



Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0377.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	13-Dec-2010
Complete List of Authors:	Raponi, Michela; 2Human Genetics DivisionUniversity of Southampton, Human Genetics DivisionUniversity; University of Southampton, Genetics Kralovicova, Jana; University of Southampton, Genetics Copson, Ellen; University of Southampton, Genetics Divina, Petr; Institute of Molecular Genetics, Bioinformatics Eccles, Diana; School of Medicine, University of Southampton, Princess Anne Hospital, Wessex Clinical Genetics Service; University of Southampton, Genetics Johnson, Peter; University of Southampton, Genetics Baralle, Diana; University of Cambridge, Medical Genetics; University of Southampton, Genetics Vorechovsky, Igor; University of Southampton, Human Genetics
Key Words:	RNA, BRCA1, mutation, gene, splicing

SCHOLARONE™
Manuscripts

**Prediction of single-nucleotide substitutions that result in exon skipping:
identification of a splicing silencer in *BRCA1* exon 6**

Michela Raponi^{1,*}, Jana Kralovicova^{1,*}, Ellen Copson^{1,*}, Petr Divina², Diana Eccles¹,
Peter Johnson¹, Diana Baralle¹, Igor Vorechovsky¹

¹University of Southampton School of Medicine, Southampton, United Kingdom

²Institute of Molecular Genetics, Prague, Czech Republic

^{*}The first three authors contributed equally to this work.

Correspondence to

Dr. I. Vorechovsky

University of Southampton School of Medicine

Division of Human Genetics, MP808

Tremona Road

Southampton SO16 6YD

United Kingdom

Email: i.vorechovsky@southampton.ac.uk

ABSTRACT

Missense, nonsense and translationally silent mutations can inactivate genes by altering the inclusion of mutant exons in mRNA, but their overall frequency amongst disease-causing exonic substitutions is unknown. Here, we have tested missense and silent mutations deposited in the *BRCA1* mutation databases of unclassified variants for their effects on exon inclusion. **Analysis of 21 *BRCA1* variants using minigene assays** revealed a single exon-skipping mutation c.231G>T. Comprehensive mutagenesis of an adjacent 12-nt segment showed that this **silent mutation** resulted in a higher level of exon skipping than 35 other single-nucleotide substitutions. Exon inclusion levels of mutant constructs correlated significantly with predicted splicing enhancers/silencers, prompting the development of two online utilities freely available at <http://www.dbass.org.uk>. EX-SKIP quickly estimates which allele is more susceptible to exon skipping whereas HOT-SKIP examines all possible mutations at each exon position and identifies candidate exon-skipping positions/substitutions. We demonstrate that the distribution of **exon-skipping and disease-associated substitutions previously identified in coding regions** was biased toward top-ranking HOT-SKIP mutations. **Finally, we show that proteins 9G8, SC35, SF2/ASF, Tra2 and hnRNP A1 were associated with significant alterations of *BRCA1* exon 6 inclusion in the mRNA.** Together, these results facilitate prediction of exonic substitutions that reduce exon inclusion in mature transcripts.

Key words: RNA, *BRCA1*, mutation, splicing, gene

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

INTRODUCTION

Exonic single-nucleotide substitutions are the most common type of mutations in human genetic disease [Cooper and Krawczak, 1993]. Missense (amino acid altering) and nonsense (stop codon creating) mutations form by far the largest group of mutations deposited in the Human Gene Mutation Database, accounting for ~56% of over 92,000 alterations recorded until May 2010 [Stenson et al., 2009]. However, nonsense, missense and also translationally silent mutations can inactivate genes by altering the inclusion of the mutant exon in mature transcripts [Cartegni et al., 2002; Cooper and Mattox, 1997; Pagani et al., 2005; Valentine, 1998]. The most common outcome of these DNA alterations is exon skipping, which is usually dramatic in severe disease phenotypes as it is associated with diminished expression of canonical mRNAs. On the other hand, single-nucleotide variants in the coding region may also induce subtle changes in the relative expression of exon-lacking or exon-containing transcripts or activate or repress decoy splice sites [Cartegni et al., 2002; Cooper et al., 2009]. The exact proportion of such DNA alterations among disease-causing exonic substitutions is, however, unknown because routine mutation screening has been largely based on DNA analysis and RNA samples have not been examined in the majority of these cases.

Exonic substitutions often create or eliminate short elements that inhibit or activate exon inclusion or splice-site selection, termed splicing silencers or enhancers, respectively. These motifs are abundantly present both in *exons* (ESSs, ESEs) and *introns* (ISSs, ISEs) and have been derived by computational and/or experimental approaches [Fairbrother et al., 2002; Liu et al., 2000; Wang et al., 2004; Zhang and Chasin, 2004]. However, the effect of exonic mutations and variants on splicing is strongly dependent on

the sequence context and does not always correlate with *in silico* prediction [Goren et al., 2006; Lastella et al., 2006; Raponi et al., 2007; Skoko et al., 2008]. Although weak correlation with exon inclusion levels in minigenes was observed [Kralovicova and Vorechovsky, 2007], the predictive value of *ab initio* computation tools is currently suboptimal.

Splicing defects resulting from exonic substitutions have been found in a growing number of disease genes, such as *CFTR* [Pagani et al., 2005; Raponi et al., 2007], *MLH1* [Auclair et al., 2006; Stella et al., 2001; Tournier et al., 2008], *ATM* [Teraoka et al., 1999], *NF1* [Ars et al., 2000], *SMN2* [Lorson and Androphy, 2000] or *BRCA1* [Liu et al., 2001]. *BRCA1* (OMIM #113705) is one of the two most important breast cancer susceptibility genes [King et al., 2003], encoding a 1863 amino-acid protein involved in DNA damage repair and transcription regulation [Gowen et al., 1998]. Inheritance of a protein truncating mutation in this tumour suppressor gene is associated with a 40-80% lifetime risk of breast cancer [King et al., 2003]. Over 550 distinct missense mutations have been reported in patients with a strong family history of breast cancer and deposited in the Breast Cancer Information Core database (BIC; <http://research.nhgri.nih.gov/bic/>), but the functional significance of many of these changes is unclear, with profound implications for genetic counseling.

In the present study, we have systematically tested 21 missense and silent mutations selected from the *BRCA1* mutation databases of unclassified variants for their effects on exon inclusion in the mRNA using two non-overlapping 3-exon minigenes. We have identified a single exon-skipping mutation c.231G>T. Systematic mutagenesis of a 12-nt segment surrounding this alteration showed that this change gave the highest exon

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

skipping levels among 36 minigene substitutions and revealed a regular pattern of exon inclusion variations across this motif. We also show that recognition of this exon is influenced by heterogenous nuclear ribonucleoprotein (hnRNP) A1 and serine/arginine (SR) proteins SC35, 9G8 and Tra2. In addition, we have developed novel online utilities (termed EX-SKIP and HOT-SKIP) for *ab initio* prediction of exonic substitutions that induce exon skipping. These results reveal a new splicing regulatory element in *BRCA1* exon 6 and facilitate identification of mutations or variants in coding sequences that lead to exon exclusion from mature transcripts.

MATERIAL AND METHODS

Selection of *BRCA1* variants

The BIC database, an open-access on-line repository of mutations and polymorphisms of the inherited breast cancer genes, was searched to identify *BRCA1* missense and silent mutations reported prior to February 2006 (Supp. Table S1). Mutation nomenclature was based on cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the *BRCA1* GenBank reference sequence U14680 (www.hgvs.org/mutnomen). The initiation codon was codon 1. The number of cases per mutation varied between 1 and 222. For further experimental study, we selected exons that were flanked by smaller introns that would allow us to construct unabridged splicing reporters containing a high frequency of substitutions *in vivo*.

Construction of splicing reporters

OLIGO[®] online software (<http://www.oligo.net>) was used to design PCR primers to amplify three-exon *BRCAl* minigenes 5-7 and 17-19 (Supp. Table S2). Amplification was performed with anonymous genomic DNA, 10 pmol of each primer and a mixture of *Taq/Pfu* DNA polymerases. PCR products were visualized on agarose gels, DNA fragments were extracted using GENECLAN[®] and cloned into the mammalian expression vector pCR3.1 (Invitrogen). Restriction enzymes used for cloning were *Bam*HI/*Xba*I (minigene 5-7) and *Hind*III/*Xho*I (minigene 17-19). Digested PCR products were purified by gel electrophoresis and ligated into linearized pCR3.1. Ligation reactions were incubated overnight at 15 °C and added to competent *E. coli* strain DH5α for transformation. Bacterial colonies were grown in LB broth containing ampicillin at a final concentration of 50 µg/ml. Plasmid DNA was extracted using Wizard[®] Plus SV Minipreps DNA Purification System (Promega). Site-directed plasmid mutagenesis was carried out using a megaprimer approach [Sarkar and Sommer, 1990]; mutagenic primers are shown in the last column of Supp. Table S3. All wild-type and mutated clones were sequenced to confirm each substitution and exclude undesired mutations.

Cell culture and transfections

Human embryonic kidney (HEK) 293 cells, HeLa cells and breast cancer cell lines MCF7 and MDA-MB-231 were grown in DMEM supplemented with 10% fetal calf serum and 2 mM L-glutamine in 6-well plates. One µg of each plasmid was transfected into 3-5x10⁵ cells per well in the serum-free medium containing Lipofectamine (Invitrogen) according to manufacturer's recommendations. Cells were grown for a total of 48 hrs post-transfection before harvesting for RNA extraction.

For RNA interference (RNAi)-mediated depletion, the HEK293 and HeLa cells were plated in 12-well plates at a density of 8×10^4 per well. The next day, the indicated siRNA duplexes were combined with the Opti-MEM medium (Invitrogen) and HiPerfect (Qiagen). The mixtures were incubated at room temperature for 20 min before adding to the wells. Twenty-four hours later, the cells received the second hit with the same amount of siRNA. Forty-eight hours after the first hit, the cells were transfected with splicing reporter constructs and harvested 24 hours later. Sequences of siRNAs were as recently published [Kralovicova and Vorechovsky, 2010]. For SF2/ASF, we have also used the siRNA duplex as published previously [Blaustein et al., 2005].

Analysis of spliced products and Western blotting

Total RNA was extracted using the Qiagen RNeasy Plus kit (Qiagen) according to the manufacturer’s instructions. One μ g of total RNA was reverse transcribed with the Reverse Transcription System (Promega) and the first strand of cDNA was amplified with vector primers PL3 and PL4 as described [Kralovicova et al., 2004]. **Endogenous transcripts were amplified using primers B1ex5F (5’-gca tgc tga aac ttc tca acc) and B1ex7R (5’-cac tct gta gaa gtc ttt tgg cac).** PCR products were separated on agarose or polyacrylamide gels as indicated. Signal intensity was measured from at least two transfection experiments using tools described previously [Lei and Vorechovsky, 2005]. Western blot analysis was carried out with the indicated antibodies as published [Kralovicova and Vorechovsky, 2010].

