therapeutics *1*, 347-355.

Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol *3*, 195-205.

Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., and Burd, C.G. (1993). hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem *62*, 289-321.

Dunckley, M.G., Manoharan, M., Villiet, P., Eperon, I.C., and Dickson, G. (1998). Modification of splicing in the dystrophin gene in cultured Mdx muscle cells by antisense oligoribonucleotides. Human molecular genetics *7*, 1083-1090.

Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S.M., Driver, D.A., Berg, R.H., Kim, S.K., Norden, B., and Nielsen, P.E. (1993). PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. Nature *365*, 566-568.

Eperon, I.C., Makarova, O.V., Mayeda, A., Munroe, S.H., Cáceres, J.F., Hayward, D.G., and Krainer, A.R. (2000). Selection of Alternative 5 ´ Splice Sites: Role of U1 snRNP and Models for the Antagonistic Effects of SF2/ASF and hnRNP A1. Molecular and Cellular Biology *20*, 8303-8318.

Estibeiro, P., and Godfray, J. (2001). Antisense as a neuroscience tool and therapeutic agent. Trends in neurosciences *24*, S56-62.

Fabbro, M., Rodriguez, J.A., Baer, R., and Henderson, B.R. (2002). BARD1 induces BRCA1 intranuclear foci formation by increasing RING-dependent BRCA1 nuclear import and inhibiting BRCA1 nuclear export. The Journal of biological chemistry *277*, 21315-21324.

Fabbro, M., Savage, K., Hobson, K., Deans, A.J., Powell, S.N., McArthur, G.A., and Khanna, K.K. (2004). BRCA1-BARD1 Complexes Are Required for p53Ser-15 Phosphorylation and a G1/S Arrest following Ionizing Radiation-induced DNA Damage. Journal of Biological Chemistry *279*, 31251-31258.

Fairbrother, W.G., and Chasin, L.A. (2000). Human genomic sequences that inhibit splicing. Mol Cell Biol *20*, 6816-6825.

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

Fisette, J.-F., Toutant, J., Dugré-Brisson, S., Desgroseillers, L., and Chabot, B. (2010). hnRNP A1 and hnRNP H can collaborate to modulate 5 ´ splice site selection. RNA (New York, NY *16*, 228-238.

Friedberg, E.C. (2008). A brief history of the DNA repair field. Cell research *18*, 3-7.

Friedman, L.S., Ostermeyer, E.A., Szabo, C.I., Dowd, P., Lynch, E.D., Rowell, S.E., and King, M.C. (1994). Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. Nature genetics *8*, 399-404.

Giulietti, M., Piva, F., D'Antonio, M., D'Onorio De Meo, P., Paoletti, D., Castrignanò, T., D'Erchia, A.M., Picardi, E., Zambelli, F., Principato, G., *et al.* (2013). SpliceAid-F: a database of human splicing factors and their RNA-binding sites. Nucleic acids research *41*, D125-D131.

Goina, E., Skoko, N., and Pagani, F. (2008). Binding of DAZAP1 and hnRNPA1/A2 to an exonic splicing silencer in a natural BRCA1 exon 18 mutant. Molecular and cellular biology *28*, 3850-3860.

Gonzalez-Santos, J.M., Wang, A., Jones, J., Ushida, C., Liu, J., and Hu, J. (2002). Central Region of the Human Splicing Factor Hprp3p Interacts with Hprp4p. Journal of Biological Chemistry *277*, 23764-23772.

Gooding, C., Roberts, G.C., and Smith, C.W. (1998). Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. RNA (New York, NY *4*, 85-100.

Goyenvalle, A., Babbs, A., van Ommen, G.J., Garcia, L., and Davies, K.E. (2009). Enhanced exon-skipping induced by U7 snRNA carrying a splicing silencer sequence: Promising tool for DMD therapy. Molecular therapy : the journal of the American Society of Gene Therapy *17*, 1234-1240.

Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev *10*, 233-243.

Graveley, B.R., Hertel, K.J., and Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. Embo J *17*, 6747-6756.

Griffey, R.H., Monia, B.P., Cummins, L.L., Freier, S., Greig, M.J., Guinosso, C.J., Lesnik, E., Manalili, S.M., Mohan, V., Owens, S., *et al.* (1996). 2'-O-aminopropyl ribonucleotides: a zwitterionic modification that enhances the exonuclease resistance and biological activity of antisense oligonucleotides. Journal of medicinal chemistry *39*, 5100-5109.

Grosso, A.R., Martins, S., and Carmo-Fonseca, M. (2008). The emerging role of splicing factors in cancer. EMBO reports *9*, 1087-1093.

Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K., and Hutton, M. (1999). 5´ Splice Site Mutations in tau Associated with the Inherited Dementia FTDP-17 Affect a Stem-Loop Structure That Regulates Alternative Splicing of Exon 10. Journal of Biological Chemistry *274*, 15134-15143.

Gudmundsdottir, K., and Ashworth, A. (2006). The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. Oncogene *25*, 5864-5874.

Gunzl, A., Palfi, Z., and Bindereif, A. (2002). Analysis of RNA-protein complexes by oligonucleotide-targeted RNase H digestion. Methods *26*, 162-169.

Gutmann, D.H., Geist, R.T., Rose, K., and Wright, D.E. (1995). Expression of two new protein isoforms of the neurofibromatosis type 1 gene product, neurofibromin, in muscle tissues. Developmental dynamics : an official publication of the American Association of Anatomists *202*, 302-311.

Haber, J.E. (1998). The many interfaces of Mre11. Cell *95*, 583-586.
Hall, S.L., and Padgett, R.A. (1996). Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. Science *271*, 1716-1718.

Hanamura, A., Caceres, J.F., Mayeda, A., Franza, B.R., Jr., and Krainer, A.R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. RNA (New York, NY *4*, 430-444.

Hartwell, L.H., and Weinert, T.A. (1989). Checkpoints: controls that ensure the order of cell cycle events. Science *246*, 629-634.

Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., Ogata, H., and Ohta, T. (2001). The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. The Journal of biological chemistry *276*, 14537-14540.

Hicks, M.J., Mueller, W.F., Shepard, P.J., and Hertel, K.J. (2010). Competing upstream 5' splice sites enhance the rate of proximal splicing. Mol Cell Biol *30*, 1878-1886.

Honore, B., Rasmussen, H.H., Vorum, H., Dejgaard, K., Liu, X., Gromov, P., Madsen, P., Gesser, B., Tommerup, N., and Celis, J.E. (1995). Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. The Journal of biological chemistry *270*, 28780-28789.

Horowitz, D.S., and Krainer, A.R. (1994). Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. Trends Genet *10*, 100-106.

House, A.E., and Lynch, K.W. (2006). An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. Nat Struct Mol Biol *13*, 937-944.

Hua, Y., Sahashi, K., Hung, G., Rigo, F., Passini, M.A., Bennett, C.F., and Krainer, A.R. (2010). Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. Genes Dev *24*, 1634-1644.

Huber, L.J., Yang, T.W., Sarkisian, C.J., Master, S.R., Deng, C.X., and Chodosh, L.A. (2001). Impaired DNA damage response in cells expressing an exon 11-deleted murine Brca1 variant that localizes to nuclear foci. Mol Cell Biol *21*, 4005-4015.

Hurst, L.D., and Pál, C. (2001). Evidence for purifying selection acting on silent sites in BRCA1. Trends in Genetics *17*, 62-65.

Hutchins, J.R., and Clarke, P.R. (2004). Many fingers on the mitotic trigger: post-translational regulation of the Cdc25C phosphatase. Cell cycle *3*, 41-45.

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., *et al.* (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature *393*, 702-705.

Hwang, D.Y., and Cohen, J.B. (1996). U1 snRNA promotes the selection of nearby 5' splice sites by U6 snRNA in mammalian cells. Genes Dev *10*, 338-350.

I. Orban, T., and Olah, E. (2001). Purifying selection on silent sites – a constraint from splicing regulation? Trends in Genetics *17*, 252-253.

Izaurralde, E., and Mattaj, I.W. (1995). RNA export. Cell *81*, 153-159.

Jenkins, J.L., Agrawal, A.A., Gupta, A., Green, M.R., and Kielkopf, C.L. (2013). U2AF65 adapts to diverse pre-mRNA splice sites through conformational

selection of specific and promiscuous RNA recognition motifs. Nucleic acids research *41*, 3859-3873.

Jiang, C.C., Lai, F., Tay, K.H., Croft, A., Rizos, H., Becker, T.M., Yang, F., Liu, H., Thorne, R.F., Hersey, P., *et al.* (2010). Apoptosis of human melanoma cells induced by inhibition of B-RAFV600E involves preferential splicing of bimS. Cell death & disease *1*, e69.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol Cell *12*, 5-14.

Kashima, T., and Manley, J.L. (2003). A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. Nature genetics *34*, 460-463.

Kashima, T., Rao, N., David, C.J., and Manley, J.L. (2007). hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. Human molecular genetics *16*, 3149-3159.

Kass, E.M., Helgadottir, H.R., Chen, C.-C., Barbera, M., Wang, R., Westermark, U.K., Ludwig, T., Moynahan, M.E., and Jasin, M. (2013). Double-strand break repair by homologous recombination in primary mouse somatic cells requires BRCA1 but not the ATM kinase. Proceedings of the National Academy of Sciences *110*, 5564-5569.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. Gene *514*, 1-30.

Kielkopf, C.L., Lucke, S., and Green, M.R. (2004). U2AF homology motifs: protein recognition in the RRM world. Genes Dev *18*, 1513-1526.

Kim, S.S., Cao, L., Lim, S.C., Li, C., Wang, R.H., Xu, X., Bachelier, R., and Deng, C.X. (2006). Hyperplasia and spontaneous tumor development in the gynecologic system in mice lacking the BRCA1-Delta11 isoform. Mol Cell Biol *26*, 6983-6992.

Kinzler, K.W., and Vogelstein, B. (1997). Cancer-susceptibility genes. Gatekeepers and caretakers. Nature *386*, 761, 763.

Koed Doktor, T., Schroeder, L.D., Vested, A., Palmfeldt, J., Andersen, H.S., Gregersen, N., and Andresen, B.S. (2011). SMN2 exon 7 splicing is inhibited by binding of hnRNP A1 to a common ESS motif that spans the 3´ splice site. Human Mutation *32*, 220-230.

Konarska, M.M., Grabowski, P.J., Padgett, R.A., and Sharp, P.A. (1985). Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature *313*, 552-557.

Koonin, E.V., Altschul, S.F., and Bork, P. (1996). BRCA1 protein products ... Functional motifs. Nature genetics *13*, 266-268.

Korlimarla, A., Bhandary, L., Prabhu, J.S., Shankar, H., Sankaranarayanan, H., Kumar, P., Remacle, J., Natarajan, D., and Sridhar, T.S. (2013). Identification of a non-canonical nuclear localization signal (NLS) in BRCA1 that could mediate nuclear localization of splice variants lacking the classical NLS. Cellular & molecular biology letters *18*, 284-296.

Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J., and Nogues, G. (2004). Multiple links between transcription and splicing. RNA (New York, NY *10*, 1489-1498.

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

Kramer, A., Gruter, P., Groning, K., and Kastner, B. (1999). Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. The Journal of cell biology *145*, 1355-1368.

Kramer, A., and Utans, U. (1991). Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. EMBO J *10*, 1503-1509.

Kuo, P.H., Doudeva, L.G., Wang, Y.T., Shen, C.K., and Yuan, H.S. (2009). Structural insights into TDP-43 in nucleic-acid binding and domain interactions. Nucleic acids research *37*, 1799-1808.

Ladd, A.N., and Cooper, T.A. (2002). Finding signals that regulate alternative splicing in the post-genomic era. Genome Biol *3*, reviews0008.

Langford, C.J., Klinz, F.J., Donath, C., and Gallwitz, D. (1984). Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. Cell *36*, 645-653.

Lavigueur, A., La Branche, H., Kornblihtt, A.R., and Chabot, B. (1993). A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev *7*, 2405-2417.

Li, X., and Heyer, W.D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. Cell research *18*, 99-113.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem *79*, 181-211.

Lim, L.P., and Burge, C.B. (2001). A computational analysis of sequence features involved in recognition of short introns. Proceedings of the National Academy of Sciences of the United States of America *98*, 11193-11198.

Liu, H.X., Cartegni, L., Zhang, M.Q., and Krainer, A.R. (2001). A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nature genetics *27*, 55-58.

Liu, H.X., Chew, S.L., Cartegni, L., Zhang, M.Q., and Krainer, A.R. (2000). Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. Mol Cell Biol *20*, 1063-1071.

Liu, H.X., Zhang, M., and Krainer, A.R. (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev *12*, 1998-2012.

Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.J. (1951). Protein measurement with the Folin phenol reagent. The Journal of biological chemistry *193*, 265-275.

Lu, M., Conzen, S.D., Cole, C.N., and Arrick, B.A. (1996). Characterization of functional messenger RNA splice variants of BRCA1 expressed in nonmalignant and tumor-derived breast cells. Cancer research *56*, 4578-4581.

Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P., and Luhrmann, R. (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. EMBO J *21*, 1148-1157.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499-506.

Maniatis, T., and Tasic, B. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature *418*, 236-243.

Maniccia, A.W., Lewis, C., Begum, N., Xu, J., Cui, J., Chipitsyna, G., Aysola, K., Reddy, V., Bhat, G., Fujimura, Y., *et al.* (2009). Mitochondrial localization, ELK-1 transcriptional regulation and growth inhibitory functions of BRCA1, BRCA1a, and BRCA1b proteins. Journal of cellular physiology *219*, 634-641.

Manley, J.L., and Krainer, A.R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). Genes & Development *24*, 1073-1074.

Mann, C.J., Honeyman, K., Cheng, A.J., Ly, T., Lloyd, F., Fletcher, S., Morgan, J.E., Partridge, T.A., and Wilton, S.D. (2001). Antisense-induced exon skipping

and synthesis of dystrophin in the mdx mouse. Proceedings of the National Academy of Sciences of the United States of America *98*, 42-47.

Manoharan, M. (1999). 2'-carbohydrate modifications in antisense oligonucleotide therapy: importance of conformation, configuration and conjugation. Biochimica et biophysica acta *1489*, 117-130.

Maroney, P.A., Romfo, C.M., and Nilsen, T.W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. Mol Cell *6*, 317-328.

Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J.F., Revil, T., and Chabot, B. (2007). hnRNP proteins and splicing control. Adv Exp Med Biol *623*, 123-147.

Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol *6*, 386-398.

Mayeda, A., Screaton, G.R., Chandler, S.D., Fu, X.-D., and Krainer, A.R. (1999). Substrate Specificities of SR Proteins in Constitutive Splicing Are Determined by Their RNA Recognition Motifs and Composite Pre-mRNA Exonic Elements. Molecular and Cellular Biology *19*, 1853-1863.

McEachern, K.A., Archey, W.B., Douville, K., and Arrick, B.A. (2003). BRCA1 splice variants exhibit overlapping and distinct transcriptional transactivation activities. Journal of cellular biochemistry *89*, 120-132.

McKie, A.B., McHale, J.C., Keen, T.J., Tarttelin, E.E., Goliath, R., van Lith-Verhoeven, J.J.C., Greenberg, J., Ramesar, R.S., Hoyng, C.B., Cremers, F.P.M., *et al.* (2001). Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). Human molecular genetics *10*, 1555-1562.

Mercado, P.A., Ayala, Y.M., Romano, M., Buratti, E., and Baralle, F.E. (2005). Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene. Nucleic acids research *33*, 6000-6010.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., *et al.* (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science *266*, 66-71.

Millevoi, S., Bernat, S., Telly, D., Fouque, F., Gladieff, L., Favre, G., Vagner, S., and Toulas, C. (2010). The c.5242C>A BRCA1 missense variant induces exon skipping by increasing splicing repressors binding. Breast cancer research and treatment *120*, 391-399.

Minn, A.J., Boise, L.H., and Thompson, C.B. (1996). Bcl-x Antagonizes the Protective Effects of Bcl-x. Journal of Biological Chemistry *271*, 6306-6312.

Misteli, T., and Spector, D.L. (1997). Protein phosphorylation and the nuclear organization of pre-mRNA splicing. Trends Cell Biol *7*, 135-138.

Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., and Kunkel, L.M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics *2*, 90-95.

Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364-368.

Moulton, H.M., and Moulton, J.D. (2010). Morpholinos and their peptide conjugates: Therapeutic promise and challenge for Duchenne muscular dystrophy. Biochimica et Biophysica Acta (BBA) - Biomembranes *1798*, 2296-2303.

Mullan, P.B., Quinn, J.E., and Harkin, D.P. (2006). The role of BRCA1 in transcriptional regulation and cell cycle control. Oncogene *25*, 5854-5863.

Mullen, M.P., Smith, C.W., Patton, J.G., and Nadal-Ginard, B. (1991). Alpha-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice. Genes Dev *5*, 642-655.

Muntoni, F., and Wood, M.J. (2011). Targeting RNA to treat neuromuscular disease. Nature reviews Drug discovery *10*, 621-637.

Muslin, A.J., and Xing, H. (2000). 14-3-3 proteins: regulation of subcellular localization by molecular interference. Cellular signalling *12*, 703-709.

Neuhausen, S.L., and Marshall, C.J. (1994). Loss of heterozygosity in familial tumors from three BRCA1-linked kindreds. Cancer research *54*, 6069-6072.
Newman, A.J. (1997). The role of U5 snRNP in pre-mRNA splicing. EMBO J *16*, 5797-5800.

O'Donovan, P.J., and Livingston, D.M. (2010). BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. Carcinogenesis *31*, 961-967.

Orban, T.I., and Olah, E. (2001). Expression profiles of BRCA1 splice variants in asynchronous and in G1/S synchronized tumor cell lines. Biochem Biophys Res Commun *280*, 32-38.

Orban, T.I., and Olah, E. (2003). Emerging roles of BRCA1 alternative splicing. Molecular pathology : MP *56*, 191-197.

Ostareck-Lederer, A., Ostareck, D.H., and Hentze, M.W. (1998). Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2. Trends Biochem Sci *23*, 409-411.

Ou, S.H., Wu, F., Harrich, D., García-Martínez, L.F., and Gaynor, R.B. (1995). Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. Journal of virology *69*, 3584-3596.

Owen, N., Zhou, H., Malygin, A.A., Sangha, J., Smith, L.D., Muntoni, F., and Eperon, I.C. (2011a). Design principles for bifunctional targeted oligonucleotide enhancers of splicing. Nucleic acids research *39*, 7194-7208.

Owen, N., Zhou, H., Malygin, A.A., Sangha, J., Smith, L.D., Muntoni, F., and Eperon, I.C. (2011b). Design principles for bifunctional targeted oligonucleotide enhancers of splicing. Nucleic acids research *39*, 7194-7208.

Pagani, F., and Baralle, F.E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet *5*, 389-396.

Pagani, F., Buratti, E., Stuani, C., and Baralle, F.E. (2003a). Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. The Journal of biological chemistry *278*, 26580-26588.

Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., Casals, T., and Baralle, F.E. (2003b). New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Human molecular genetics *12*, 1111-1120.

Park, E., Han, J., Son, G.H., Lee, M.S., Chung, S., Park, S.H., Park, K., Lee, K.H., Choi, S., Seong, J.Y., *et al.* (2006). Cooperative actions of Tra2alpha with 9G8 and SRp30c in the RNA splicing of the gonadotropin-releasing hormone gene transcript. The Journal of biological chemistry *281*, 401-409.

Park, V.M., Kenwright, K.A., Sturtevant, D.B., and Pivnick, E.K. (1998). Alternative splicing of exons 29 and 30 in the neurofibromatosis type 1 gene. Hum Genet *103*, 382-385.

Passini, M.A., Bu, J., Richards, A.M., Kinnecom, C., Sardi, S.P., Stanek, L.M., Hua, Y., Rigo, F., Matson, J., Hung, G., *et al.* (2011). Antisense oligonucleotides delivered to the mouse CNS ameliorate symptoms of severe spinal muscular atrophy. Science translational medicine *3*, 72ra18.

Passoni, M., De Conti, L., Baralle, M., and Buratti, E. (2012). UG repeats/TDP-43 interactions near 5' splice sites exert unpredictable effects on splicing modulation. Journal of molecular biology *415*, 46-60.

Pavlicek, A., Noskov, V.N., Kouprina, N., Barrett, J.C., Jurka, J., and Larionov, V. (2004). Evolution of the tumor suppressor BRCA1 locus in primates: implications for cancer predisposition. Human molecular genetics *13*, 2737-2751.

Petersen, M., Nielsen, C.B., Nielsen, K.E., Jensen, G.A., Bondensgaard, K., Singh, S.K., Rajwanshi, V.K., Koshkin, A.A., Dahl, B.M., Wengel, J., *et al.* (2000). The conformations of locked nucleic acids (LNA). Journal of molecular recognition : JMR *13*, 44-53.

Pinol-Roma, S., and Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. Nature *355*, 730-732.

Piva, F., Giulietti, M., Burini, A.B., and Principato, G. (2012). SpliceAid 2: A database of human splicing factors expression data and RNA target motifs. Human Mutation *33*, 81-85.

Piva, F., Giulietti, M., Nocchi, L., and Principato, G. (2009). SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. Bioinformatics *25*, 1211-1213.

Pollard, A.J., Krainer, A.R., Robson, S.C., and Europe-Finner, G.N. (2002). Alternative Splicing of the Adenylyl Cyclase Stimulatory G-protein Gαs Is Regulated by SF2/ASF and Heterogeneous Nuclear Ribonucleoprotein A1 (hnRNPA1) and Involves the Use of an Unusual TG 3´-Splice Site. Journal of Biological Chemistry *277*, 15241-15251.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501-512.

Qin, Y., Xu, J., Aysola, K., Begum, N., Reddy, V., Chai, Y., Grizzle, W.E., Partridge, E.E., Reddy, E.S., and Rao, V.N. (2011). Ubc9 mediates nuclear localization and growth suppression of BRCA1 and BRCA1a proteins. Journal of cellular physiology *226*, 3355-3367.

Quesada, V., Conde, L., Villamor, N., Ordonez, G.R., Jares, P., Bassaganyas, L., Ramsay, A.J., Bea, S., Pinyol, M., Martinez-Trillos, A., *et al.* (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nature genetics *44*, 47-52.

Raker, V.A., Plessel, G., and Luhrmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. Embo J *15*, 2256-2269.

Raponi, M., Buratti, E., Dassie, E., Upadhyaya, M., and Baralle, D. (2009). Low U1 snRNP dependence at the NF1 exon 29 donor splice site. The FEBS journal *276*, 2060-2073.

Raponi, M., Douglas, A.G., Tammaro, C., Wilson, D.I., and Baralle, D. (2012). Evolutionary constraint helps unmask a splicing regulatory region in BRCA1 exon 11. PloS one *7*, e37255.

Raponi, M., Kralovicova, J., Copson, E., Divina, P., Eccles, D., Johnson, P., Baralle, D., and Vorechovsky, I. (2011). Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. Hum Mutat *32*, 436-444.

Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231-1245.

Raynard, S., Niu, H., and Sung, P. (2008). DNA double-strand break processing: the beginning of the end. Genes & Development *22*, 2903-2907.

Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. Genes Dev *3*, 2113-2123.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. Curr Opin Cell Biol *12*, 340-345.

Reed, R., and Maniatis, T. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. Cell *41*, 95-105.

Reed, R., and Maniatis, T. (1988). The role of the mammalian branchpoint sequence in pre-mRNA splicing. Genes Dev *2*, 1268-1276.

Roca, X., and Krainer, A.R. (2009). Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. Nat Struct Mol Biol *16*, 176-182.

Roscigno, R.F., Weiner, M., and Garcia-Blanco, M.A. (1993). A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. The Journal of biological chemistry *268*, 11222-11229.

Ruffner, H., and Verma, I.M. (1997). BRCA1 is a cell cycle-regulated nuclear phosphoprotein. Proceedings of the National Academy of Sciences *94*, 7138-7143.

Ruskin, B., Greene, J.M., and Green, M.R. (1985). Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. Cell *41*, 833-844.

Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317-331.

Sander, B., Golas, M.M., Makarov, E.M., Brahms, H., Kastner, B., Lührmann, R., and Stark, H. (2006). Organization of Core Spliceosomal Components U5 snRNA Loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as Revealed by Electron Cryomicroscopy. Molecular Cell *24*, 267-278.

Sanford, J.R., Ellis, J., and Caceres, J.F. (2005). Multiple roles of arginine/serine-rich splicing factors in RNA processing. Biochem Soc Trans *33*, 443-446.

Sauliere, J., Sureau, A., Expert-Bezancon, A., and Marie, J. (2006). The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. Mol Cell Biol *26*, 8755-8769.

Schaal, T.D., and Maniatis, T. (1999). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol *19*, 261-273.

Scott, A., Petrykowska, H.M., Hefferon, T., Gotea, V., and Elnitski, L. (2012). Functional analysis of synonymous substitutions predicted to affect splicing of the CFTR gene. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society *11*, 511-517.

Senapathy, P. (1988). Possible evolution of splice-junction signals in eukaryotic genes from stop codons. Proceedings of the National Academy of Sciences of the United States of America *85*, 1129-1133.

Sephton, C.F., Cenik, C., Kucukural, A., Dammer, E.B., Cenik, B., Han, Y., Dewey, C.M., Roth, F.P., Herz, J., Peng, J., *et al.* (2011). Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. The Journal of biological chemistry *286*, 1204-1215.

Shen, H., and Green, M.R. (2004). A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. Mol Cell *16*, 363-373.

Shen, H., Kan, J.L., and Green, M.R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell *13*, 367-376.