EX-SKIP and HOT-SKIP construction

Both tools were implemented as a common gateway interface script written in Perl. Each utility performs a local search of the user's exonic sequence for putative ESSs/ESEs (PESSs/PESEs) [Zhang and Chasin, 2004], FAS-ESSs [Wang et al., 2004], RESCUE-ESEs [Fairbrother et al., 2002; Fairbrother et al., 2004], neighbourhood inference (NI) [Stadler et al., 2006] and exon/intron identity elements (EIE/IEs) [Zhang et al., 2008]. The HOT-SKIP algorithm generates all possible substitutions at each exonic position, extracts 15-bp and 11-bp oligomers with the substituted base at the central position and computes the appropriate ESS/ESE profile for each oligomer. The profile is presented in a table that can be sorted online according to a chosen element or their counts/ratios. Both tools are available at <http://www.dbass.org.uk> or at <http://ex-skip.img.cas.cz/> (EX-SKIP) or <http://hot-skip.img.cas.cz/> (HOT-SKIP).

The performance of HOT-SKIP was tested with a set of 37 previously reported mutations that resulted in exon skipping and genetic disease and with a similar number of control mutations that did not lower exon inclusion (Supp. Table S4). Each mutation was within the exon but outside the splice site consensus sequences, ie. the first and the last three exon positions. Set 2 consisted of *CFTR* exon 12 inclusion levels measured after transfection of the wild-type and 42 mutated minigenes, as published previously [Pagani et al., 2005; Pagani et al., 2003] (Supp. Table S5).

RESULTS

Identification of a silent mutation in *BRCAl* that induced exon skipping

A search of the BIC database for disease-associated exonic substitutions identified a total of 487 alterations. For further investigation, we selected 38 non-synonymous and 2

synonymous substitutions in *BRCA1* exons 5, 6, 10 and 18 because they were flanked by relatively short introns and their length was below the average (78, 89, 77 and 78 bp, respectively), thus facilitating DNA manipulation (Supp. Table S3).

Splicing of the wild-type reporter construct containing exons 5-7 produced a mixture of correctly spliced products and transcripts with skipped exon 6 in HEK293 cells (Fig. 1A-C). Examination of reporters mutated at each exonic variant revealed that the c.231G>T mutation significantly increased exon skipping as compared to the wild-type construct, while the remaining exon 6 variants did not produce any splicing defects (Fig. 1A-C). A similar level of exon 6 skipping was found in other cell types, including breast cancer cell lines MDA-MB-231 and MCF7 and HeLa cells (Fig. 1D). The exclusion of this exon from exogenous pre-mRNAs was negligible (Fig. 1C) and we did not observe exon 6 skipping in endogenous transcripts in any cell line using primers B1ex5F and B1ex7R located in flanking exons.

In contrast to the wild-type minigene 5-7, reporter construct 17-19 did not induce exon skipping (Supp. Figure S1). As the second intron of minigene 17-19 was partially retained, we decided to mutate this construct at each variant. Although none of the mutants generated significant exon skipping, mutation c.5138T>C improved splicing of the second intron, perhaps as a result of removal of a predicted ESS octamer AAATGGGT [Zhang and Chasin, 2004] and overlapping FAS-ESS hexamers TAGTTA, GTAGTT, GG TAGT and GGGTAG [Wang et al., 2004] from the wildtype construct. Finally, minigenes containing exons 5 and 19 as middle exons were either not informative or their cloning was unsuccessful.

Taken together, the introduction of 21 naturally occurring *BRCA1* variants in middle exons of two non-overlapping splicing reporter constructs revealed a single (1/21, ~5%) case of exon skipping resulting from the mutation c.231G>T. As minigene assays generally recapitulate *in vivo* splicing patterns of ubiquitously expressed pre-mRNAs, this mutation is likely to diminish expression of canonical *BRCA1* transcripts *in vivo*, although the patient RNA was not available to firmly establish its pathogenic character.

Characterization of a silencer element in *BRCA1* exon 6

The c.231G>T mutation may induce exon 6 skipping by creating a splicing silencer or eliminate an exonic enhancer. A systematic search for splicing regulatory elements revealed that this mutation increased the predicted ESS/ESE ratio about 70-fold (Table 1). To identify residues that control exon 6 inclusion in this region and to systematically examine contribution of each nucleotide, we mutated each position of a 12-nt region surrounding this mutation. Interestingly, measurements of exon inclusion levels of the wild-type and 36 mutated constructs showed that the c.231G>T transversion gave the highest percentage of transcripts lacking exon 6 (Fig. 2A). This mutation creates a TAG trinucleotide, which are highly enriched in splicing silencers [Wang et al., 2004; Vorechovsky, 2010]. In addition, exon skipping/inclusion levels across the mutated region exhibited a pattern of regular fluctuations that were only sporadically interrupted by more erratic changes, such as 224T or 231T, suggesting that the outcome of these substitutions can be predicted. As expected, the average level of exon skipping was highest for alterations that introduced uracil in the pre-mRNA (Fig. 2B). This nucleotide is overrepresented in introns and in splicing silencers [Wang et al., 2004; Vorechovsky,

2010], including FAS-ESSs [Wang et al., 2004] and PESSs [Zhang and Chasin, 2004], but N>T substitutions in plasmid reporter constructs did not increase exon skipping at each position (Fig. 2A).

Correlation of the ESS/ESE profile and exon inclusion

To determine which of the previously determined ESSs or ESEs correlate best with exon inclusion levels, the number of RESCUE-ESEs [Fairbrother et al., 2002; Fairbrother et al., 2004], FASS-ESSs [Wang et al., 2004], PESEs/PESSs [Zhang and Chasin, 2004], NI elements [Stadler et al., 2006] and EIE/IIEs [Zhang et al., 2008] was calculated for each segment. In addition, we employed the ESEfinder [Cartegni et al., 2003; Smith et al., 2006] to correlate exon inclusion levels with predicted scores of putative ESEs for serine/arginine-rich (SR) proteins that may interact with splicing regulatory sequences in exons [Sanford et al., 2005; Zhong et al., 2009].

For the 12-nt segment of exon 6, the highest correlation with enhancers was observed for the scores of the improved SF2/ASF consensus (Fig. 2C, Supp. Tables S6 and S7). Altogether, the ESEfinder predictions identified a total of 669 potential ESE motifs, with 34 potential ESEs within exons 5, 6, 10 and 18 (Supp. Table S7). However, putative ESEs for the remaining SR proteins did not significantly correlate with exon inclusion. A significant correlation was also found for EIEs ($r=0.43$), followed by a simple sum of all enhancers ($r=0.37$, Supp. Table S8). Conversely, the highest correlation between exon skipping and silencers was found for counts of FAS-ESSs (hex2 subset) ($r=0.27$), followed by trusted NI ESSs ($r=0.25$; Supp. Table S8).

A role for RNA secondary structure in exon skipping?

Of 36 minigene substitutions, the c.231G>T mutation gave the highest levels of exon skipping that were not matched by any other point mutation in the 12-nt segment, which is likely to contain short regulatory motifs engaged in intra- or inter-molecular contacts. This exclusivity could be due to a loss of a purine tetramer from the predicted terminal loop that might become less accessible to a ligand (Fig. 3A). Loss of the single-stranded configuration of this motif was consistent with the drop of PU (probability of unpaired) values [Hiller et al., 2007] from 0.8 for guanine to 0.07 for uracil. In an attempt to address this hypothesis, we introduced mutation c.238A>C in both the wild-type and the c.231G>T constructs. In the wild-type minigene, cytosine would base-pair with guanine, reducing the PU value to less than 0.01; in contrast, the PU value would increase to 0.49 in the double mutant c.231G>T/c.238A>C. However, transient transfections showed that exon inclusion levels for these pre-mRNAs were higher than for the wild-type reporter construct (Fig. 3B), which makes such a simple explanation unlikely. Prediction of putative binding sequences using currently available online tools [Cartegni et al., 2003; Piva et al., 2009] has identified SR proteins SRp40 and SC35 as candidate factors (data not shown), perhaps suggesting that the c.231G>T mutation could primarily alter protein binding, probably involving uracil in the predicted stem.

SR proteins in *BRCA1* exon 6 recognition

To determine the importance of *trans*-acting factors in exon 6 recognition and identify candidate ligands that interact with the novel silencer, we transiently transfected both wild-type and mutated reporter constructs into HeLa and HEK293 cell lines

individually depleted of 13 proteins known to influence splicing. The analysis of spliced products from the mutated reporter revealed an increased exon 6 skipping in cells lacking 9G8 and Tra2 (Fig. 4A). In contrast, exon skipping was reduced in cells depleted of hnRNP A1/A2 and SC35, while only minor increase of exon inclusion was observed in cells depleted of SRp40 (Supp. Figure S2). Conversely, overexpression of hnRNP A1 and SC35 increased exon 6 skipping, whereas overexpressed 9G8 and Tra2 enhanced exon 6 inclusion (Fig. 4B and data not shown).

Although downregulation of SF2/ASF was more efficient in HeLa cells than HEK293 cells (Fig. 4C), we did not see an increase of exon skipping in depleted cells, as predicted by the ESEfinder for the 12-nt segment (Fig. 2C), nor a decrease upon overexpression of this multifunctional protein, both in the wild-type and mutant *BRCA1* reporters (Fig. 4B). In fact, overexpression of SF2/ASF resulted in increased exon skipping for each reporter construct (Fig. 4B).

Taken together, hnRNP A1 and SR proteins 9G8, SC35, SF2/ASF and Tra2 were associated with significant exon 6 skipping/inclusion and are likely to regulate splicing this exon *in vivo*.

HOT-SKIP and EX-SKIP: online tools predicting exon skipping mutations

As combination of independently derived splicing regulatory motifs provided a better predictor of the c.231G>T mutation than individual elements (Table 1), we have developed two online tools, termed EX-SKIP and HOT-SKIP, that compute the ESS/ESE profile using a single click of the mouse and, as we show below, facilitate identification of exon-skipping alterations.

EX-SKIP is a simple utility that compares the ESE/ESS profile of two or more short sequences, typically a wild-type and a mutated allele, to quickly determine which exonic variant has a higher chance to skip this exon. In contrast, HOT-SKIP systematically examines all possible substitutions at each exonic position, computes the ESS/ESE profile for each alteration and identifies ‘hot’ positions and substitutions that are most likely to skip this exon.

The input to both utilities is an exonic sequence (strictly in upper case) in the FASTA format flanked by short intronic sequences (in lower case) on either side to account for adjacent auxiliary splicing motifs. Multiple FASTA format sequences are permitted up to a total length of 4,000 bp.

The output of EX-SKIP is a simple table showing the number of each ESSs/ESEs and their sum/ratio. The table is followed by a short statement which of the submitted allele is more likely to skip the exon. By contrast, the HOT-SKIP output is a large table listing all possible point mutations at each exon position. The user can sort the ESE/ESS data online by clicking on chosen elements or their sum/ratio in table headers. Substitutions with the highest values of ESSs and the lowest values of ESEs are likely to represent ‘hot-spots’ for exon skipping. The corresponding positions are highlighted in red in the exonic sequence, with the intensity of red color reflecting the values of ESS/ESE ratios. Thus, HOT-SKIP considers all possible exonic substitutions, except for changes in the first and the last three positions of the exon, ie. outside splice-site consensus signals. As an example, the ESS/ESE profile was calculated for each *BRCA1* exon (Supp. Table S9).