Shepard, P.J., and Hertel, K.J. (2009). The SR protein family. Genome Biol *10*, 242.

Shimoni-Sebag, A., Lebenthal-Loinger, I., Zender, L., and Karni, R. (2013). RRM1 domain of the splicing oncoprotein SRSF1 is required for MEK1-MAPK-ERK activation and cellular transformation. Carcinogenesis.

Shodai, A., Morimura, T., Ido, A., Uchida, T., Ayaki, T., Takahashi, R., Kitazawa, S., Suzuki, S., Shirouzu, M., Kigawa, T., *et al.* (2013). Aberrant Assembly of RNA Recognition Motif 1 Links to Pathogenic Conversion of TAR DNA-binding Protein of 43 kDa (TDP-43). Journal of Biological Chemistry *288*, 14886-14905.

Siliciano, P.G., and Guthrie, C. (1988). 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes Dev *2*, 1258-1267.

Simard, M.J., and Chabot, B. (2002). SRp30c Is a Repressor of 3´ Splice Site Utilization. Molecular and Cellular Biology *22*, 4001-4010.

Singh, N.N., Androphy, E.J., and Singh, R.N. (2004). An extended inhibitory context causes skipping of exon 7 of SMN2 in spinal muscular atrophy. Biochemical and Biophysical Research Communications *315*, 381-388.

Skordis, L.A., Dunckley, M.G., Yue, B., Eperon, I.C., and Muntoni, F. (2003a). Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. Proceedings of the National Academy of Sciences of the United States of America *100*, 4114-4119.

Skordis, L.A., Dunckley, M.G., Yue, B., Eperon, I.C., and Muntoni, F. (2003b). Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. Proceedings of the National Academy of Sciences *100*, 4114-4119.

Smith, C.W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci *25*, 381-388.

Soret, J., Gabut, M., and Tazi, J. (2006). SR proteins as potential targets for therapy. Progress in molecular and subcellular biology *44*, 65-87.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315-326.

Stamm, S. (2008). Regulation of alternative splicing by reversible protein phosphorylation. The Journal of biological chemistry *283*, 1223-1227.

Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature *409*, 539-542.

Strong, M.J., Volkening, K., Hammond, R., Yang, W., Strong, W., Leystra-Lantz, C., and Shoesmith, C. (2007). TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. Molecular and Cellular Neuroscience *35*, 320-327.

Sy, S.M.H., Huen, M.S.Y., and Chen, J. (2009). PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proceedings of the National Academy of Sciences *106*, 7155-7160.

Tacke, R., Chen, Y., and Manley, J.L. (1997). Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proceedings of the National Academy of Sciences of the United States of America *94*, 1148-1153.

Talukdar, I., Sen, S., Urbano, R., Thompson, J., Yates, J.R., 3rd, and Webster, N.J. (2011). hnRNP A1 and hnRNP F modulate the alternative splicing of exon 11 of the insulin receptor gene. PloS one *6*, e27869.

Tammaro, C., Raponi, M., Wilson, D.I., and Baralle, D. (2012). BRCA1 exon 11 alternative splicing, multiple functions and the association with cancer. Biochem Soc Trans *40*, 768-772.

Tarn, W.Y., and Steitz, J.A. (1994). SR proteins can compensate for the loss of U1 snRNP functions in vitro. Genes Dev *8*, 2704-2717.

Taylor, J.K., Zhang, Q.Q., Wyatt, J.R., and Dean, N.M. (1999). Induction of endogenous Bcl-xS through the control of Bcl-x pre-mRNA splicing by antisense oligonucleotides. Nature biotechnology *17*, 1097-1100.

Teigelkamp, S., Achsel, T., Mundt, C., Gothel, S.F., Cronshagen, U., Lane, W.S., Marahiel, M., and Luhrmann, R. (1998). The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. RNA (New York, NY *4*, 127-141.

Teplova, M., Wallace, S.T., Tereshko, V., Minasov, G., Symons, A.M., Cook, P.D., Manoharan, M., and Egli, M. (1999). Structural origins of the exonuclease resistance of a zwitterionic RNA. Proceedings of the National Academy of Sciences of the United States of America *96*, 14240-14245.

Terns, M.P., and Terns, R.M. (2001). Macromolecular complexes: SMN — the master assembler. Current Biology *11*, R862-R864.

Tian, H., and Kole, R. (1995). Selection of novel exon recognition elements from a pool of random sequences. Mol Cell Biol *15*, 6291-6298.

Tollervey, J.R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., Konig, J., Hortobagyi, T., Nishimura, A.L., Zupunski, V., *et al.* (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nature neuroscience *14*, 452-458.

Tosi, M., Stamm, S., and Baralle, D. (2010). RNA splicing meets genetic testing: detection and interpretation of splicing defects in genetic diseases. European journal of human genetics : EJHG *18*, 737-738.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.

Turner, I.A., Norman, C.M., Churcher, M.J., and Newman, A.J. (2004). Roles of the U5 snRNP in spliceosome dynamics and catalysis. Biochem Soc Trans *32*, 928-931.

Turunen, J.J., Niemelä, E.H., Verma, B., and Frilander, M.J. (2013). The significant other: splicing by the minor spliceosome. Wiley Interdisciplinary Reviews: RNA *4*, 61-76.

Twyffels, L., Gueydan, C., and Kruys, V. (2011). Shuttling SR proteins: more than splicing factors. FEBS Journal *278*, 3246-3255.

Umen, J.G., and Guthrie, C. (1995). The second catalytic step of pre-mRNA splicing. RNA (New York, NY *1*, 869-885.

Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996a). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. Science *273*, 1706-1709.

Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996b). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected] [published erratum appears in Science 1996 Oct 4;274(5284):21]. Science *273*, 1706-1709.

Valcarcel, J., and Green, M.R. (1996). The SR protein family: pleiotropic functions in pre-mRNA splicing. Trends Biochem Sci *21*, 296-301.

van Deutekom, J.C., Janson, A.A., Ginjaar, I.B., Frankhuizen, W.S., Aartsma-Rus, A., Bremmer-Bout, M., den Dunnen, J.T., Koop, K., van der Kooi, A.J., Goemans, N.M., *et al.* (2007). Local Dystrophin Restoration with Antisense Oligonucleotide PRO051. New England Journal of Medicine *357*, 2677-2686.

Vaughn, J., Davis, P., Jarboe, M., Huper, G., Evans, A., Wiseman, R., Berchuck, A., Iglehart, J., Futreal, P., and Marks, J. (1996). BRCA1 expression is induced

before DNA synthesis in both normal and tumor-derived breast cells. Cell Growth Differ *7*, 711-715.

Venkitaraman, A.R. (2002). Cancer susceptibility and the functions of BRCA1 and BRCA2. Cell *108*, 171-182.

Vibe-Pedersen, K., Kornblihtt, A.R., and Baralle, F.E. (1984). Expression of a human alpha-globin/fibronectin gene hybrid generates two mRNAs by alternative splicing. EMBO J *3*, 2511-2516.

Vithana, E.N., Abu-Safieh, L., Allen, M.J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N.D., Willis, C., Moore, A.T., *et al.* (2001). A Human Homolog of Yeast Pre-mRNA Splicing Gene, PRP31, Underlies Autosomal Dominant Retinitis Pigmentosa on Chromosome 19q13.4 (RP11). Molecular Cell *8*, 375-381.

Wagner, E.J., and Garcia-Blanco, M.A. (2001). Polypyrimidine tract binding protein antagonizes exon definition. Mol Cell Biol *21*, 3281-3288.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wan, Y., and Wu, C.J. (2013). SF3B1 mutations in chronic lymphocytic leukemia. Blood *121*, 4627-4634.

Wang, J., Pegoraro, E., Menegazzo, E., Gennarelli, M., Hoop, R.C., Angelini, C., and Hoffman, E.P. (1995). Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. Human molecular genetics *4*, 599-606.

Wang, X.W., Zhan, Q., Coursen, J.D., Khan, M.A., Kontny, H.U., Yu, L., Hollander, M.C., O'Connor, P.M., Fornace, A.J., Jr., and Harris, C.C. (1999). GADD45 induction of a G2/M cell cycle checkpoint. Proceedings of the National Academy of Sciences of the United States of America *96*, 3706-3711.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S.J., and Qin, J. (2000). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev *14*, 927-939.

West, S.C., CHAPPELL, C., HANAKAHI, L.A., MASSON, J.-Y., MCILWRAITH, M.J., and VAN DYCK, E. (2000). Double-strand Break Repair in Human Cells. Cold Spring Harbor Symposia on Quantitative Biology *65*, 315-322.

Will, C.L., and Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol *13*, 290-301.

Will, C.L., Rumpler, S., Klein Gunnewiek, J., van Venrooij, W.J., and Luhrmann, R. (1996). In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. Nucleic acids research *24*, 4614-4623.

Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Luhrmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. Embo J *21*, 4978-4988.

Williams, J.H., Schray, R.C., Patterson, C.A., Ayitey, S.O., Tallent, M.K., and Lutz, G.J. (2009). Oligonucleotide-mediated survival of motor neuron protein expression in CNS improves phenotype in a mouse model of spinal muscular atrophy. The Journal of neuroscience : the official journal of the Society for Neuroscience *29*, 7633-7638.

Winton, M.J., Igaz, L.M., Wong, M.M., Kwong, L.K., Trojanowski, J.Q., and Lee, V.M. (2008). Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. The Journal of biological chemistry *283*, 13302-13309.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061-1070.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832-835.

Wyatt, J.R., Sontheimer, E.J., and Steitz, J.A. (1992). Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. Genes & Development *6*, 2542-2553.

Xu, C.F., Chambers, J.A., Nicolai, H., Brown, M.A., Hujeirat, Y., Mohammed, S., Hodgson, S., Kelsell, D.P., Spurr, N.K., Bishop, D.T., *et al.* (1997). Mutations and alternative splicing of the BRCA1 gene in UK breast/ovarian cancer families. Genes, chromosomes & cancer *18*, 102-110.

Yarden, R.I., and Brody, L.C. (1999). BRCA1 interacts with components of the histone deacetylase complex. Proceedings of the National Academy of Sciences *96*, 4983-4988.

Yin, X., Jin, N., Gu, J., Shi, J., Zhou, J., Gong, C.X., Iqbal, K., Grundke-Iqbal, I., and Liu, F. (2012). Dual-specificity tyrosine phosphorylation-regulated kinase 1A (Dyrk1A) modulates serine/arginine-rich protein 55 (SRp55)-promoted Tau exon 10 inclusion. The Journal of biological chemistry *287*, 30497-30506.

Young, P.J., DiDonato, C.J., Hu, D., Kothary, R., Androphy, E.J., and Lorson, C.L. (2002). SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2β1. Human molecular genetics *11*, 577-587.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

Zerbe, L.K., Pino, I., Pio, R., Cosper, P.F., Dwyer-Nield, L.D., Meyer, A.M., Port, J.D., Montuenga, L.M., and Malkinson, A.M. (2004). Relative amounts of antagonistic splicing factors, hnRNP A1 and ASF/SF2, change during neoplastic lung growth: implications for pre-mRNA processing. Molecular carcinogenesis *41*, 187-196.

Zhang, J., and Powell, S.N. (2005). The role of the BRCA1 tumor suppressor in DNA double-strand break repair. Molecular cancer research : MCR *3*, 531-539.

Zhang, M.Q. (1998). Statistical Features of Human Exons and Their Flanking Regions. Human molecular genetics *7*, 919-932.

Zhang, X.H.-F., Leslie, C.S., and Chasin, L.A. (2005). Dichotomous splicing signals in exon flanks. Genome Research *15*, 768-779.

Zhang, Y., Madl, T., Bagdiul, I., Kern, T., Kang, H.-S., Zou, P., Mäusbacher, N., Sieber, S.A., Krämer, A., and Sattler, M. (2013). Structure, phosphorylation and U2AF65 binding of the N-terminal domain of splicing factor 1 during 3 ´ -splice site recognition. Nucleic acids research *41*, 1343-1354.

Zheng, Z.M., He, P.J., and Baker, C.C. (1999). Function of a bovine papillomavirus type 1 exonic splicing suppressor requires a suboptimal upstream 3' splice site. Journal of virology *73*, 29-36.

Zhong, X.Y., Wang, P., Han, J., Rosenfeld, M.G., and Fu, X.D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. Mol Cell *35*, 1-10.

Zhou, H.-L., Hinman, M.N., Barron, V.A., Geng, C., Zhou, G., Luo, G., Siegel, R.E., and Lou, H. (2011). Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. Proceedings of the National Academy of Sciences *108*, E627-E635.

Zhou, Z., and Fu, X.-D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma *122*, 191-207.

Zhu, J., Mayeda, A., and Krainer, A.R. (2001). Exon Identity Established through Differential Antagonism between Exonic Splicing Silencer-Bound hnRNP A1 and Enhancer-Bound SR Proteins. Molecular Cell *8*, 1351-1361.

Zhuang, Y., and Weiner, A.M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell *46*, 827-835.