Prediction of exon skipping mutations by HOT-SKIP

To illustrate the utility of this tool, we computed the ESS/ESE ratio for all possible point mutations at each position of 33 exons that sustained a total of 37 alterations, each resulting in exon skipping **observed in** genetic diseases (Supp. Table S4). Mutations were ordered according to the ESS/ESE ratio for each exon. A subset of mutations that lacked any predicted ESSs was sorted according to the sum of ESEs. We found that 20/37 (54%) of these mutations were in the first HOT-SKIP quartile, while 27/37 (73%) were in the first two quartiles (Fig. 5A). The bias in their distribution toward top HOT-SKIP positions was statistically significant ($P=0.0001$ and $P<0.003$, respectively, binomial tests). Interestingly, several of the remaining 10 mutations that were in the third and fourth quartiles were located in the vicinity of top-ranking positions. For example, as many as 10 mutations located up to 10 nt downstream of the disease-associated G>A substitution at position 37 of *ADA* exon 7 [Ozsahin et al., 1997] were ranked among the first 12 positions with the highest ESS/ESE ratios. Similarly, mutations introducing guanine at position 195-197 of the *BRCA2* exon 18 were ranked by HOT-SKIP among the top 5 exon-skipping mutations, just 6 nt from the disease-causing substitution at position 189 [Fackenthal et al., 2002]. Exonic positions that were mutated in cryptic exons remained in the top quartile if mutated to other nucleotides [Dear et al., 2006; Ishii et al., 2002].

These observations suggest that in a subset of these cases, the ESS/ESE profile did not correctly predict mutations that were likely to reside within a shared secondary structure. This leads to the prediction that the ESS/ESE profile may be more useful for less structured pre-mRNAs. In addition, HOT-SKIP may not perform well for mutations

very close to natural splice sites. For example, two of the 10 ‘non-performing’ mutations, located at position 7 of *MLH1* exon 3 [McVety et al., 2006] and at position 5 of *GHI* exon 3 [Ryther et al., 2003]), were very close to 3’ splice sites, and the 3’YAG splice-site consensus is likely to interfere with accurate prediction of exonic auxiliary sequences. In addition, these mutations were in exonic segments recognized by the auxiliary factor of the U2 small nuclear ribonucleoprotein. The small subunit of this factor may bind up to 12 nt of the exon and has a preference for uracil [Wu et al., 1999].

Because formation of cross-exon spliceosomal network is likely to be less complete for longer exons, we plotted exon length against the HOT-SKIP rank, but no correlation was found ($r=0.027$, $P>0.05$; Fig. 5A).

In contrast to exon-skipping substitutions, a similar number of exonic mutations that did not lower exon inclusion (Supp. Table S4), revealed the opposite trend. Only less than 9% mutations were in the first HOT-SKIP quartile ($p<0.01$), while 18/46 (39%) were in the first two quartiles (Fig. 5B).

Finally, we compared previously published exon inclusion data for 42 mutations in *CFTR* exon 12 that were determined *ex vivo* [Pagani et al., 2005; Pagani et al., 2003]. We found a significant negative correlation of exon 12 inclusion with the ESS/ESE ratio ($r=-0.28$, $P=0.03$) as well as for the total number of ESSs (-0.35), but not for ESEs or any of the remaining individual elements, except for trusted NI ESSs that gave the most significant correlation ($r=-0.40$, $P=0.004$). Thus, these data suggest that the best prediction method may vary from exon to exon and highlight the versatility of HOT-SKIP to address this in future studies.

DISCUSSION

This work has shown that (i) exonic point mutations that result in substantive exon skipping are infrequent in *BRCA1* exons 6 and 18; (ii) prediction of exon-skipping mutations is facilitated by the ESS/ESS profile that takes into account independently derived elements and can now be easily determined online; (iii) exon 6 inclusion levels in the *BRCA1* mRNA were not irregular but showed an undulating pattern across a systematically mutated segment that contained the newly identified ESS; and (iv) modification of intracellular levels of hnRNP A1 and SR proteins 9G8, Tra2 and SC35 was associated with exon 6 skipping/inclusion, with each protein giving the opposite ratios of exon-lacking/exon-containing transcripts in depleted/overexpressing cell lines.

We believe that the first conclusion is not limited to the two *BRCA1* exons studied here. Although nonsense, missense and translationally silent mutations can inactivate genes by skipping the mutant exons, most mutations that initially exemplified this point were in fact in the splice-site consensus (Table 1 in [Cartegni et al., 2002]). Second, the low frequency of such mutations has been supported by independent screens of candidate exonic variants in other genes. For example, *BRCA2* substitutions deposited in BIC and screened by RT-PCR and minigene assays did not show a single example of aberrant splicing [Whiley et al., 2010]. On the other hand, more systematic studies [Sanz et al., 2010] confirmed previous findings [Pros et al., 2008; Teraoka et al., 1999] that in large genes with many introns up to a half of all disease gene mutations affected splicing, although only a few were in exons. For example, 2 of four exonic unclassified variants in *BRCA2* showed aberrant splicing [Bonnet et al., 2008]. In *BRCA1*, only several exon-skipping point mutations have been reported in patients with breast cancer, including a

nonsense mutation c.5199G>T and missense variant c.5242C>A in exon 18 [Mazoyer et al., 1998; Millevoi et al., 2010] and missense mutations c.5434C>G and c.5453A>G in exon 23 [Gaildrat et al. 2010; Rouleau et al. 2010], each pointing to exonic splicing regulatory sequences. Thus, although the overall fraction of substitutions that result in low exon inclusion levels is likely to be relatively small, they should always be considered, particularly in exons weakly included in mRNAs, as exemplified by the *CFTR* exon 12 where almost a third of all point mutations induced exon skipping in a minigene system, ranging between full inclusion and full exclusion [Pagani et al., 2005] (Supp. Table S5). In contrast, our systematic minigene substitutions around c.231G>T altered exon inclusion levels within a rather limited range of 39-86%, revealing a rather regular pattern of fluctuations (Fig. 2A). Disregarding the effect of the c.231G>T mutation, these changes were translated only to ~1.6-fold difference between maximum and minimum exon skipping.

Apart from canonical splicing signals, the extent to which exonic substitutions lower exon inclusion levels can be influenced by more distant intronic sequences that may form intra-molecular contacts with the exon. For example, we observed no exon skipping for the mutation c.5242C>A in our 17-19 construct (Supplementary Fig. S1), which can be attributed to a deletion of an intron 17 segment that may repress exon 18 inclusion in the mRNA. The observed small variations in intron 18 retention may in fact signify variable exon inclusion levels that are not detected by the minigene with the shortened intron, but may be observed *in vivo*. Thus, the variability in exon inclusion should be always interpreted in the context of adjacent segments, including flanking introns and exons and the strength of their splice sites and splicing regulatory signals.

Before concluding pathogenicity, the results of minigene assays should be complemented with the analysis of spliced products in fresh RNA samples extracted from affected individuals and appropriate controls.

Remarkably, the ESE/ESS profile correctly predicted a putative exon-skipping mutation over a considerable distance from its exonic position (Table 1) and suggested that a combination of independently derived elements had a better predictive and perhaps practical value. This observation is exploited by new online tools that are publicly available and should facilitate detection of DNA variants that alter splicing in conjunction with utilities that assess the strength of natural splice sites [Senapathy et al., 1990; Yeo and Burge, 2004] and branch points [Kol et al., 2005]. Recently developed web servers that distinguish exon skipping and cryptic splice site activation *ab initio* [Divina et al., 2009] and that integrate numerous methods for estimating the intrinsic strength of splice sites and for defining splicing regulatory sequences and their change upon mutation [Schwartz et al., 2009] represent just the starting point toward more accurate and sensitive algorithms in the future.

Our screen of splicing factors known to influence exon inclusion has identified candidates that may bind to the newly identified ESS. Although they may exert this influence by contacting the RNA, this will require *in vitro* evidence to exclude alternative explanations. Likewise, the opposite effects of individual SR proteins on exon inclusion (Fig. 4) are consistent with the notion that these multifunctional factors act both as splicing activators and inhibitors [Kanopka et al., 1996]. Finally, the discrepancy between the prediction that SF2/ASF would act as an activator, based on ~13% of the total exonic sequence (Fig. 2C), and the experimentally determined inclusion levels in cells with

modified SF2/ASF levels (Fig. 4A-C) clearly point to additional interactions involved in the SF2/ASF-mediated regulation of this exon.

Taken together, screening of 21 unclassified exonic variants in *BRCA1* has identified a single case of exon skipping. Systematic mutagenesis surrounding the new ESS in *BRCA1* exon 6 revealed only modest variations in exon inclusion for the majority of changes, supporting the view that exonic point mutations that result in substantial exon skipping are relatively infrequent. The newly developed online tools should facilitate detection of exonic alterations that reduce exon inclusion in mRNAs.

ACKNOWLEDGEMENTS

This study was supported by grants from the EURASNET, Action Medical Research, CRUK and the JDRF International.

REFERENCES

Ars E, Serra E, Garcia J, Kruyer H, Gaona A, Lazaro C, Estivill X. 2000. Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum Mol Genet* 9:237-247.

Auclair J, Busine MP, Navarro C, Ruano E, Montmain G, Desseigne F, Saurin JC, Lasset C, Bonadona V, Giraud S and others. 2006. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. *Hum Mutat* 27:145-154.

Blaustein M, Pelisch F, Tanos T, Munoz MJ, Wengier D, Quadrana L, Sanford JR, Muschietti JP, Kornblihtt AR, Caceres JF and others. 2005. Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol* 12:1037-1044.

Bonnet C, Krieger S, Vezain M, Rousselin A, Tournier I, Martins A, Berthet P, Chevrier A, Dugast C, Layet V and others. 2008. Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *J Med Genet* 45:438-446.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3:285-298.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31:3568-3571.

Cooper DN, Krawczak M. 1993. *Human Gene Mutation*. Oxford: BIOS Scientific Publishers.

Cooper TA, Mattox W. 1997. The regulation of splice-site selection, and its role in human disease. *Am J Hum Genet* 61:259-266.

Cooper TA, Wan L, Dreyfuss G. 2009. RNA and disease. *Cell* 136:777-793.

Dear A, Daly J, Brennan SO, Tuckfield A, George PM. 2006. An intronic mutation within *FGB* (IVS1+2076 a>g) is associated with afibrinogenemia and recurrent transient ischemic attacks. *J Thromb Haemost* 4:471-472.

Divina P, Kvitkovicova A, Vorechovsky I. 2009. *Ab initio* prediction of cryptic splice-site activation and exon skipping. *Eur J Hum Genet* 17:759-756.

Fackenthal JD, Cartegni L, Krainer AR, Olopade OI. 2002. BRCA2 T2722R is a deleterious allele that causes exon skipping. *Am J Hum Genet* 71:625-631.

Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* 297:1007-1013.

Fairbrother WG, Yeo GW, Yeh R, Goldstein P, Mawson M, Sharp PA, Burge CB. 2004. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res* 32:W187-190.

Gaildrat P, Krieger S, Thery JC, Killian A, Rousselin A, Berthet P, Frebourg T, Hardouin A, Martins A, Tosi M. 2010. The BRCA1 c.5434C->G (p.Pro1812Ala) variant induces a deleterious exon 23 skipping by affecting exonic splicing regulatory elements. *J Med Genet* 47:398-403.

- Goren A, Ram O, Amit M, Keren H, Lev-Maor G, Vig I, Pupko T, Ast G. 2006. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol Cell* 22:769-781.
- Gowen LC, Avrutskaya AV, Latour AM, Koller BH, Leadon SA. 1998. BRCA1 required for transcription-coupled repair of oxidative DNA damage. *Science* 281:1009-1012.
- Hiller M, Zhang Z, Backofen R, Stamm S. 2007. Pre-mRNA Secondary Structures Influence Exon Recognition. *PLoS Genet* 3:e204.
- Ishii S, Nakao S, Minamikawa-Tachino R, Desnick RJ, Fan JQ. 2002. Alternative splicing in the alpha-galactosidase A gene: increased exon inclusion results in the Fabry cardiac phenotype. *Am J Hum Genet* 70:994-1002.
- Kanopka A, Muhlemann O, Akusjarvi G. 1996. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* 381:535-538.
- King MC, Marks JH, Mandell JB. 2003. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* 302:643-646.
- Kol G, Lev-Maor G, Ast G. 2005. Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. *Hum Mol Genet* 14:1559-1568.
- Kralovicova J, HOUNGNINOU-MOLANGO S, KRAMER A, VORECHOVSKY I. 2004. Branch sites haplotypes that control alternative splicing. *Hum Mol Genet* 13:3189-3202.
- Kralovicova J, Vorechovsky I. 2007. Global control of aberrant splice site activation by auxiliary splicing sequences: evidence for a gradient in exon and intron definition. *Nucleic Acids Res* 35:6399-6413.
- Kralovicova J, Vorechovsky I. 2010. Allele-dependent recognition of the 3' splice site of *INS* intron 1. *Hum Genet* 128:383-400.
- Lastella P, Surdo NC, Resta N, Guanti G, Stella A. 2006. In silico and in vivo splicing analysis of MLH1 and MSH2 missense mutations shows exon- and tissue-specific effects. *BMC Genomics* 7:243.
- Lei H, Vorechovsky I. 2005. Identification of splicing silencers and enhancers in sense *Alus*: a role for pseudo-acceptors in splice site repression. *Mol Cell Biol* 25:6912-6920.
- Liu HX, Cartegni L, Zhang MQ, Krainer AR. 2001. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat Genet* 27:55-58.
- Liu HX, Chew SL, Cartegni L, Zhang MQ, Krainer AR. 2000. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol* 20:1063-1071.
- Lorson CL, Androphy EJ. 2000. An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. *Hum Mol Genet* 9:259-265.
- Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* 101:7287-7292.
- Mazoyer S, Puget N, Perrin-Vidoz L, Lynch HT, Serova-Sinilnikova OM, Lenoir GM. 1998. A BRCA1 nonsense mutation causes exon skipping. *Am J Hum Genet* 62:713-715.

McVety S, Li L, Gordon PH, Chong G, Foulkes WD. 2006. Disruption of an exon splicing enhancer in exon 3 of *MLH1* is the cause of HNPCC in a Quebec family. *J Med Genet* 43:153-156.

Millevoi S, Bernat S, Telly D, Fouque F, Gladieff L, Favre G, Vagner S, Toulas C. 2010. The c.5242C>A *BRCA1* missense variant induces exon skipping by increasing splicing repressors binding. *Breast Cancer Res Treat* 120:391-399.

Ozsahin H, Arredondo-Vega FX, Santisteban I, Fuhrer H, Tuchscheid P, Jochum W, Aguzzi A, Lederman HM, Fleischman A, Winkelstein JA and others. 1997. Adenosine deaminase deficiency in adults. *Blood* 89:2849-2855.

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 USA* 102:6368-6372.

Pagani F, Stuani C, Tzetis M, Kanavakis E, Efthymiadou A, Doudounakis S, Casals T, Baralle FE. 2003. New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in *CFTR* exon 12. *Hum Mol Genet* 12:1111-1120.

Piva F, Giulietti M, Nocchi L, Principato G. 2009. SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. *Bioinformatics* 25:1211-1213.

Pros E, Gomez C, Martin T, Fabregas P, Serra E, Lazaro C. 2008. Nature and mRNA effect of 282 different *NF1* point mutations: focus on splicing alterations. *Hum Mutat* 29:E173-E193.

Raponi M, Baralle FE, Pagani F. 2007. Reduced splicing efficiency induced by synonymous substitutions may generate a substrate for natural selection of new splicing isoforms: the case of *CFTR* exon 12. *Nucleic Acids Res* 35:606-613.

Rouleau E, Lefol C, Moncoutier V, Castera L, Houdayer C, Caputo S, Bieche I, Buisson M, Mazoyer S, Stoppa-Lyonnet D and others. 2010. A missense variant within *BRCA1* exon 23 causing exon skipping. *Cancer Genet Cytogenet* 202:144-146.

Ryther RC, McGuinness LM, Phillips JA, 3rd, Moseley CT, Magoulas CB, Robinson IC, Patton JG. 2003. Disruption of exon definition produces a dominant-negative growth hormone isoform that causes somatotroph death and IGHD II. *Hum Genet* 113:140-148.

Sanford JR, Ellis J, Caceres JF. 2005. Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochem Soc Trans* 33:443-446.

Sanz DJ, Acedo A, Infante M, Duran M, Perez-Cabornero L, Esteban-Cardenosa E, Lastra E, Pagani F, Miner C, Velasco EA. 2010. A high proportion of DNA variants of *BRCA1* and *BRCA2* is associated with aberrant splicing in breast/ovarian cancer patients. *Clin Cancer Res* 16:1957-1967.

Sarkar G, Sommer SS. 1990. The "megaprimer" method of site-directed mutagenesis. *BioTechniques* 8:404-407.

Schwartz S, Hall E, Ast G. 2009. SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Res* 37:W189-192.

Senapathy P, Shapiro MB, Harris NL. 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol* 183:252-278.

- Skoko N, Baralle M, Buratti E, Baralle FE. 2008. The pathological splicing mutation c.6792C>G in NF1 exon 37 causes a change of tenancy between antagonistic splicing factors. *FEBS Lett* 582:2231-2236.
- Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. 2006. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* 15:2490-2508.
- Spies N, Nielsen CB, Padgett RA, Burge CB. 2009. Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* 36:245-254.
- Stadler MB, Shomron N, Yeo GW, Schneider A, Xiao X, Burge CB. 2006. Inference of splicing regulatory activities by sequence neighborhood analysis. *PLoS Genet* 2:e191.
- Stella A, Wagner A, Shito K, Lipkin SM, Watson P, Guanti G, Lynch HT, Fodde R, Liu B. 2001. A nonsense mutation in MLH1 causes exon skipping in three unrelated HNPCC families. *Cancer Res* 61:7020-7024.
- Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN. 2009. The Human Gene Mutation Database: 2008 update. *Genome Med* 1:13.
- Teraoka SN, Telatar M, Becker-Catania S, Liang T, Onengut S, Tolun A, Chessa L, Sanal O, Bernatowska E, Gatti RA and others. 1999. Splicing defects in the ataxia-telangiectasia gene, *ATM*: underlying mutations and consequences. *Am J Hum Genet* 64:1617-1631.
- Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, Wang Q, Buisine MP, Soret J, Tazi J and others. 2008. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat* 29:1412-1424.
- Valentine CR. 1998. The association of nonsense codons with exon skipping. *Mutation Research-Reviews in Mutation Research* 411:87-117.
- Vorechovsky I. 2010. Transposable elements in disease-associated cryptic exons. *Hum Genet* 127:135-154.
- Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell* 119:831-845.
- Wang Z, Xiao X, Van Nostrand E, Burge CB. 2006. General and specific functions of exonic splicing silencers in splicing control. *Mol Cell* 23:61-70.
- Whiley PJ, Pettigrew CA, Brewster BL, Walker LC, Investigators K, Spurdle AB, Brown MA. 2010. Effect of BRCA2 sequence variants predicted to disrupt exonic splice enhancers on BRCA2 transcripts. *BMC Med Genet* 11:80.
- Wu S, Romfo CM, Nilsen TW, Green MR. 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 402:832-835.
- Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11:377-394.
- Zhang C, Li WH, Krainer AR, Zhang MQ. 2008. RNA landscape of evolution for optimal exon and intron discrimination. *Proc Natl Acad Sci USA* 105:5797-5802.
- Zhang XH, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* 18:1241-1250.
- Zhong X-Y, Wang P, Han J, Rosenfeld MG, Fu X-D. 2009. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell* 35:1-10.

FIGURE LEGENDS

Figure 1 Identification of the *BRCA1* missense mutation that induced exon skipping

Legend: **A:** 5-7 minigene reporter construct. Exons are shown as boxes, introns as lines. Aberrantly spliced products, including 2 minor cryptic 5' splice sites (cr5'ss) observed 22 nt upstream and 186 nt downstream of canonical 5' splice site, are shown as dotted lines. Repetitive elements in introns are schematically shown as grey (AluJb), white (3' end of L2) and black (MER104) rectangles; their orientation is denoted by arrowheads. Cloning primers are shown as grey arrows. **B:** exon 6 sequence (upper case) with flanking intronic sequences (lower case). Substitutions (bold) are shown above mutated residues (highlighted). c.231G>T substitution creates the TACTAG hexamer (underlined) predicted by FAS-ESS [Wang et al., 2006] and also the CTAGAT splicing regulatory motif (highlighted in grey) predicted by a computational method based on the conservation of wobble positions between human and mouse orthologous exons [Goren et al., 2006]. **C:** Splicing products of the *BRCA1* 5-7 minigene determined by reverse-transcription PCR and separated by PAGE. RNA products are schematically shown to the right. The use of minor cryptic 5'ss-22 as shown in panel A is denoted by a shorter exon 5. Minigene mutations (top) correspond to panel B. **D:** the splicing patterns of both the wild-type and mutant minigenes are maintained in other cell types. RNA products separated on agarose gels are shown to the right.

Figure 2 Systematic mutagenesis of a new splicing regulatory motif in *BRCA1* exon 6

Legend: **A:** exon 6 inclusion/skipping levels of the wild-type (WT) and 36 mutated minigenes. Designation of substitutions (horizontal axis) corresponds to the numbering shown in Fig. 1B. The BIC-derived transversion is boxed. Error bars represent standard deviations of two transfections experiments; **B:** The average exon skipping of X>G, X>T, X>C and X>A substitutions. Error bars represent standard deviations; **C:** A high correlation between the improved SF2/ASF ESE scores and exon inclusion. **Predicted SR ESEs were calculated using the ESEfinder (http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese_finder.cgi?process=home) as a sum of all scores at each position of the wild-type (WT) sequence and 36 mutated constructs. The scores for the original SF2/ASF matrix and other SR proteins are shown in Supp. Table S6.**

Figure 3 Predicted RNA secondary structure of the c. 231G/T alleles

Legend: **A:** predicted secondary structure of the wild type (WT) and mutated (positions 231 and 238) pre-mRNA. Mutations are denoted by black and grey arrows, respectively. Secondary structure predictions were carried out using the dynamic programming algorithm implemented in the RNAstructure program (v. 4.6) [Mathews et al., 2004]. The algorithm is based on free energy minimization with nearest neighbor parameters and includes free energy increments for coaxial stacking of helices when they are either adjacent or separated by a single mismatch [Mathews et al., 2004]. **B:** The splicing pattern of four mutated reporter constructs upon transient transfections in HEK293 cells. EI, exon inclusion.