Zhuang, Y.A., Goldstein, A.M., and Weiner, A.M. (1989). UACUAAC is the preferred branch site for mammalian mRNA splicing. Proceedings of the National Academy of Sciences *86*, 2752-2756.

OPEN ACCESS Freely available online

PLoS one

Evolutionary Constraint Helps Unmask a Splicing Regulatory Region in BRCA1 Exon 11

Michela Raponi[1☯], Andrew G. L. Douglas[1,2☯], Claudia Tammaro[1], David I. Wilson[1], Diana Baralle[1,2]*

1 Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom, **2** Wessex Clinical Genetics Service, Southampton University Hospitals NHS Trust, Southampton, United Kingdom

Abstract

Background: Alternative splicing across exon 11 produces several *BRCA1* isoforms. Their proportion varies during the cell cycle, between tissues and in cancer suggesting functional importance of *BRCA1* splicing regulation around this exon. Although the regulatory elements driving exon 11 splicing have never been identified, a selective constraint against synonymous substitutions (silent nucleotide variations that do not alter the amino acid residue sequence) in a critical region of *BRCA1* exon 11 has been reported to be associated with the necessity to maintain regulatory sequences.

Methodology/Principal Findings: Here we have designed a specific minigene to investigate the possibility that this bias in synonymous codon usage reflects the need to preserve the *BRCA1* alternative splicing program. We report that in-frame deletions and translationally silent nucleotide substitutions in the critical region affect splicing regulation of *BRCA1* exon 11.

Conclusions/Significance: Using a hybrid minigene approach, we have experimentally validated the hypothesis that the need to maintain correct alternative splicing is a selective pressure against translationally silent sequence variations in the critical region of *BRCA1* exon 11. Identification of the *trans*-acting factors involved in regulating exon 11 alternative splicing will be important in understanding *BRCA1*-associated tumorigenesis.

Citation: Raponi M, Douglas AGL, Tammaro C, Wilson DI, Baralle D (2012) Evolutionary Constraint Helps Unmask a Splicing Regulatory Region in BRCA1 Exon 11. PLoS ONE 7(5): e37255. doi:10.1371/journal.pone.0037255

Editor: Ludmila Prokunina-Olsson, National Cancer Institute, National Institutes of Health, United States of America

Received January 16, 2012; **Accepted** April 18, 2012; **Published** May 16, 2012

Funding: This work has been supported by EURASNET (http://www.eurasnet.info/) and Cancer research UK (19649) (http://www.cancerresearchuk.org/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: D.Baralle@soton.ac.uk

☯ These authors contributed equally to this work.

Introduction

Pathogenic mutations in the *BRCA1* gene are associated with a high risk of breast and ovarian cancer. Women heterozygous for such mutations have a lifetime risk of up to around 80% of developing breast cancer and up to 40% of developing ovarian cancer [1]. The effects of common deleterious mutations, such as exonic insertions and deletions, nonsense substitutions and substitutions at invariant consensus splice sites (AG and GT), are relatively easy to predict. However, the effects of synonymous, translationally silent substitutions on splicing are generally much less well understood and require investigation through functional studies. Such substitutions are called 'silent' as they do not directly change the amino acid sequence of the protein. However, these synonymous substitutions (as well as non synonymous substitutions) can still have a deleterious effect at the RNA level by creating or disrupting secondary structures or regulatory sequences. This in turn may alter splicing fidelity, with consequent loss of function or production of new antagonistic protein isoforms.

BRCA1 is known to undergo alternative splicing of a number of its exons, including the large and functionally important exon 11 [2]. Alternative splicing of exon 11 yields a full length isoform (FL) and also shorter isoforms through use of an alternative intra-exonic splice donor site, D(11q), or through complete skipping of exon 11, D(11). Alternative splicing can also exclude exons 9 and 10 from the mature mRNA, D(9,10). These isoforms maintain the original *BRCA1* open reading frame and allow functional protein production.

The control of the ratio of splicing isoforms produced within a cell requires regulation by *trans*-acting splicing factors that recognise and bind specific pre-mRNA sequences. Their expression changes in cancer as well as in different tissues [3,4]. Since *BRCA1* isoforms vary in quantity during the cell cycle and within different tissues [5–8], including tumour tissue, it is expected that important splicing regulatory elements may be found at critical gene regions (e.g. within exon 11) that allow the binding of relevant splicing factors. Mutations in these sequence elements would disrupt normal splicing and potentially lead to disease. Therefore there would be an evolutionary selection pressure not only against codon-altering mutations but also against synonymous mutations at these sites. This process, known as purifying selection, is thought to explain the bias in synonymous codon usage observed at specific genomic sites across evolutionarily divergent species. The observation of purifying selection at translationally silent sites may therefore reflect the presence of splicing regulatory elements or regions of critical RNA secondary structure [9–10].

A bias towards the usage of particular codons, rather than their synonymous counterparts, has previously been reported in a

critical region of *BRCA1* spanning the 3′ end of exon 10 and the 5′ end of exon 11 [11]. Several 'hotspots' (codons 195, 196, 215, 231, 244 and 313) have been identified around this critical region [11]. These are putative sites of purifying selection with a high ratio of nonsynonymous to synonymous substitutions.

In order to investigate *BRCA1* splicing involving exon 11, we developed a minigene incorporating the whole of *BRCA1* exons 8 to 11 and part of exon 12. Using this hybrid minigene approach, we have experimentally validated the hypothesis that maintenance of correct alternative splicing is the cause of selection against silent sequence variations in the *BRCA1* critical region spanning exon 11.

Results

Minigene splicing assay for BRCA1 exon 11

In order to assay the effects of synonymous substitutions on *BRCA1* splicing, we constructed a pB1 wild-type minigene that incorporated sequence from *BRCA1* exon 8 up to the first 89 nucleotides of exon 12 (Figure 1a). Intronic sequences were shortened but still contained at least 190 nucleotides of native intronic sequence at either end. Intron 8 is 460 bp (original size 2485 bp); Intron 9 is 379 bp (original size 1321 bp); Intron 10 is 684 bp (original size 985 bp); Intron 11 is 402 bp (original size 402 bp). The pB1 minigene was transiently transfected into MCF7 breast cancer cell lines. RNA was extracted and minigene-specific cDNA synthesised using the specific primer pCSrev. *BRCA1* splicing products were analysed by RT-PCR using primers specific for *BRCA1* FL, D(11) and D(11q) isoforms (Figure 1b). Electrophoresis of RT-PCR *BRCA1* splicing products revealed the presence of the three isoforms FL, D(11) and D(11q); showing comparable outcomes with that of endogenous BRCA1 in non-transfected MCF7 cells (Figure 1c). In order to validate the minigene for splicing assays, we introduced the nucleotide substitution c.696A>G previously reported to affect exon 11 splicing. As shown in Figure 1c, introduction of this change to the minigene caused an increase of the D(11) isoform as described in the literature [12].

The c.696A>G mutated minigene was also tested in normal mammary epithelial cells (HMEpC, invitrogen) and breast cancer cells MDA-MB-231 (ATCC). All cell lines tested resembled the increase in D11 isoform observed in MCF7 cells (data not shown).

Furthermore, a minigene construct containing the genomic sequence from exons 9–12 had the same pattern of splicing Another version of the minigene behaved in exactly the same manner when we inserted the entire genomic sequence from exons 9 to 12 (data not shown).

Synonymous substitutions of codons thought to undergo purifying selection

Mutations at codons 195, 196, 215, 231, 244 and 313 (which are proposed to exhibit purifying selection) were assayed for their effect on *BRCA1* splicing. 13 synonymous substitutions were introduced by site-directed mutagenesis (Supporting Table S1):

The resulting mutated minigenes were transiently transfected into MCF7 cell lines and splicing analysed as described (Figure 2). Mutations that do not affect splicing would be expected to give a similar isoform band pattern to the wild-type sequence. If splicing of exon 11 is affected, a relative decrease of one isoform should be accompanied by a reciprocal increase in one or more of the other isoforms, or the appearance of a novel isoform. In addition one or more new isoforms could appear.

All three synonymous substitutions at codon 231 appear to alter splicing. These changes lie 23 nucleotides distal to the start of exon 11 and significantly reduce the levels of FL and D(11q) in favour of increased levels of D(11) isoform. This suggests that sequence changes at this site alter the splicing of exon 11, through reduced recognition of the intron 10 acceptor site, resulting in its exclusion from the final mRNA. This finding points towards an important sequence element, such as an exonic splicing enhancer, located around codon 231.

Codon 244 lies 60 nucleotides downstream from the start of exon 11 and is 55 nucleotides proximal to the 11q splice donor site. The T>C synonymous substitution at the third position of this codon seems to enhance levels of the D(11q) isoform. A similar effect was also observed in normal mammary epithelial cells (HMEpC supplied by ECACC) (data not shown).

Codon 313 is situated 150 nucleotides 3′ (downstream) of the 11q splice donor site. Synonymous mutations in this codon do not appear to alter exon 11 splicing. Codons 195 and 196 are 9 and 6 nucleotides away from the 3′ end of exon 9 respectively and synonymous mutations at these sites would therefore most likely affect splicing of exon 9. Codon 215 lies 23 nucleotides from the 3′ end of exon 10 and so any splicing effects may be expected to affect this exon. Our assay does not show any significant alteration in relative exon 11 splice isoform abundance with synonymous mutations at these sites. We therefore evaluated their effect on skipping of exon 9 and 10. Splicing products revealed presence of four isoforms {FL, D(9), D(10) and D(9&10)} showing comparable outcomes with that of the pB1WT minigene (supporting figure S1).

Other synonymous changes reported in breast cancer patients

Since the c.693G>A substitution at codon 231 affects splicing of *BRCA1* exon 11 and has previously been reported in a patient with breast cancer [12], we decided to evaluate the effects of additional synonymous substitutions in *BRCA1* exon 11 (c.825C>T codon 275, c.828A>G codon 276 and c.795T>C codon 265; supporting table S1) that had been identified in patients with breast cancer ascertained by the local regional genetics service in whom no other pathogenic mutation had been found. After introducing the substitutions into the minigene, MCF7 cell lines were transfected and RNA analysed for FL, D(11q) and D(11) isoforms.

The T>C substitution within codon 265 lies just 8 nucleotides 3′ of the 11q splice donor site. Alteration of the native sequence GTAGTTCT to GTAGTTCC results in reduction of the D(11q) isoform (Figure 3 and supporting table S1). This suggests that the sequence at codon 265 influences the use of this alternative splice site. The proximity to the splice donor site makes it likely that the sequence forms part of the splice site itself. The mutation appears to weaken the splice site, decreasing its usage. The synonymous substitution in codon 275 lies 38 nucleotides 3′ of the 11q splice donor site. Its presence appears to increase the relative amount of D(11q) while decreasing the FL isoform but not D(11)(Figure 3 and supporting table S1). This suggests that the substitution favours the usage of D11(q) donor site competing with the FL isoform donor site. The substitution in codon 276 just 3 nucleotides further downstream does not seem to have this same effect. This suggests that codon 276 contains an important sequence element such as an ESE that favours use of the D(11q) splice site. A similar effect on splicing for these three synonymous changes was also observed in HEK 293 cells (data not shown).