Figure 4 *Trans-acting factors that influence recognition of BRCA1 exon 6*

Legend: **A:** RNAi-mediated depletion of selected SR-like proteins. SC, scrambled controls. Spliced products are shown to the right. The extended experiment with additional SR proteins and hnRNPs is shown in Supp. Figure S2. **B:** Splicing of the wild-type and mutated *BRCA1* reporter constructs in cells overexpressing a subset of SR proteins. **C:** Western blot analysis showing the extent of SF2/ASF RNAi-mediated depletion in transfected HEK293 and Hela cells. siRNAs sequences are denoted as K [Kralovicova and Vorechovsky, 2010] and N [Blaustein et al., 2005]. Final concentration of siRNAs in each well was 100 nM (50 nM each for a combination of N/K siRNAs). SC, scrambled controls.

Figure 5 *Prediction of exon-skipping substitutions by HOT-SKIP*

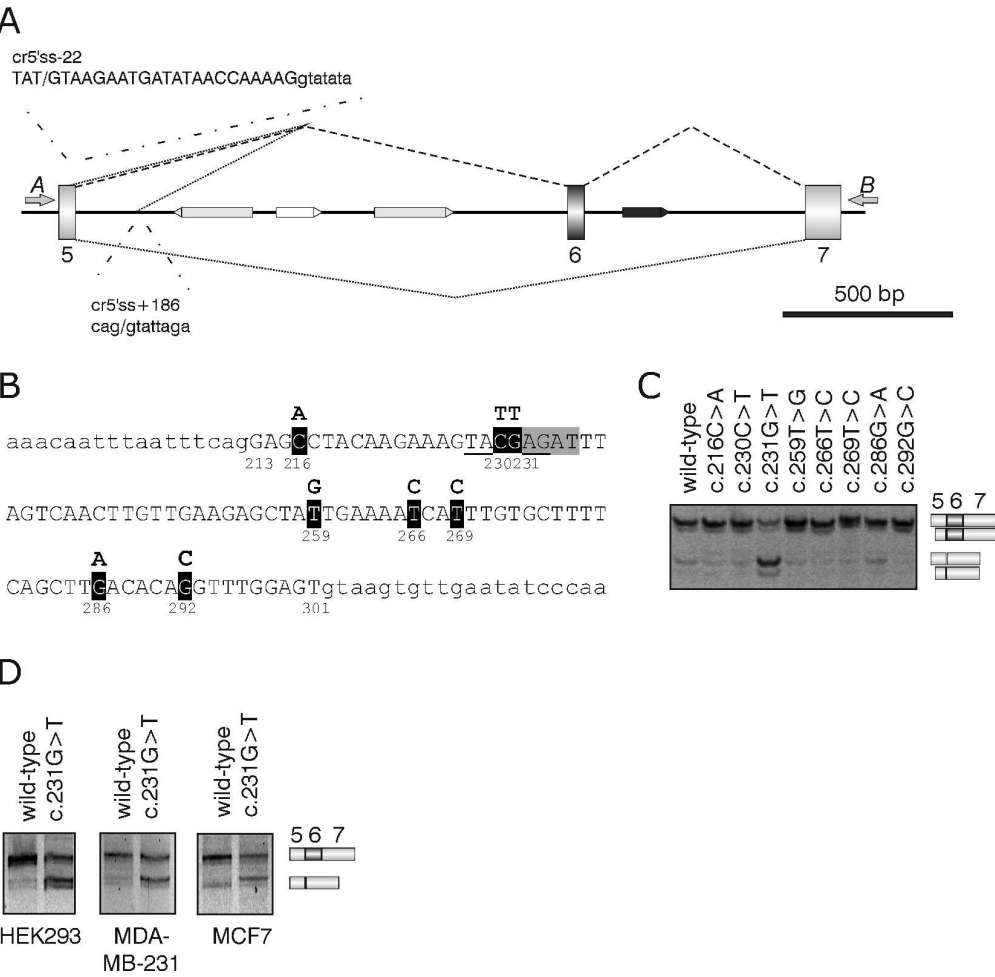
Legend: Each point represents a disease-causing mutation that resulted in exon skipping (A) or a mutation that did not alter splicing (B). Full details of these cases are shown in Supp. Table S4. HOT-SKIP percentiles were calculated as $(a-4)*400/n$, where n is exon length in nucleotides and a is the HOT-SKIP rank of the mutation among all possible exon substitutions, as determined by sorting the ESS/ESE ratios.

TABLES

Table 1 Counts of predicted ESSs and ESEs in the wild-type and c.231G>T alleles

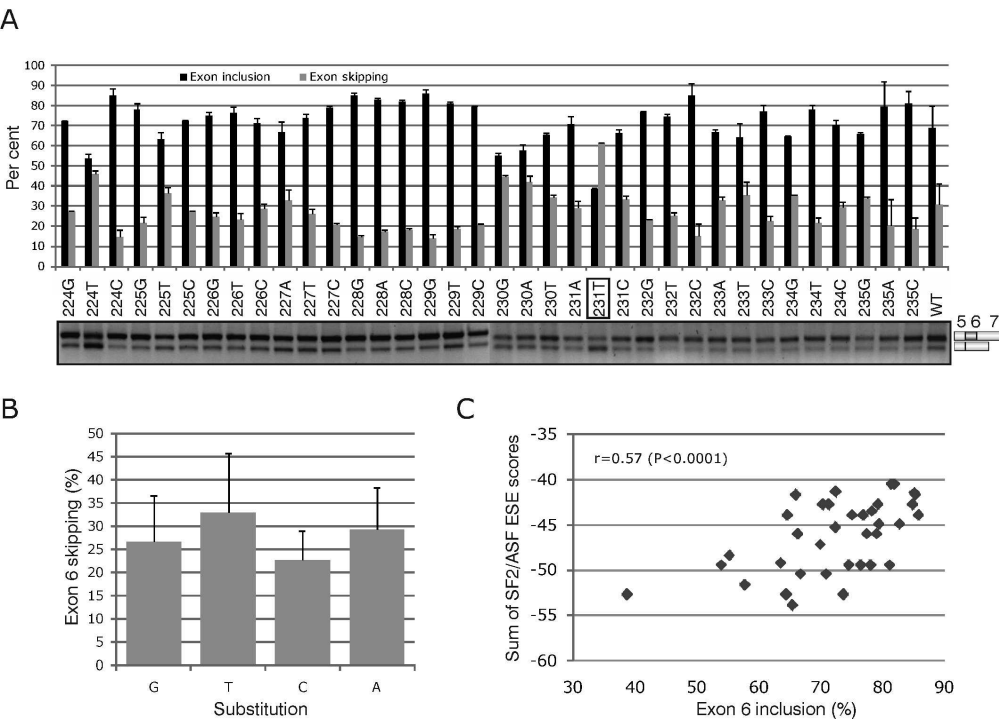
	Exon 6 segment	RESCUE-ESE	FAS-ESS	PESE	PESS	EIE	IIE	ESE total	ESS total	ESS/ESE
Wild type	AGTACG ₂₃₁ AGATT	0	0	13	0	6	1	19	1	0.053
Mutant	AGTACT ₂₃₁ AGATT	0	1	2	9	1	1	3	11	3.667

Fig. 1



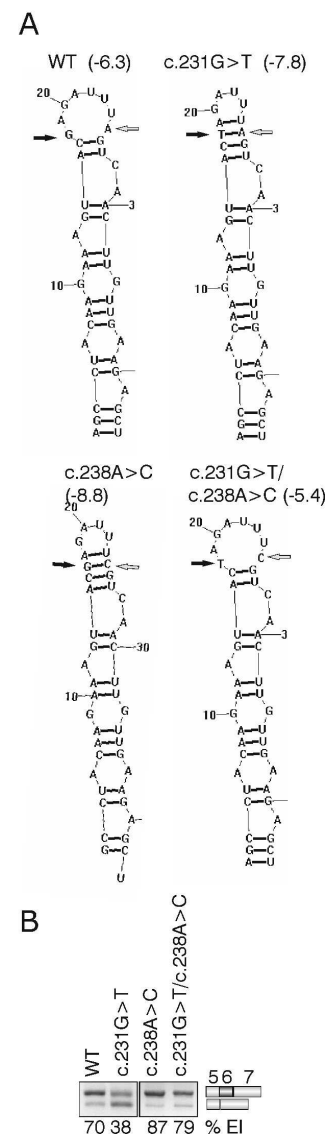
140x151mm (600 x 600 DPI)

Fig. 2



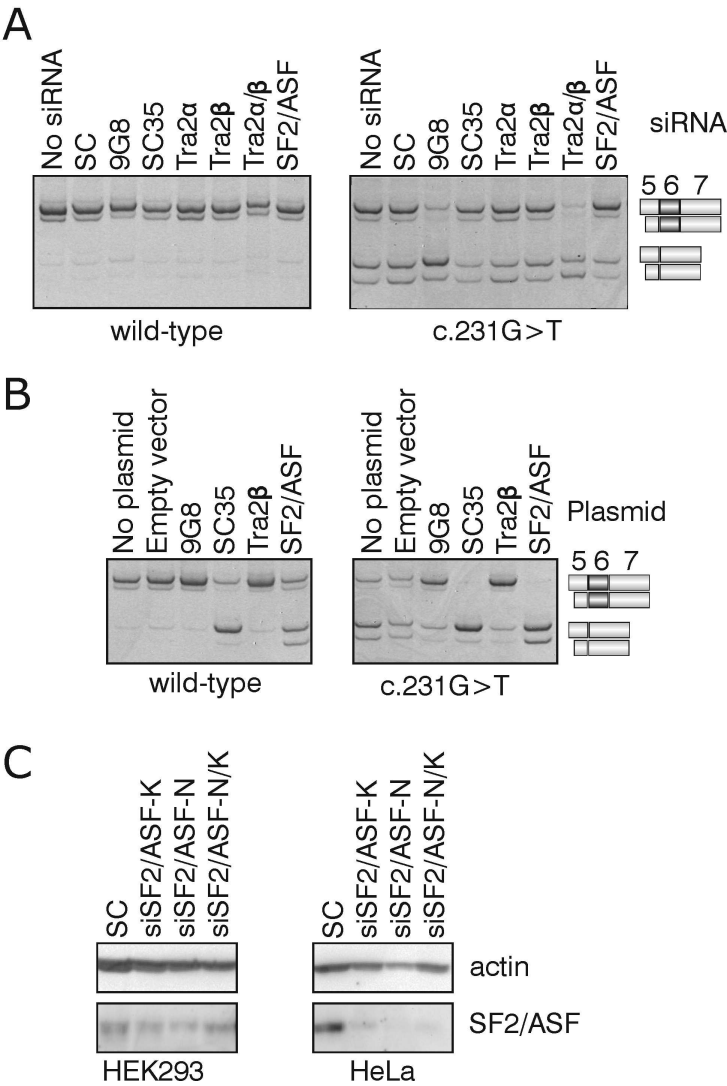
179x142mm (600 x 600 DPI)

Fig. 3



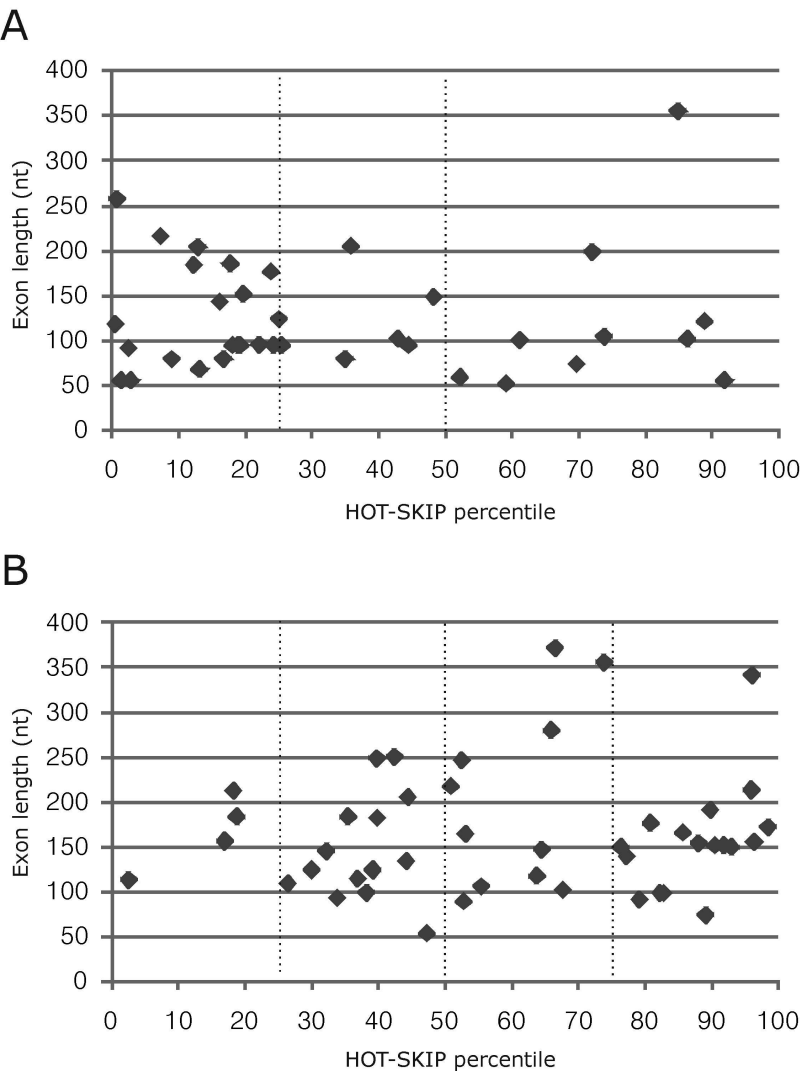
53x199mm (600 x 600 DPI)

Fig. 4



88x135mm (600 x 600 DPI)

Fig. 5



99x142mm (600 x 600 DPI)

SUPPLEMENTARY MATERIAL

**Prediction of single-nucleotide substitutions that result in exon skipping:
identification of a splicing silencer in *BRCA1* exon 6**

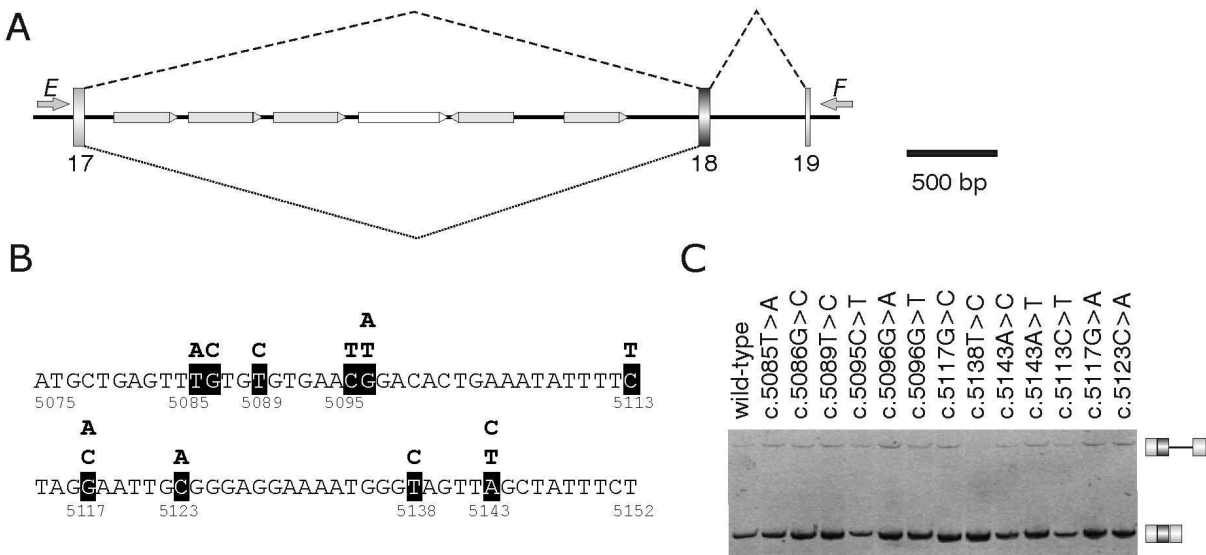
Michela Raponi^{1,*}, Jana Kralovicova^{1,*}, Ellen Copson^{1,*}, Petr Divina², Diana Eccles¹, Peter Johnson¹, Diana Baralle¹, Igor Vorechovsky¹

¹University of Southampton School of Medicine, Southampton, United Kingdom

²Institute of Molecular Genetics, Prague, Czech Republic

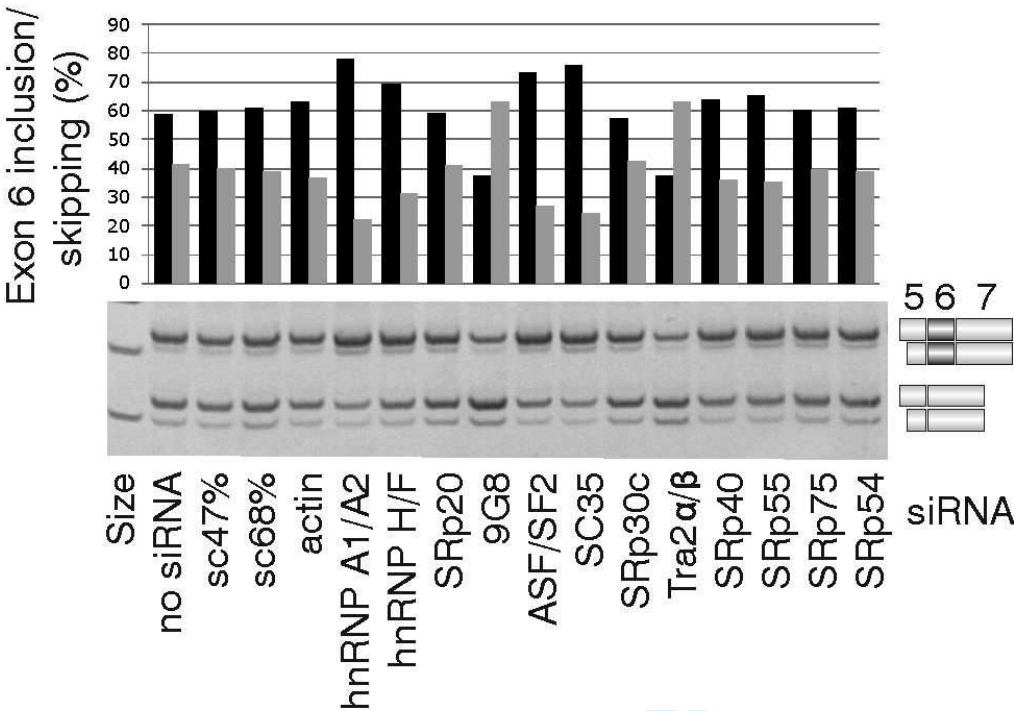
SUPPLEMENTARY FIGURES

Supplementary Figure S1 Splicing pattern of the 17-19 reporter construct



Legend: **A:** Exons are shown as numbered boxes, introns as lines. Normal and aberrantly spliced products are shown as dotted lines above and below the pre-mRNA, respectively. Repetitive elements in introns are schematically shown as grey (*Alu*) and white (L1) rectangles; their orientation is denoted by arrowheads. Cloning primers are shown as grey arrows. **B:** Exon 18 sequence and single-nucleotide substitutions introduced in mutant reporter constructs. **C:** RNA products of the 17-19 minigene (schematically shown to the right) visualized by reverse-transcription PCR and separated by PAGE. Minigene mutations shown at the top correspond to positions in panel **B**.

Supplementary Figure S2 RNAi-mediated screen of *trans*-acting factors involved in exon 6 recognition



Legend: sc47%, sc68%, scrambled controls with the indicated GC content. Spliced products are schematically shown to the right, measurements of two fragments with the strongest signal intensity (exon 6 inclusion and skipping) is at the top.

SUPPLEMENTARY TABLES

Supplementary Table 1 Exon-by-exon breakdown of *BRCA1* missense/silent mutations in BIC

Exon ¹	Exon size (bp)	Size of 5' intron (bp)	Size of 3' intron (bp)	Number of missense/silent changes per exon	Number of missense/silent changes per bp
1	378		621		
2	99	621	8237	13	0.131
3	54	8237	9192	13	0.241
5	78	9192	1499	15	0.192
6	89	1499	606	7	0.079
7	140	606	4241	8	0.057
8	106	4241	2485	8	0.075
9	46	2485	1321	3	0.065
10	77	1321	985	4	0.052
11	3426	985	402	250	0.073
12	89	402	8638	5	0.056
13	172	8638	5789	14	0.008
14	127	5789	1966	8	0.063
15	191	1966	3092	15	0.079
16	311	3092	3232	33	0.106
17	88	3232	3656	14	0.159
18	78	3656	500	15	0.192
19	41	500	6197	5	0.122
20	84	6197	5934	15	0.179
21	55	5934	1868	8	0.145
22	74	1868	1417	8	0.108
23	61	1417	1840	11	0.180
24	1502	1840		15	0.010

Supplementary Table 2 Primers for cloning *BRCA1* minigene reporter constructs

Minigene	Forward primer name and sequence (5'-3')	Reverse primer name and sequence (5'-3')
5-7	A gata <u>ggatcc</u> atgctgaaacttctcaaccag	B gata <u>tctagaggaaggatttcgggttcact</u>
	E gataa <u>agc</u> ttttaatttcagatgctcgtga	F gata <u>ctcgagatctctggtagtttgaaca</u>

Legend: restriction sites are underlined.