D11q intensity seems to be higher in figure 3 than 1 and 2. Although we observed a certain degree of variability between biological replicates (at different transfection dates), changes in isoform proportion (comparing different variants) were consistent. In addition, we did not observe variability between technical

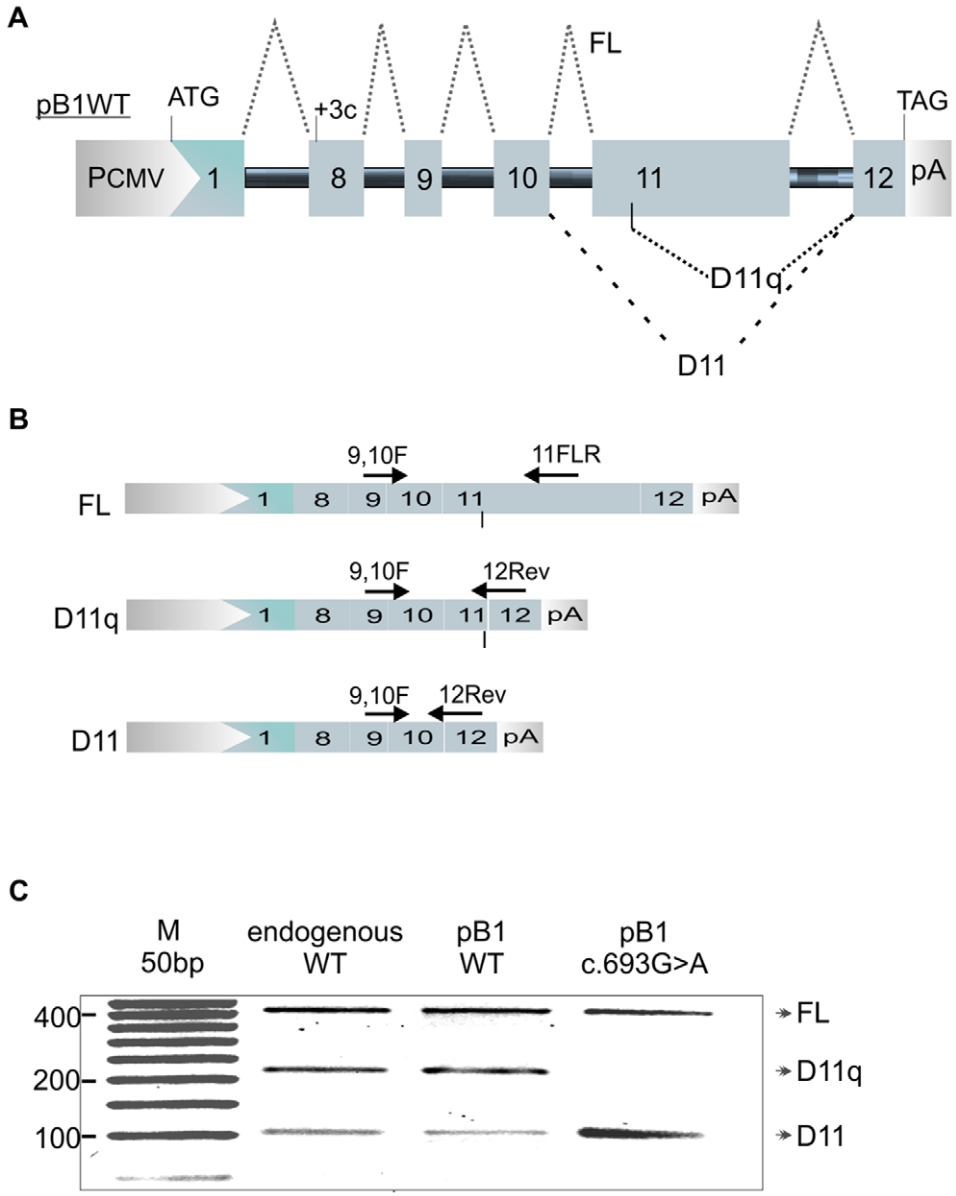


Figure 1. Minigene splicing assay of BRCA1 exon 11. A. The pB1 wild type (WT) version of the minigene is shown. PCMV = promoter of the pCDNA3 vector. ATG = start codon. TAG = stop codon. +3C = insertion of cytosine as the third nucleotide in exon 8. pA = poly A signal. 1 = exon 1 of the alfa globin gene. BRCA1 exons from 8 to 12 are numbered. The black solid line represents introns. Dotted lines show alternative splicing of exon 11. **B**. The three splicing isoforms FL, D11q and D11 and the position of specific oligos used for detection, are shown. **C**. Detection of BRCA1 exon 11 splicing isoforms for: _MCF7 endogenous BRCA1 (endogenous WT). _pB1 WT minigene transfected in MCF7 (pB1 WT). _pB1 minigene carrying the c.696G>A nucleotide substitution transfected in MCF7 (pB1 c.696G>A).
doi:10.1371/journal.pone.0037255.g001

replicates (minigene transfected on the same date). This could well be explained by the fact that the BRCA1 isoform proportions vary during the cell cycle [5] so that minimal changes in cell status (e.g. the number of passages) may account for the change in intensity of D11q isoforms between biological replicates.

In addition, these types of effects should be routinely tested with quantitative qRT-PCR methods.

Targeted deletions of the critical region

In order to identify putative splicing regulatory elements in the 'critical region' of *BRCA1* exon 11, we undertook a comparative genomic analysis of the *BRCA1* sequence of 9 eutherian mammals using the Ensembl genome browser (http://www.ensembl.org/index.html). The data show several conserved sequences along the 'critical region' (supporting Figure S2). In order to evaluate whether these conserved regions include splicing regulatory sequences we used SFmap (http://sfmap.technion.ac.il/) [13] for the prediction of splicing factor binding sites (supporting Figure S2).

We then performed a pB1 minigene deletion analysis of the most conserved regions containing putative binding sites for splicing regulatory proteins. The D1, D2, D3 and D4 deletions are highlighted in Figure 4a. Hybrid minigenes carrying each deletion were transiently transfected into MCF7 cells and splicing was

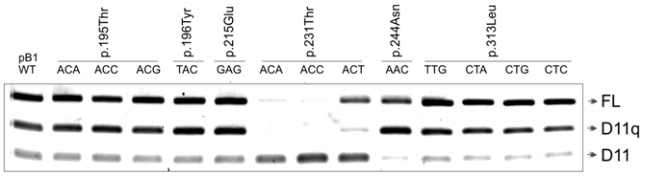


Figure 2. Minigene splicing assay of synonymous substitutions. Effect of synonymous *BRCA1* substitutions on splicing products full-length, D(11q) and D(11). RT-PCR products from transfection experiments using minigenes carrying codon substitution are shown.
doi:10.1371/journal.pone.0037255.g002

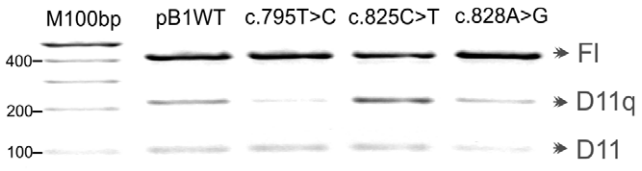


Figure 3. Minigene splicing assay of patients synonymous substitutions. Effect of synonymous *BRCA1* substitutions on splicing products full-length, D(11q) and D(11). RT-PCR products from transfection experiments using mutated minigenes are shown.
doi:10.1371/journal.pone.0037255.g003

analysed (Figure 4b). The results show that *BRCA1* exon 11 splicing is unchanged with deletion D1 and D4. Deletion D3 has a weak effect in reducing levels of the D(11q) isoform in favour of D(11). Transfection with the hybrid minigene carrying deletion D2 has the strongest effect, inducing almost complete skipping of exon 11.

The putative splicing factors predicted to bind in D2 and D3 by SFmap (http://sfmap.technion.ac.il/) and SpliceAid2 (www.introni.it/spliceaid.html) are shown in supporting Figure S3.

Discussion

That synonymous mutations are under evolutionary constraint due to splicing requirements has been demonstrated experimentally [14]. Hurst and Pal found a pronounced peak in the ratio of non-synonymous to synonymous substitutions between codons 200 and 300 of *BRCA1* exons 10 and 11 when comparing human/dog and mouse/rat gene alignments [11]. The presence of this so-called 'critical region' reflects an unusually low rate of translationally silent synonymous sequence changes, suggesting that purifying selection may be acting on this region to select against synonymous changes, possibly in order to preserve splicing regulatory elements. In this study, we have experimentally verified this hypothesis by testing synonymous substitutions at codon sites proposed by Hurst and Pal to have the highest non-synonymous to synonymous substitution ratios.

Of the 6 codons analysed, only synonymous substitutions at codon 231 and codon 244 affected splicing of *BRCA1* exon 11. The substitution at codon 244 increased levels of the D(11q) isoform to the detriment of the full and D(11)isoforms. However, it is difficult to speculate upon the relevance of this aberrant splicing effect. This is for two main reasons. Firstly, the effect on the splicing ratio is only relatively weak. Secondly, the D(11q) isoform has been shown to play a role in apoptosis [15]. However, a recent article has shown that, following knockdown of the nuclear chaperon Ubc9 (ubiquitin conjugating enzyme 9), D(11q) accumulation in the cytoplasm promotes growth and survival of breast cancer cells [16]. This may signify that an increase of D(11q) isoform may have deleterious effects in circumstances where Ubc9 is compromised.

Substitutions at codon 231 caused skipping of exon 11 with a marked increase in amounts of the D(11) isoform. Overexpression of this isoform in mouse epithelial mammary cells has been shown to cause atypical duct hyperplasia [17]. This could explain the biased synonymous codon usage observed at codon 231 and may reflect the necessity to preserve a regulatory sequence that protects against aberrant splicing, which would otherwise predispose to breast cancer.

Substitutions at codons 195, 196, 215 and 313 did not have any discernible effect on isoform levels. However, there may be other ways by which synonymous changes can exert an effect [18]. For instance tRNAs complimentary to different codons vary in their relative concentrations within a cell, meaning that synonymous codons may not be equally represented in terms of the relative abundance of their complimentary tRNAs [19]. This variable cellular availability of specific tRNAs can affect translational efficiency of protein product.

Nevertheless, the possibility that substitutions at these codon positions affect regulatory elements of splicing should not be

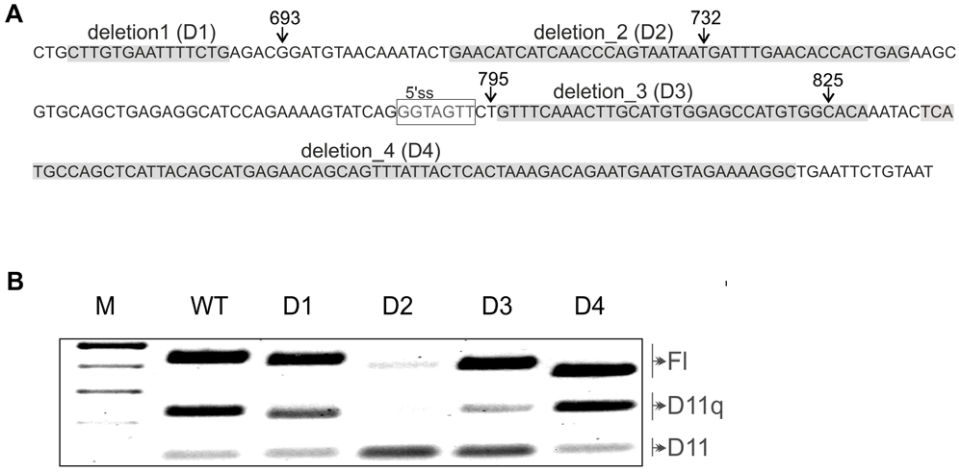


Figure 4. Minigene splicing assay of BRCA1 exon 11. A. Sequence of the critical region in exon 11 showing the deletion 1,2,3,4 (highlighted). The donor site (5′ss) generating D11q isoform is boxed. Arrows indicate nucleotide positions of the variations reported in Figure 2 and 3 to affect splicing. **B**. Transient transfection results for the hybrid minigenes carrying deletions.
doi:10.1371/journal.pone.0037255.g004

excluded. In fact protective variants might have occurred during evolution that compensate for aberrant splicing [14,20].

The c.693G>A substitution at codon 231 has previously been reported to cause exon 11 skipping in a patient with breast cancer [12]. This synonymous change in our minigene gave the same splicing outcome, proving the validity of the pB1 minigene as a splicing assay for *BRCA1* exon 11 variants. We therefore tested three further synonymous variants found in breast cancer patients to investigate their effect on *BRCA1* splicing. Our findings show that the mutations c.795T>C and c.825C>T decrease and increase levels of the D(11q) isoform respectively. However, for the same reasons mentioned above it is not possible to classify these variants as pathogenic mutations causing aberrant splicing since the role of D(11q) is not clear. In fact, the D(11q) isoform has been described as both causing apoptosis and causing cancer [15,16]. With these opposing roles in mind, we can hypothesise that an increased abundance of this isoform may have competing and contrasting effects in controlling cell proliferation or cell death depending on a patient's personal sequence context.

Both artificial and natural variants analysed in this study cause multiple splicing effects (Supporting Table S1). For instance down-regulation of the D(11q) isoform is not always accompanied by up-regulation of other isoforms. This suggests that different regulatory elements are affected that control one or all of the following events: usage of the exon 11 acceptor site, usage of the exon 11 donor site, usage of the D(11q) isoform donor site, and competition between the two donor sites.

Despite the fact that only synonymous substitutions in 2 out of 6 codons affected exon 11 splicing, the deletion analysis of *BRCA1* exon 11 appears to experimentally validate the hypothesis that maintenance of correct alternative splicing is the cause of selection against silent sequence variations in the *BRCA1* critical region. In fact, 2 out of 4 of these deletions affected the proportion of *BRCA1* splicing isoforms. In summary, deletions 2, and 3 changed *BRCA1* splicing ratios in different and opposing directions.

Deletion 2, in particular, caused almost complete skipping of exon 11 and was predicted to lose splicing regulatory sequences for binding of several splicing enhancer proteins (SC35, SRp20/30/40, NOVA 1 and YB1). Deletion 3, just next to the D(11q) donor site, decreased D(11q) isoform levels, probably due to the loss of a putative binding site for TIA1 (a donor site modulator) predicted by the SpliceAid analysis. NOVA 1 binding was predicted to bind both regions corresponding to deletion 2 and 3. Binding of NOVA1 at the end of an exon and beginning of an intron was proposed to induce exon inclusion [21]. In this case binding of NOVA1 upstream and downstream D11q isoform donor site might regulate production of this isoform.