Supplementary Table 3 Summary of missense/silent mutations in *BRCA1* exons 5, 6, 10 and 18

Exon	Mutation ¹	Predicted amino-acid	Number of reports	Source	Mutagenic primer ²
5	c.135A>T	K45N	3	Myriad	ttataatttatagTTTTTCATGCTG
5	c.139T>G	C47G	1	Casey	atttatagATTTCGCATGCTGAAAC
5	c.140G>T	C47F	2	Myriad	tttatagATTTCATGCTGAAAC
5	c.154C>T	L52F	5	Myriad, Casey	ATGCTGAAACTTTTCAACCAGAAG
5	c.181T>G	C61G	222	Myriad, others	GGGCCTTCACAGGTCCTTTATGTA
5	c.181T>C	C61R	1	Bond	GGGCCTTCACAGGTCCTTTATGTA
5	c.182G>A	C61Y	6	Myriad	GCCTTCACAGTATCCTTTATGTA
5	c.189A>T	L63F	1	Ostrow	AGTGTCTTTTGTGAAGAATGA
5	c.190T>G	C64G	4	Myriad, others	AGTGTCTTTTGTGAAGAATGAT
5	c.190T>C	C64R	11	Myriad, Radice	AGTGTCTTTTGTGAAGAATGAT
5	c.191G>A	C64Y	19	Myriad, others	AGTGTCTTTTGTGAAGAATGATA
5	c.199G>T	D67Y	8	Myriad	TATGTAAGAATTATATAACCAAAAG
5	c.201T>G	D67E	3	Patmisiriwat	ATGTAAGAATGAGATAACCAAAAG
5	c.203T>A	I68K	1	Myriad	TAAGAATGATAAAACCAAAAGgta
5	c.211A>G	R71G	35	Myriad, others	TATAACCAAAAGgtatataattgg
6	c.216C>A	S72R	1	Myriad	TaatttcagGAGACTACAAGAAAGT
6	c.230C>T	T77M	2	Myriad, Radice	ACAAGAAAGTATGAGATTTAGTCAA
6	c.231G>T		1	Casey	TACAAGAAAGTACTAGATTTAGTCA
6	c.259T>G	L87V	2	Myriad	GTGGAAGAGCTAGTGAATATCATT
6	c.266T>C	I89T	2	Myriad, Caldes	GCTATTGAAAACGATTGTGCTTTT
6	c.269T>C	I90T	1	Weber	TATTGAAAATCACTGTGCTTTTC
6	c.286G>A	D96N	1	Myriad	TGCTTTTCAGCTTACACAGGTTTG
6	c.292G>C	G98R	1	Diez	CAGCTTGACACAGTTTGGAGTgta
10	c.612G>C	L204F	6	Myriad, Lidereau	GAGATCAAGAATTCTACAAATCAC
10	c.637A>G	R213G	1	Myriad	CTCAAGGAACCGGGATGAATC
10	c.641A>G	D214G	11	Myriad, Cheetham	AAGGAACCAAGGGTGAAATCAGTT
10	c.661G>T	A221S	2	Myriad	GTTT GATTCTTCAAAAAAG gtaa
18	c.5085T>A	F1695L	1	Myriad	agATGCTGAGTTAGTGTGTGAACGG
18	c.5086G>C	V1696L	1	Myriad	gATGCTGAGTTTGTGTGAACGGGA
18	c.5089T>C	C1697R	2	Borg, Bergthorsson	GCTGAGTTGTGCGTGAACGGACAC
18	c.5095C>T	R1699W	13	Myriad, others	TTGTGTGTGAATGGACACTG AAAT
18	c.5096G>A	R1699Q	11	Myriad, Abbs	TTGTGTGTGAACAGACACTGAAATA
18	c.5096G>T	R1699L	1	Myriad	TTGTGTGTGAACAGACACTG AAATA
18	c.5113C>T		1	Murphy, Voglino	CTGAAATATTTTITAGGAATTGCGG
18	c.5117G>A	G1706E	7	Myriad, others	AAATATTTTCTAGCAATTGCGGGAG
18	c.5117G>C	G1706A	6	Myriad, Cheetham	AAATATTTTCTAGCAATTGCGGGAG
18	c.5123C>A	A1708E	45	Myriad, others	TTCTAGGAATTGAGGGAGGAAATG
18	c.5138T>C	V1713A	3	Myriad, Struewing	GAGGAAAATGGGAGTTAGCTATT
18	c.5143A>C	S1715R	1	Borg	AATGGGTAGTTGCTATTCTgtaa
18	c.5143A>T	S1715C	1	Myriad	AATGGGTAGTTGCTATTCTgtaa

Legend: ¹, the first nucleotide of the start codon is 1. ², mutated nucleotides are highlighted in grey. Large case indicates exonic sequences, small case indicates intronic sequences.

Supplementary Table 4 Exons skipped as a result of pathogenic exonic mutations and a control group

Gene	Exon	Phenotype	Mutation	Reference	HOT-SKIP rank	HOT-SKIP percentile	ID ¹	15-mer included	15-mer excluded	Exon length
RET	11	Hirschsprung disease	E11+62C>T	Auricchio, A. et al. (1999) Am. J. Hum. Genet. 64, 1216-1221.	9	0.89	S1	CCTTCATCGTCTCGG	CCTTCATTGTCTCGG	257
TNFRSF5	5	Hyper IgM syndrome	E5+5A>T	Ferrari, S. et al. (2001) Proc. Natl. Acad. Sci. USA 98, 12614-12619.	80	22.22	S2	cagCTACAGGGGTTT	cagCTACTGGGGTTT	94
PTPRC	4	T cell activation	E4+76C>G	Thude, H. et al. (1995) Eur. J. Immunol. 25, 2101-2106.	560	72.16	S3	TCTCACCCGCAAGCA	TCTCACCGCAAGCA	198
HPRT1	3	HPRT deficiency	E3+29A>T	Steingrimsdottir, H. et al. (1992) Nucleic Acids Res. 20, 1201-1208.	129	17.92	S4	TGTGATGAAGGAGAT	TGTGATGTAGGAGAT	184
FBN1	52	Marfan syndrome	E52+41C>T	Liu, W.G. et al. (1997) Nat. Genet. 16, 328-329.	33	13.31	S5	GGATCATCGTGGGAC	GGATCATTGTGGGAC	66
HMBS	3	Acute intermittend porphyria	E3+33C>G	Llewellyn, D.H. et al. (1996) J. Med. Genet. 33, 437-438.	3	1.50	S6	TGATTCGCGTGGGTA	TGATTCGGGTGGGTA	54
GLA	Cryptic	Fabry disease	Ecr+54A>G	Ishii, S. et al. Am J Hum Genet, 70, 994-1002.	111	52.36	S7	CCCACTAAAGTgtaa	CCCACTAGAGTgtaa	57
FGB	Cryptic	Afibrinogenemia	Ecr+23G>A	Dear, A. et al. J. Thromb. Haemost, 4, 471-472.	109	59.24	S8	CACAGATGAAGCAAC	CACAGATAAGCAAC	50
MLH1	3	Hereditary non-polyposis colorectal cancer	E3+7G>T	McVety, S. et al. (2006) J. Med. Genet. 43, 153-156.	233	61.32	S9	gAAAGAAGATCTGGA	gAAAGAATATCTGGA	99
DMD	27	Muscular dystrophy	E27+27G>T	Shiga, N. et al. (1997) J. Clin. Invest. 100, 2204-2210.	89	12.43	S11	ACAAAAAGAAGCGAA	ACAAAAATAAGCGAA	183
COL4A5	Cryptic	Alport syndrome	Ecr+45A>G	King, K. et al. (2002) Hum. Genet. 111, 548-554.	277	48.43	S12	TAACAAGATGGAGAG	TAACAAGGTGGAGAG	147
GH1	3	Isolated growth hormone Deficiency, type II	E3+5A>G	Ryther, R.C. et al. (2003) Hum. Genet. 113, 140-148.	413	89.01	S13	tagGAAGAAGCCTAT	tagGAAGGAGCCTAT	120
NF1	30	Neurofibromatosis, type I	E30+174G>T	Zatkova, A. et al. (2004) Hum. Mutat. 24, 491-501.	104	13.07	S14	ATTTTGGGAAGAGTG	ATTTTGTGAAGAGTG	203
ADA	7	Adenosine deaminase deficiency	E7+37G>A	Ozsahin, H. et al. (1997) Blood 89, 2849-2855.	190	69.85	S15	TGTCCACGCCGGGGA	TGTCCACACGGGGA	72
ATM	44	Ataxia telangiectasia	E44+58G>A	Teraoka, S.N. et al. (1999) Am. J. Hum. Genet. 64, 1617-1631.	293	73.99	S16	TGACCTCGAACAGC	TGACCTCAAAACAGC	103
ATP7A	4	Menkes disease	E4+103G>A	Vulpe, C. et al. (1993) Nat. Genet. 3, 7-13.	288	36.00	S17	GGGAGATGGAATCAA	GGGAGATAGAATCAA	204
BRCA1	18	Breast and ovarian cancer	E18+18G>T	Mazoyer, S. et al. (1998) Am. J. Hum. Genet. 62, 713-715.	50	16.89	S18	TGTGTGTGAACGGAC	TGTGTGTTAACGGAC	78
F8	19	Haemophilia	E19+47C>T	Theophilus, B.D. et al. (2001) Haemophilia 7, 381-391.	3	0.66	S19	AATTTGGCGGTGGGA	AATTTGTGGGTGGGA	117
HEXB	11	Sandhoff disease	E11+8C>T	Wakamatsu, N. et al. (1992) J. Biol. Chem. 267, 2406-2413.	164	23.98	S20	CTTGCGCCGGGCACA	CTTGCGCTGGGCACA	175
IVD	2	Isovaleric acidemia	E2+4C>T	Vockley, J., et al. (2000) Am. J. Hum. Genet. 66, 356-367.	9	2.62	S21	ttagCTTCGTCAGAC	ttagCTTTGTCAGAC	90
MAPT	10	Frontotemporal dementia	E10+15G>T	Jiang, Z et al. (2000) Mol. Cell. Biol. 20, 4036-4048.	87	24.44	S23	TAATTAAGAAGAAGC	TAATTAATAAGAAGC	93
MAPT	10	Frontotemporal dementia	E10+29C>T	Jiang, Z et al. (2000) Mol. Cell. Biol. 20, 4036-4048.	91	25.56	S24	CTGGATCCTAGCAAC	CTGGATCTTAGCAAC	93
MAPT	10	Frontotemporal dementia	E10+66C>T	Jiang, Z et al. (2000) Mol. Cell. Biol. 20, 4036-4048.	65	18.26	S25	AGGATAACATCAAAC	AGGATAATATCAAAC	93
MLH1	17	Hereditary non-polyposis colorectal cancer	E17+79C>T	Nystrom-Lahti, M., et al. (1999) Genes Chromos. Cancer 26, 372-375.	68	19.10	S26	CAITCTTCGACTAGC	CAITCTTTACTAGCC	93
MLH1	17	Hereditary non-polyposis colorectal cancer	E17+90G>C	Nystrom-Lahti, M., et al. (1999) Genes Chromos. Cancer 26, 372-375.	69	19.38	S27	ATTCTTCGACTAGCC	ATTCTTCCACTAGCC	93
PMM2	5	Carbohydrate-deficient glycoprotein syndrome	E5+68G>A	Vuillaumier-Barrot, S. et al. (1999) Hum. Mutat. 14, 543-544.	332	86.46	S28	CAGCCAAGAAGAAGC	CAGCCAAAAGAAGC	100
SMN1	7	Spinal muscular atrophy	E7+25G>T	Lorson, C.L. & Androphy, E.J. (2000) Hum. Mol. Genet. 9, 259-265.	184	92.00	S29	AAAAGAAGGAAGGTG	AAAAGAATGAAGGTG	54