We did not observe strong correlation on the splicing effect of nucleotide changes occurring inside the region of deletion 2 (c.732T>C) and deletion 3 (c.825C>T). This suggests that composite regulatory elements of splicing might be present in this region of BRCA1 exon 11 as has been previously suggested for CFTR exon 12 [22].

In order to fully understand the effects of altered splice isoform ratios in *BRCA1*-related cancer, it will be necessary to know the roles of each isoform individually and to understand the combined roles of different isoforms both in health and in tumorigenesis. Part of this understanding will require confirmation of which splicing factors are involved in regulating alternative splicing of *BRCA1*, including those suggested by this study that may be involved in the splicing of exon 11. Knowledge of these mechanisms would provide a framework for the development of new therapeutic agents capable of manipulating *BRCA1* splicing and treating *BRCA1*-related cancers (eg. breast ovarian and prostate cancer). At some level it may be that cancer predisposition is not so much down to whether individual isoforms are simply present or absent but rather that subtle alterations to a complex and nuanced isoform profile are particularly relevant. Such an isoform environment may interact with other cellular pathways to make conditions favourable or otherwise towards tumour development. If this is the case, it will provide an added challenge of complexity to our understanding of tumorigenesis. However, it may also allow novel and innovative approaches to manipulating the cellular environment in order to prevent and treat cancer.

Materials and Methods

Construction of the minigene

The pB1 WT minigene is shown in Figure 4. It consists of 6 exons (including part of their flanking introns) cloned in a modified version of the pCDNA3(+) vector (Invitrogen) under the control of the CMV promoter. The polylinker of the pCDNA3+ vector has been replaced with an adaptor containing the appropriate restriction sites for cloning of exon 1 of the α-globin gene with its 3′ flanking intronic region together with the relevant *BRCA1* genomic region from exon 8 to exon 12. Introns have been shortened by PCR amplification with oligonucleotides carrying a non complementary tail for specific restriction digestion and subsequent cloning in pCDNA3+.

Exon 1 of the α-globin gene is used as the first exon of the minigene as it provides a strong splice donor site at its 3′ end as well as an ATG start codon at its 5′ end.

Using specific oligonucleotides and a two-step PCR mutagenesis method [23], a stop codon was created in exon 12 and a single nucleotide insertion was created in exon 8 in order to maintain the correct reading frame. Several unique restriction sites are maintained in the sequence in order to facilitate subsequent mutagenesis and deletion analysis. Mutated minigenes and minigenes carrying deletions were generated through a two-step PCR overlap extension [24] using the pB1 WT construct as a template. The identity of all minigenes was checked by sequencing. The minigene complete sequence and oligonucleotidess used for cloning and mutagenesis are available upon request.

Cell Culture

Human breast cancer cell lines, MCF7 (ATCC number: HTB-22^{TM}), were grown in DMEM medium with 4500 mg/L glucose, pyruvate and L-glutamine supplemented with 1% penicillin/streptomycin and 10% fetal bovine serum. Cells were incubated at 37°C in a 5% CO2 atmosphere.

Transfection

Minigene plasmid vector was transfected into MCF7 cell lines using the FuGENE 6 transfection reagent from Roche. 100 µl of DMEM serum-free medium containing 2 µg of vector DNA and 3 µl of FuGENE reagent was incubated for 15 minutes at room temperature before the mixture was added in 6 cm-well cell cultures (50% confluent) in the presence of 10% fetal bovine serum.

RNA Extraction and RT-PCR

40 hours after transfection, RNA was extracted from cells using the RNeasy-plus kit from QIAGEN following the manufacturer's instructions. To analyse alternative splicing of *BRCA1* exon 11 in the pB1 minigene, RT-PCR was performed with 1.5 µg of total RNA using the pCSrev primer [5′ GCAACTAGAAGGCACAGTCGAGG 3′] to exclusively target only RNA products from the pB1 minigene. Given the large size of exon 11 (3426

nucleotides), the QIAGEN LongRange RT PCR kit was used following the manufacturer's instructions. 1/4 of the resulting cDNA was amplified in a PCR reaction using primers specific for the desired splice isoforms and PCR products were analysed by gel electrophoresis in 1.5% agarose. The forward primer [9–10F: 5′ ACTTATTGCAGTGTGGGAGA 3′] hybridises to the junction between exons 9 and 10. Reverse primers are a mixture of one specific for FL [11FLR: 5′ GGAGTCCGCCTATCATTACATG 3′] and one specific for D(11q) and D(11) [12R: 5′ CCA-GATGCTGCTTCACCCT 3′]. 11FLR hybridises within exon 11 distal to the 11q splice site. 12R hybridises to proximal exon 12 and overlaps the exon 10/12 junction and also the exon 11q/12 junction.

For the analysis of *BRCA1* alternative splicing of exons 9 and 10, RT-PCR was performed on 1 μg total RNA using random primers with the Promega kit. *BRCA1*-minigene specific PCR was performed on resulting cDNA using the forward specific primer alpha-8F [5′ GAGGCCCTGGAGAGGAcAA 3′] and the reverse primer 11+81Rev [5′ TCTCAGTGGTGTTCAAATCA 3′]. Alpha-8F hybridises to the junction between exon 1 of the α-globin gene and exon 8 of *BRCA1*. The c nucleotide in lowercase represents the insertion made in exon 8 of the pB1 minigene. 11+81Rev hybridises to exon 11 upstream of the donor site producing D(11q).

In order to eliminate heteroduplexes from mixed-template 1/10^{th} of PCR products were subjected to 'reconditioning PCR' for 6 cycles [25].

In silico analysis

Putative splicing regulatory sequences in *BRCA1* exon 11 were predicted using the computational tools SFmap [13] and SpliceAid2 [26], which enable accurate prediction and mapping of known splicing factor binding sites.

The following calculation parameters were chosen for SFmap:

- Scoring function: COS(WR);
- Medium stringency (Threshold [Significant] at p-value<0.005; [Suboptimal] at p-value<0.05);
- Window size: 50.

Supporting Information

Figure S1 Minigene splicing assay of BRCA1 exon 9 and 10. A. The pB1 wild type (WT) version of the minigene is shown. PCMV = promoter of the pCDNA3 vector. ATG = start codon. TAG = stop codon. +3C = insertion of cytosine as the third nucleotide in exon 8. pA = poly A signal. 1 = exon 1 of the alfa globin gene. BRCA1 exons from 8 to 12 are numbered. The black solid line represents introns. Dotted lines show alternative splicing of exon 9 and 10. **B**. Detection of BRCA1 splicing isoforms FL (inclusion of exon 9 and 10); D9 (skipping of exon 9); D10 (skipping of exon 10); D9,10 (skipping of exon 9 and 10).
(TIF)

Figure S2 BRCA1 alignments between nine eutherian mammals. The sequence of exon 11 critical region is in blue. Black characters represent the last 6 nucleotides of intron 10. The alternative donor site (5′ss) in exon 11 which gives rise to the Δ(11q) isoform is in white characters. Red are the nucleotide variations with respect to the human sequence. Ca. = Canis_familiaris; Eq. = Equus_caballus; Bo. = Bos_taurus; Ho. = Homo_sapiens; Pa. = Pan_troglodytes; Po. = Pongo_pygmaeus; Ma. = Macaca_mulatta; Mu. = Mus_musculus; Ra. = Rattus_norvegicus. Splicing regulatory proteins (predicted with SFmap) putatively binding to the human sequence are shown at the top of each sequence. The one corresponding to splicing regulatory motifs that are most conserved are highlighted. The list of putative sequences for splicing regulatory proteins and relative scores as reported in SFmap are listed.
(DOC)

Figure S3 Splicing factors predicted to bind. The sequence of deletion 2 and deletion 3 region are shown. Splicing factors predicted (by SFmap and/or SpliceAid) to bind these regions are listed.
(XLS)

Table S1 Information about all synonymous variants tested. The table includes the amino acid number, the nucleotide change, the dbSNP rs number, genomic location in hg19 coordinates (Chr 17 position),.PMID (for existing PubMed records) and our interpretation of the effects caused by the variations on different splicing events tested (isoform regulation).
(XLS)

Acknowledgments

We thank the staff at the Wessex Regional Genetics Laboratory and Prof Diana Eccles for their suggestions and support.

Author Contributions

Performed the experiments: MR AD CT DW DB. Analyzed the data: MR AD CT DW DB. Contributed reagents/materials/analysis tools: DB MR. Wrote the paper: MR AD CT DW DB.

References

1. Ryan PD (2010) Genetics of hereditary breast cancer. In: Chung DC, Haber DA, eds. Principles of clinical cancer genetics: a handbook from the Massachusetts General Hospital. New York: Springer. pp 41–52.
2. Orban TI, Olah E (2003) Emerging roles of BRCA1 alternative splicing. Mol Pathol 56: 191–197.
3. Grosso AR, Martins S, Carmo-Fonseca M (2008) The emerging role of splicing factors in cancer. EMBO Rep 9(11): 1087–93.
4. de la Grange P, Gratadou L, Delord M, Dutertre M, Auboeuf D (2010) Splicing factor and exon profiling across human tissues. Nucleic Acids Res 38(9): 2825–38.
5. Orban TI, Olah E (2001) Expression Profiles of BRCA1 Splice Variants in Asynchronous and in G1/S Synchronized Tumor Cell Lines. Biochem Biophys Res Commun 280(1): 32–8.
6. Lu M, Conzen SD, Cole CN, Arrick BA (1996) Characterization of functional messenger RNA splice variants of BRCA1 expressed in nonmalignant and tumor-derived breast cells. Cancer Res 56: 4578–81.
7. Thakur S, Zhang HB, Peng Y, Le H, Carroll B, et al. (1997) Localization of BRCA1 and a splice variant identifies the nuclear localization signal. Mol Cell Biol 17: 444–52.
8. Wilson CA, Payton MN, Elliott GS, Buaas FW, Cajulis EE, et al. (1997) Differential subcellular localization, expression and biological toxicity of BRCA1 and the splice variant BRCA1-delta11b. Oncogene 14: 1–16.
9. Chamary JV, Hurst LD (2009) The price of silent mutations. Sci Am 300: 46–53.
10. Orban TI, Olah E (2001) Purifying selection on silent sites - a constraint from splicing regulation? Trends Genet 17: 252–253.
11. Hurst LD, Pal C (2001) Evidence for purifying selection acting on silent sites in BRCA1. Trends Genet 17: 62–65.
12. Brandão RD, van Roozendaal K, Tserpelis D, García EG, Blok MJ (2011) Characterisation of unclassified variants in the BRCA1/2 genes with a putative effect on splicing. Breast Cancer Res Treat 129(3): 971–82.
13. Paz I, Akerman M, Dror I, Kosti I, Mandel-Gutfreund Y (2010) SFmap: a web server for motif analysis and prediction of splicing factor binding sites. Nucleic Acids Res 38: W281–W285.
14. Pagani F, Raponi M, Baralle FE (2005) Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc Natl Acad Sci U S A 102(18): 6368–72.

15. Yuli C, Shao N, Rao R, Aysola P, Reddy V, et al. (2007) BRCA1a has antitumor activity in TN breast, ovarian and prostate cancers. Oncogene 26(41): 6031–7.
16. Qin Y, Xu J, Aysola K, Begum N, Reddy V, Chai Y, et al. (2011) Ubc9 mediates nuclear localization and growth suppression of BRCA1 and BRCA1a proteins. J Cell Physiol 226(12): 3355–67.
17. Bachelier R, Vincent A, Mathevet P, Magdinier F, Lenoir GM, et al. (2002) Retroviral transduction of splice variant Brca1-Delta11 or mutant Brca1-W1777Stop causes mouse epithelial mammary atypical duct hyperplasia. Virchows Arch 440(3): 261–6.
18. Sauna ZE, Kimchi-Sarfaty C (2011) Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet 12(10): 683–91.
19. Dittmar KA, Goodenbour JM, Pan T (2006) Tissue-specific differences in human transfer RNA expression. PLoS Genet 2(12): e221.
20. Tang W, Fu YP, Figueroa JD, Malats N, Garcia-Closas M, et al. (2012) Mapping of the UGT1A locus identifies an uncommon coding variant that affects mRNA expression and protects from bladder cancer. Hum Mol Genet 15;21(8): 1918–30.
21. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456(7221): 464–9.
22. Pagani F, Stuani C, Tzetis M, Kanavakis E, Efthymiadou A, et al. (2003) New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum Mol Genet 12(10): 1111–20.
23. Good L, Nazar RN (1992) An improved thermal cycle for two-step PCR-based targeted mutagenesis. Nucleic Acids Res 20: 4934.
24. Lee J, Lee HJ, Shin MK, Ryu WS (2004) Versatile PCR-mediated insertion or deletion mutagenesis. Biotechniques 36(3): 398–400.
25. Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucleic Acids Res 30(9): 2083–8.
26. Piva F, Giulietti M, Burini AB, Principato G (2012) SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. Hum Mutat 33(1): 81–5.

BRCA1 exon 11 alternative splicing, multiple functions and the association with cancer

Claudia Tammaro, Michela Raponi, David I. Wilson and Diana Baralle[1]

Human Development and Health Academic Unit, Faculty of Medicine, University of Southampton, Institute of Developmental Sciences Building, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, U.K.

Abstract

BRCA1 (breast cancer early-onset 1) alternative splicing levels are regulated in a cell-cycle- and cell-type-specific manner, with splice variants being present in different proportions in tumour cell lines as well as in normal mammary epithelial cells. The importance of this difference in the pathogenesis of breast cancer has yet to be determined. Developing an understanding of the impact of BRCA1 isoform ratio changes on cell phenotype will be of value in breast cancer and may offer therapeutic options. In the present paper, we describe the splicing isoforms of *BRCA1* exon 11, their possible role in cancer biology and the importance of maintaining a balanced ratio.

Background

Splicing is an important mechanism which contributes to pre-mRNA maturation. Control of this process determines which pre-mRNA sequences are removed (introns) and which are spliced to form the mature RNA (exons). The machinery driving the splicing process is the spliceosome, a complex of many interacting proteins and snRNAs (small nuclear RNAs) [1].

In order to perform accurate splicing, the spliceosome must recognize the donor site at the exon–intron junction and the acceptor site at the intron–exon junction. These elements are necessary, but not always sufficient. Splice site strength often depends on nearby sequence elements known as splicing enhancers and silencers [2].

Different combinations of splice site selections allow a large number of mRNA and protein isoforms to be generated from a lower number of genes; this process, known as alternative splicing, is a critical mechanism and generates transcriptome and proteome complexity. Alternative splicing has been found to be associated with diseases, including cancer, and can influence cell proliferation, motility and response to drugs [3].

Mutations in splicing regulatory sequences as well as alteration in the levels of splicing regulatory proteins may affect alternative splicing. BRCA1 (breast cancer early-onset 1) mutations predicted to affect BRCA1 function have been found in 40–45 % cases of hereditary breast cancer. In addition, a large number of BRCA1 sequence variants have been found in patients but their clinical significance is unknown. These unclassified variants could include potential splicing mutations [4].

Key words: alternative splicing, breast cancer early-onset 1 (BRCA1), cancer.

Abbreviations used: BRCA1, breast cancer early-onset 1; BARD1, BRCA1-associated RING domain protein; BRCT, BRCA1 C-terminus; NLS, nuclear localization signal.

[1]To whom correspondence should be addressed (email D.Baralle@soton.ac.uk).

Several mRNA splicing isoforms have been identified for the *BRCA1* gene in normal tissues. Among those that result from exon-skipping events and retain the translational reading frame are variants that skip exon 5, exon 11 (all of it or most of it), exons 2–10, exons 9–11, exons 14–17 and exons 14–18 [5]. Some of these splicing isoforms have been associated with breast and ovarian cancer [6]. In particular, the relative levels of BRCA1 isoforms associated with exon 11 alternative splicing appears to be different between normal and cancer tissues/cell lines as well as between the phases of the cell cycle [7]. These isoforms include BRCA1 full-length (inclusion of all exons), Δ11 (skipping of exon 11), Δ11q (partial skipping of exon 11, throughout the use of a donor site within exon 11), Δ9,10,11q (skipping of exons 9, 10 and partial skipping of exon 11) and IRIS isoforms (skipping of exons 12–24, but retaining a short segment from intron 11) (Figures 1, 2 and 3). With the exception of IRIS, it is unclear whether altered levels of exon 11 splicing isoforms play a pathogenic role in cancer.

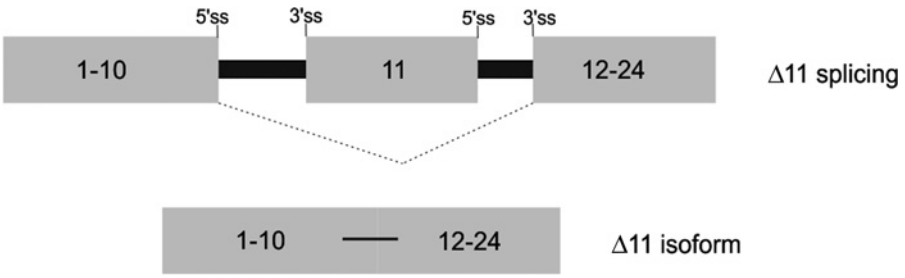
BRCA1 protein domains and multiple roles

The breast cancer susceptibility gene *BRCA1* is a tumour-suppressor gene that was first identified on the basis of its linkage to early-onset breast and ovarian cancer in women [8]. The *BRCA1* gene has 24 exons, including two untranslated exons, and encodes a protein of 1863 amino acids with three distinct regions of protein interaction: the RING domain, the RAD51-interaction domain and the BRCT (BRCA1 C-terminus) domain [8].

The N-terminal RING finger domain has been found to bind to several proteins, including formation of heterodimers with BARD1 (BRCA1-associated RING domain protein) [9]. Both BRCA1 and BARD1 contain the RING finger domain that seems to be important for several tumour-suppression functions [10].

Figure 1 | **BRCA1 Δ11 isoform**

BRCA1 exons 1–24 are numbered. 5′ss and 3′ss represent the donor and the acceptor sites respectively. The black line represents the introns and the grey square represents the exons. The broken line shows the alternative splicing of exon 11.



The RAD51-interaction domain is enclosed by exon 11 and is involved in repair of double-strand DNA breaks [11]. It also contains multiple protein-binding sites, in addition to those for the RAD51 and RAD50 complex [12]. Loss of RAD51 binding may increase the risk of cancer as there could be an increment in the amount of damaged DNA.

The C-terminus of BRCA1 contains an acidic domain that can function as a transcriptional activation domain [8,13]. This domain contains a tandem repeat of approximately 95 amino acids called BRCT [14]. This domain is also found in proteins such as BARD1 which is involved in cell-cycle control and DNA repair [15].

BRCA1 has been implicated in diverse cancer-related activities, including roles in cell-cycle progression, DNA repair, DNA damage-responsive cell-cycle checkpoints, transcription regulation, ubiquitination, chromosome re-modelling and apoptosis [16].

The presence of different BRCA1 splicing isoforms that are naturally expressed in various cellular settings may be the reason that BRCA1 is involved in such different activities.

The Δ11 isoform

The Δ11 isoform is composed of 21 coding exons, arising from skipping of exon 11 (Figure 1). Exon 11 is a large exon of 3.4 kb. The acceptor site (at the intron–exon junction) and the donor site (at the exon–intron junction) of exon 11 are predicted to have weak splice site consensus sequences and therefore may be also surrounded by additional regulatory sequences in order to be recognized. We have shown recently that such regulatory sequences exist at the beginning of exon 11 and they are important for regulating BRCA1 alternative splicing [17]. The identification of which factors bind to these sequences would provide insight into the mechanism of BRCA1 alternative splicing regulation and especially the maintenance of BRCA1 isoform ratios.

Although studies have attempted to determine the role of the Δ11 isoform, its importance in cancer has yet to be determined and has been implicated in both cell death and cell proliferation (Table 1).

Table 1 | **Δ11 isoform**

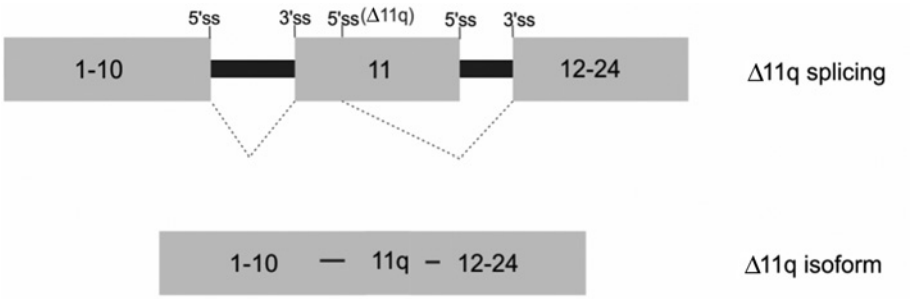
Manipulation	Effect	Reference(s)
Overexpression	Mouse embryonic fibroblasts fail to maintain mitosis and undergo apoptosis	[32]
Exclusive expression	Reduced DNA damage-induced phosphorylation. Impaired γ irradiation-induced RAD51 focus formation. Slow cell growth, quick to reach senescence and the accumulation of aneuploidy. Decreased expression of genes involved in the spindle checkpoint/defective spindle checkpoint. Embryonic lethality (rescued by elimination of one p53 allele). Often accompanied by alterations in p53	[19–21,33]
Depletion	Mammary gland abnormalities and uterine hyperplasia with spontaneous tumour formation in mice. Abnormal centrosome amplification and reduction of the G_1 population in cells	[22]

Both BRCA1-null mice and mice that exclusively express BRCA1 Δ11 undergo embryo lethality [18,19]. However, embryos with exclusive expression of BRCA1 Δ11 die later, suggesting that, at least in part, the Δ11 isoform can compensate for the lack of the other BRCA1 isoforms during early embryogenesis.

Embryos exclusively expressing BRCA1 Δ11 (Δ11/Δ11 embryos) and also heterozygous for a p53 allele are able to survive, probably due to a reduction of the apoptotic barrier. These mice are susceptible to premature aging and

Figure 2 | **BRCA1 Δ11q isoform**

BRCA1 exons 1–24 are numbered. 5′ss and 3′ss represent the donor and the acceptor sites respectively. The black line represents the introns and the grey square represents the exons. The broken line shows the alternative splicing of exon 11q, which uses another 5′ splice site.



also tumour formation [19,20]. In addition to these mouse models, *in vitro* studies have utilized cells that exclusively express BRCA1 Δ11.

Studies of mouse embryonic fibroblasts homozygous for the allele BRCA1 Δ11 (Δ11/Δ11) showed poor proliferation, suggesting a role of the BRCA1 Δ11 isoform in preventing tumour formation. However, these cells (BRCA1 Δ11/Δ11 fibroblasts) when immortalized proliferated faster than control immortalized fibroblasts [20]. A consideration is that the Δ11 isoform is inefficient at binding the protein RAD51 [21]. This can alter the capacity of Δ11/Δ11 cells to repair DNA double-strand breaks and cause an accumulation of defects in the cell cycle. This genetic instability and accumulation of defects may well explain the inconsistent behaviour of Δ11/Δ11 cells and mice that have been shown to potentially induce both apoptosis and a tendency to malignant transformation.

Besides studying the exclusive expression of the Δ11 isoform, Kim et al. [22] investigated mice lacking the Δ11 isoform. Female mice showed hyperplasia and spontaneous tumour in the gynaecological system, suggesting that the BRCA1 Δ11 isoform is involved in repressing tumour formation [22]. However, the cDNA knockin approach employed in this study, to specifically block BRCA1 alternative splicing from exon 10 to exon 12, could have generated mutant mice lacking not only the Δ11 isoform, but also other BRCA1 isoforms. For instance, as suggested by Kim et al. [22], the possibility cannot be excluded that depletion of BRCA1 IRIS, rather than depletion of the Δ11 isoform, is responsible for the hyperplasia and spontaneous tumour observed. As both BRCA1 IRIS and the Δ11 isoform seem to be involved in tumour formation neither hypothesis appeals.

A further possibility is that loss of balance between each isoform (e.g. the selective expression of the remaining BRCA1 isoforms or a combination of this with depletion of the Δ11 and BRCA1 IRIS isoforms) is causative of the phenotypes observed. In addition, overexpression (using retroviral transduction) of the Δ11 isoform in mammary epithelial cells showed hyperplasia in mice injected with these cells [22].

Table 2 | **Δ11q isoform**

Manipulation	Effect	Reference
Overexpression	Decreases proliferation of breast and ovarian cancer cells	[32]
Exclusive expression	Promotes growth and survival of breast cancer cells	[23]

The Δ11q isoform

The Δ11q isoform derives from the alternative choice of a donor site within human *BRCA1* exon 11 (Figure 2). It should be noted that the Δ11q isoform has not been described in mouse. Importantly, the splice site used within human *BRCA1* exon 11 to produce the Δ11q isoform is not predicted in the murine sequence.

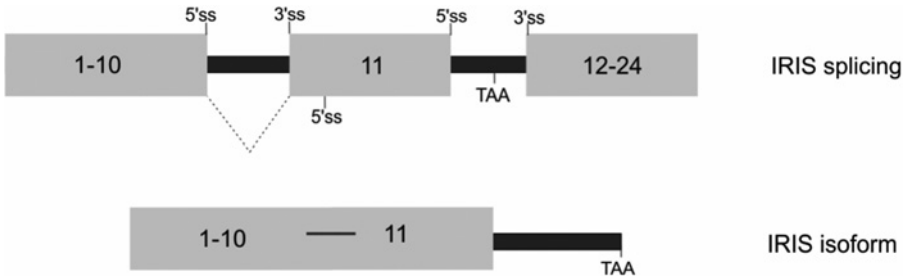
In the Δ11q isoform, part of exon 11 (nucleotides 905–4215) is excluded. As a consequence of this exclusion, the isoform (like the Δ11 isoform) lacks the NLS (nuclear localization signal). Isoforms lacking the NLS can be transported in the nucleus via an alternative mechanism that requires Ubc9 [23]. However, when Ubc9 is depleted, the Δ11q isoform becomes exclusively cytoplasmic and promotes growth and survival of breast cancer cells [23].

Although cytoplasmic Δ11q seems to play a role in tumour formation, the overexpression of BRCA1 Δ11q is able to inhibit growth of breast cancer cells [24] (Table 2). This ambiguous capacity of BRCA1 Δ11q to induce or repress cell proliferation in different contexts seems strictly related to its localization; overexpression may cause an increase in nuclear Δ11q that promotes apoptosis, whereas nuclear depletion (cytoplasmic retention) of Δ11q has the opposite effect.

To highlight further the different functional roles of BRCA1 isoforms, a study of human mammary epithelial cells has shown that the expression levels of many genes are altered in a BRCA1 isoform-specific manner [25]. Specifically, overexpression of Δ11q induced activation of

Figure 3 | **BRCA1 IRIS isoform**

BRCA1 exons 1–24 are numbered. 5′ss and 3′ss represent the donor and the acceptor sites respectively. TAA = stop codon. The black line represents the introns and the grey square represents the exons. The broken line shows the alternative splicing of BRCA1 IRIS.



22 genes, whereas overexpression of the full-length isoform (that included all BRCA1-coding exons) did not. Moreover, whereas overexpression of the full length also repressed several genes, transfection of Δ11q did not result in repression.

This BRCA1 isoform-specific gene regulation is intriguing and suggests a hypothesis that a balanced ratio of BRCA1 isoforms is required to maintain normal cell physiology.

BRCA1 IRIS

BRCA1 also encodes a 1399 residue polypeptide, termed BRCA1 IRIS. It consists of an open reading frame from codon 1 in exon 2 up to the first 34 codons in intron 11 (Figure 3). It is not completely clear whether this recently described BRCA1 IRIS isoform is the product of alternative splicing at the intron 11 donor site or simply the result of an alternative promoter usage as described previously [26]. A possible explanation is that the use of a different promoter may also affect the rate of transcription which in turn may affect splicing at the intron 11 donor site with the production of *BRCA1* IRIS mRNA.

BRCA1 IRIS plays an important positive role in DNA replication [26]. Unlike full-length BRCA1, BRCA1 IRIS does not interact with BARD1; it is exclusively chromatin-associated, present in G_0 cells and overexpression promotes cell proliferation, breast cancer and cisplatin resistance in ovarian cancer cells [26–30]. The effect on cells of BRCA1 IRIS depletion or overexpression is shown in Table 3.

Knockdown experiments of BRCA1 isoforms targeting exon 12, have claimed that down-regulation of the full-length isoform can cause overexpression of BRCA1 IRIS because of mRNA stabilization [31]. Whether this effect is related only to depletion of the full-length isoform or is also due to depletion of other isoforms that include exon 12 (e.g. Δ11 and Δ11q) needs to be determined.

The recent discovery of this oncogene-like BRCA1 IRIS isoform is of potential significance as it could represent a therapeutic target in breast cancer. In addition, discovery of BRCA1 IRIS should challenge the interpretation of various data. For instance, as suggested by ElShamy and Livingston [26], the embryonic lethality observed in BRCA Δ11/Δ11 mice was probably attributable to a loss of BRCA1 IRIS isoform rather than to the exclusive expression of the Δ11 isoform.

Table 3 | **IRIS isoform**

Manipulation	Effect	Reference(s)
Overexpression	Promotes cell proliferation. Promotes cisplatin resistance in ovarian cancer cells. Promotes breast cancer. Accelerates DNA synthesis	[26–29]
Exclusive expression	Promotes formation of aggressive and invasive breast tumours	[19–21,33]
Depletion	Marked depression of early S-phase nucleotide uptake. Slow DNA synthesis	[26]

Moreover, mRNA levels of the full-length, Δ11 and Δ11q BRCA1 isoforms could have been misinterpreted considering that detection methods using oligonucleotides or probes upstream of intron 11 do not discriminate the IRIS isoform.

Concluding remarks

The fact that the physiological and pathogenic role of most BRCA1 isoforms tends to be opposite strongly suggests that cell fate can be switched in a particular direction that is dependent on changes in the overall splicing ratio rather than changes in a specific isoform. Understanding the significance of this ratio in cancer/apoptosis may be crucial in determining cancer predisposition and would also offer therapeutic options.

Currently, unclassified variants found to affect the ratio of natural BRCA1 splicing isoforms in patients with breast

cancer are not classified as pathogenic mutations. Establishing that an unbalanced ratio can predispose an individual to cancer is fundamental in order to change the way that these variants are classified. Central to this, it will be critical a full understanding of the potential role of BRCA1 splicing.

Acknowledgements

We thank Professor Diana Eccles.

Funding

D.B., M.R. and C.T. are supported by EURASNET, Cancer Research UK and the University of Southampton.

References

1 Rappsilber, J., Ryder, U., Lamond, A.I. and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. Genome Res. **12**, 1231–1245
2 Moore, M.J. and Sharp, P.A. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature **365**, 364–368
3 Krawczak, M., Reiss, J. and Cooper, D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum. Genet. **90**, 41–54
4 Schwartz, G.F., Hughes, K.S., Lynch, H.T., Fabian, C.J., Fentiman, I.S., Robson, M.E., Domchek, S.M., Hartmann, L.C., Holland, R., Winchester, D.J. et al. (2009) Proceedings of the international consensus conference on breast cancer risk, genetics, & risk management, April, 2007. Breast J. **15**, 4–16
5 Orban, T.I. and Olah, E. (2003) Emerging roles of BRCA1 alternative splicing. Mol. Pathol. **56**, 191–197
6 Lixia, M., Zhijian, C., Chao, S., Chaojiang, G. and Congyi, Z. (2007) Alternative splicing of breast cancer associated gene *BRCA1* from breast cancer cell line. J. Biochem. Mol. Biol. **40**, 15–21
7 Orban, T.I. and Olah, E. (2001) Expression profiles of BRCA1 splice variants in asynchronous and in G1/S synchronized tumor cell lines. Biochem. Biophys. Res. Commun. **280**, 32–38
8 Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W. et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. Science **266**, 66–71
9 Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., Ogata, H. and Ohta, T. (2001) The RING heterodimer BRCA1–BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. J. Biol. Chem. **276**, 14537–14540
10 De Brakeleer, S., De Grève, J., Loris, R., Janin, N., Lissens, W., Sermijn, E. and Teugels, E. (2010) Cancer predisposing missense and protein truncating *BARD1* mutations in non-BRCA1 or BRCA2 breast cancer families. Hum. Mutat. **31**, E1175–E1185
11 Zhang, J. and Powell, S.N. (2005) The role of the BRCA1 tumor suppressor in DNA double-strand break repair. Mol. Cancer Res. **3**, 531–539
12 Deng, C.X. and Brodie, S.G. (2000) Roles of BRCA1 and its interacting proteins. BioEssays **22**, 728–737
13 Monteiro, A.N., August, A. and Hanafusa, H. (1996) Evidence for a transcriptional activation function of BRCA1 C-terminal region. Proc. Natl. Acad. Sci. U.S.A. **93**, 13595–13599
14 Williams, R.S., Green, R. and Glover, J.N. (2001) Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. Nat. Struct. Biol. **8**, 838–842
15 Vallon-Christersson, J., Cayanan, C., Haraldsson, K., Loman, N., Bergthorsson, J.T., Brøndum-Nielsen, K., Gerdes, A.M., Møller, P., Kristoffersson, U., Olsson, H. et al. (2001) Functional analysis of *BRCA1* C-terminal missense mutations identified in breast and ovarian cancer families. Hum. Mol. Genet. **10**, 353–360
16 Rosen, E.M., Fan, S., Pestell, R.G. and Goldberg, I.D. (2003) *BRCA1* gene in breast cancer. J. Cell. Physiol. **196**, 19–41
17 Raponi, M., Douglas, A.G., Tammaro, C., Wilson, D.I. and Baralle, D. (2012) Evolutionary constraint helps unmask a splicing regulatory region in *BRCA1* exon 11. PLoS ONE **7**, e37255
18 Ludwig, T., Chapman, D.L., Papaioannou, V.E. and Efstratiadis, A. (1997) Targeted mutations of breast cancer susceptibility gene homologs in mice: lethal phenotypes of *Brca1*, *Brca2*, *Brca1/Brca2*, *Brca1/p53*, and *Brca2/p53* nullizygous embryos. Genes Dev. **11**, 1226–1241
19 Xu, X., Qiao, W., Linke, S.P., Cao, L., Li, W.M., Furth, P.A., Harris, C.C. and Deng, C.X. (2001) Genetic interactions between tumor suppressors *Brca1* and *p53* in apoptosis, cell cycle and tumorigenesis. Nat. Genet. **28**, 266–271
20 Cao, L., Li, W., Kim, S., Brodie, S.G. and Deng, C.X. (2003) Senescence, aging, and malignant transformation mediated by p53 in mice lacking the *Brca1* full-length isoform. Genes Dev. **17**, 201–213
21 Huber, L.J., Yang, T.W., Sarkisian, C.J., Master, S.R., Deng, C.X. and Chodosh, L.A. (2001) Impaired DNA damage response in cells expressing an exon 11-deleted murine *Brca1* variant that localizes to nuclear foci. Mol. Cell. Biol. **21**, 4005–4015
22 Kim, S.S., Cao, L., Lim, S.C., Li, C., Wang, R.H., Xu, X., Bachelier, R. and Deng, C.X. (2006) Hyperplasia and spontaneous tumor development in the gynecologic system in mice lacking the BRCA1-Δ11 isoform. Mol. Cell. Biol. **26**, 6983–6992
23 Qin, Y., Xu, J., Aysola, K., Begum, N., Reddy, V., Chai, Y., Grizzle, W.E., Partridge, E.E., Reddy, E.S. and Rao, V.N. (2011) Ubc9 mediates nuclear localization and growth suppression of BRCA1 and BRCA1a proteins. J. Cell. Physiol. **226**, 3355–3367
24 Maniccia, A.W., Lewis, C., Begum, N., Xu, J., Cui, J., Chipitsyna, G., Aysola, K., Reddy, V., Bhat, G., Fujimura, Y. et al. (2009) Mitochondrial localization, ELK-1 transcriptional regulation and growth inhibitory functions of BRCA1, BRCA1a, and BRCA1b proteins. J. Cell. Physiol. **219**, 634–641
25 McEachern, K.A., Archey, W.B., Douville, K. and Arrick, B.A. (2003) *BRCA1* splice variants exhibit overlapping and distinct transcriptional transactivation activities. J. Cell. Biochem. **89**, 120–132
26 ElShamy, W.M. and Livingston, D.M. (2004) Identification of BRCA1-IRIS, a *BRCA1* locus product. Nat. Cell Biol. **6**, 954–967
27 Chock, K.L., Allison, J.M., Shimizu, Y. and ElShamy, W.M. (2010) BRCA1-IRIS overexpression promotes cisplatin resistance in ovarian cancer cells. Cancer Res. **70**, 8782–8791
28 Elshamy, W.M. (2010) Induction of breast cancer in wild type p53 cells by BRCA1-IRIS overexpression. Hawaii Med. J. **69**, 200–201
29 Chock, K., Allison, J.M. and Elshamy, W.M. (2010) BRCA1-IRIS overexpression abrogates UV-induced $p38^{MAPK}$/p53 and promotes proliferation of damaged cells. Oncogene **29**, 5274–5285
30 Shimizu, Y., Luk, H., Horio, D., Miron, P., Griswold, M., Iglehart, D., Hernandez, B., Killeen, J. and ElShamy, W.M. (2012) BRCA1-IRIS overexpression promotes formation of aggressive breast cancers. PLoS ONE **7**, e34102
31 Shimizu, Y., Mullins, N., Blanchard, Z. and Elshamy, W.M. (2012) BRCA1/p220 loss triggers BRCA1-IRIS overexpression via mRNA stabilization in breast cancer cells. Oncotarget **3**, 299–313
32 Bachelier, R., Vincent, A., Mathevet, P., Magdinier, F., Lenoir, G.M. and Frappart, L. (2002) Retroviral transduction of splice variant Brca1-Δ11 or mutant Brca1-W1777Stop causes mouse epithelial mammary atypical duct hyperplasia. Virchows Arch. **440**, 261–266
33 Wang, R.H., Yu, H. and Deng, C.X. (2004) A requirement for breast-cancer-associated gene 1 (*BRCA1*) in the spindle checkpoint. Proc. Natl. Acad. Sci. U.S.A. **101**, 17108–17113

Received 5 June 2012
doi:10.1042/BST20120140