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

APC	14	Familial adenomatous polyposis	E14+175C>G	Goncalves, V., et al. (2008) Mutat. Res. 662, 33-36.	63	7.46	S 30	GATATTACGGAATGT	GATATTAGGGAATGT	215
MCAD	5	Medium-chain acyl-CoA dehydrogenase deficiency	E5+76C>T	Nielsen, K.B. et al. (2007) Am. J. Hum. Genet. 80, 416-432.	167	43.04	S 31	GTTCAGACTGCTATT	GTTCAGATTGCTATT	101
PTEN	6	Bannayan-Riley Ruvalcaba syndrome	E6+19C>T	Suphapeetiporn, K. et al. (2006) Jpn. J. Clin. Oncol. 36, 814-821.	90	16.30	S 32	TCCCAGTCAGAGCGG	TCCCAGTTAGAGCGG	142
PDHA1	6	Pyruvate dehydrogenase complex deficiency	E6+82G>A	Okajima, K. et al. (2006) Mol. Genet. Metab. 87, 162-168.	159	44.66	S 33	CGATGGTGTGCTAA	CGATGGTACTGCTAA	93
DMD	38	Muscular dystrophy	E38+108 del4	Tran, V.K. et al. (2006) J. Med. Genet. 43, 924-930.	120	25.21	S 34	AAGACTTCAATAAAG	AAGACTTGAATAAAG	123
BRCA2	18	Breast cancer	E18+189 C>G	Fackenthal, J.D. et al. (2002) Am. J. Hum. Genet. 71, 625-631.	1194	85.04	S 37	GAACTTACAGATGGG	GAACTTAGAGATGGG	355
DMD	29	Muscular dystrophy	E29+19C>T	Ginjaar, I.B. et al. (2000) Eur. J. Hum. Genet. 8, 793-796.	116	19.86	S 38	TTTGATGCGACATTC	TTTGATGTGACATTC	150
SMN1	7	Spinal muscular atrophy	E7+6C>T	Kashima, T. et al. (2007) Hum. Mol. Genet. 16, 3149-3159.	6	3.00	S 39	agGGTTTCAGACAAA	agGGTTTTAGACAAA	54
BEST1	6	Vitreoretino-choroidopathy	E6+68T>C	Burgess, R. et al. (2008) J. Med. Genet. 46, 620-625.	104	35.14	S 40	CCACTGGGTGTATACA	CCACTGGCGTATACA	78
BEST1	6	Vitreoretino-choroidopathy	E6+71A>G	Burgess, R. et al. (2008) J. Med. Genet. 46, 620-625.	27	9.12	S 41	CTGGTGATACACAG	CTGGTGTGTACACAG	78
HRPT2	2	No exon skipping	E2+34C>G	Hahn, M. A. et al. (2009) J Endocrinol. 201(3):387-96.	227	55.64	N1	AGTACTACACATTGG	AGTACTAGACATTGG	106
HRPT2	5	No exon skipping	E5+36C>T	Hahn, M. A. et al. (2009) J Endocrinol. 201(3):387-96.	93	47.45	N2	AGAAGCACAGAAACC	AGAAGCATAGAAACC	53
HRPT2	7	No exon skipping	E7+152C>T	Hahn, M. A. et al. (2009) J Endocrinol. 201(3):387-96.	435	51.06	N3	TGTGACCCGAGATAT	TGTGACCTGAGATAT	217
NF1	4a	No exon skipping	E4a+62T>G	Ars, E. et al. (2000) Hum Mol Genet. 22;9(2):237-47.	673	89.97	N4	CCAGAAATCTGCCAT	CCAGAAAGCTGCCAT	191
NF1	12a	No exon skipping	E12a+76 G>A	Ars, E. et al. (2000) Hum Mol Genet. 22;9(2):237-47.	189	39.38	N5	TCAAGTGGTTGCGGG	TCAAGTGATTGCGGG	124
NF1	13	No exon skipping	E13+40C>T	Ars, E. et al. (2000) Hum Mol Genet. 22;9(2):237-47.	419	42.58	N6	GATTTGCCGACAAGC	AATTTGCTGACAAGC	250
NF1	20	No exon skipping	E20+105 C>G	Ars, E. et al. (2000) Hum Mol Genet. 22;9(2):237-47.	285	40.03	N7	CTGGCATCACTGAGG	CTGGCATGACTGAGG	182
NF1	21	No exon skipping	E21+114 C>T	Ars, E. et al. (2000) Hum Mol Genet. 22;9(2):237-47.	154	18.51	N8	GGCTGATCGGTTTGA	GGCTGATTGGTTTGA	212
NF1	25	No exon skipping	E25+5T>C	Ars, E. et al. (2000) Hum Mol Genet. 22;9(2):237-47.	310	82.45	N9	tagATACTTCAGAGT	tagATACTTCAGAGT	98
NF1	27	No exon skipping	E27a+23 C>T	Ars, E. et al. (2000) Hum Mol Genet. 22;9(2):237-47.	370	64.69	N 10	TGGAAGACGACCTTT	TGGAAGATGACCTTT	147
NF1	29	No exon skipping	E29+183 T>A	Raponi, M. et al. (2009) FEBS J. 276(7):2060-73.	1297	96.22	N 11	AGGAGTGTGAAGCCA	AGGAGTGAGAAGCCA	341
NF1	37	No exon skipping	E37+34T>A	Baralle, M. et al. (2006) FEBS Lett. 7;580(18):4449-56.	266	67.86	N 12	TGACACTTACACAG	TGACACTAACACAG	102
BRCA2	18	No exon skipping	E18+186 T>G	Bonnet, C. et al. (2008) J Med Genet. 45(7):438-46.	1039	74.00	N 13	ATTGAACCTACAGAT	ATTGAACGTACAGAT	355
BRCA2	23	No exon skipping	E23+19G>A	Whiley, P. J. et al. (2010) BMC Med Genet. 28;11:80.	341	53.28	N 14	ATTTGGCGTCCATCA	ATTTGGCATCCATCA	164
BRCA2	24	No exon skipping	E24+55A>G	Whiley, P. J. et al. (2010) BMC Med Genet. 28;11:80.	418	77.41	N 15	TCACCTCAGCAAATT	TCACCTCGGCAAAATT	139
CFTR	9	No exon skipping	E9+118G>T	Pagani, F. et al. (2003) J Biol Chem. 18;278(29):26580-8.	136	18.99	N 16	CCTGAAAGATATTAA	CCTGAAATATATTAA	183
F8	9	No exon skipping	E9+126G>A	Theophilus, B. D. et al. (2001) Haemophilia. 7(4):381-91.	663	98.66	N 17	GAATCAGGAATCTTG	GAATCAGAAATCTTG	172
F8	12	No exon skipping	E12+52C>G	Theophilus, B. D. et al. (2001) Haemophilia. 7(4):381-91.	533	90.65	N 18	TGAGAACCGAAGCTG	TGAGAACGGAAGCTG	151
F8	16	No exon skipping	E16+184G>A	Theophilus, B. D. et al. (2001) Haemophilia. 7(4):381-91.	803	96.05	N 19	CTGCAAGCCTGGGC	CTGCAAAAACCTGGGC	213
F8	18	No exon skipping	E18+139G>A	Theophilus, B. D. et al. (2001) Haemophilia. 7(4):381-91.	255	35.61	N 20	ACTGTACGAAAAAAA	ACTGTACAAAAAAA	183

<i>F8</i>	19	No exon skipping	E19+13C>G	Theophilus, B. D. et al. (2001) Haemophilia. 7(4):381-91.	289	63.94	N 21	TTTGAGACAGTGGAA	TTTGAGAGAGTGGAA	117
<i>F8</i>	22	No exon skipping	E22+4G>T	Theophilus, B. D. et al. (2001) Haemophilia. 7(4):381-91.	104	17.11	N 22	gtagGTGGATCTGTT	gtagGTGTATCTGTT	156
<i>F8</i>	23	No exon skipping	E23+77G>A	Theophilus, B. D. et al. (2001) Haemophilia. 7(4):381-91.	183	32.45	N 23	TACATCGTTTGCAC	TACATCCATTTGCAC	145
<i>F8</i>	24	No exon skipping	E24+109G>T	Theophilus, B. D. et al. (2001) Haemophilia. 7(4):381-91.	444	76.55	N 24	AAAGCTCGACTTCAC	AAAGCTCTACTTCAC	149
<i>MSH2</i>	2	No exon skipping	E2+128G>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	583	96.52	N 25	CATCCAAGAGAATG	CATCCAAGAGAATG	155
<i>MSH2</i>	3	No exon skipping	E3+229T>C	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	727	66.09	N 26	AAAGGAATGTGTTT	AAAGAACGTGTTT	279
<i>MSH2</i>	6	No exon skipping	E6+49A>G	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	231	44.42	N 27	CTTGCTGAATAAGTG	CTTGCTGGATAAGTG	134
<i>MSH2</i>	9	No exon skipping	E9+102A>G	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	145	30.21	N 28	CAACATTAATAAGTG	CAACATTGATAAGTG	124
<i>MSH2</i>	10	No exon skipping	E10+92T>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	541	92.01	N 29	TCCTTCGTAACAATA	TCCTTCGAAACAATA	151
<i>MSH2</i>	11	No exon skipping	E11+76A>G	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	312	82.98	N 30	TTGTTAAAGAAATG	TTGTTAAGGAAATG	98
<i>MSH2</i>	12	No exon skipping	E12+69C>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	510	52.69	N 31	CTTTGCTCAGTGTC	CTTTGCTAACGTGTC	246
<i>MSH2</i>	13	No exon skipping	E13+82C>T	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	359	44.65	N 32	TTTGTGCCATGTGAG	TTTGTCTATGTGAG	205
<i>MSH2</i>	14	No exon skipping	E14+32G>T	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	390	39.96	N 33	AATCATAGATGAATT	AATCATATATGAATT	248
<i>MSH2</i>	15	No exon skipping	E15+59T>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	557	80.96	N 34	CTAAGCATGTAATAG	CTAAGCAAGTAATAG	176
<i>MLH1</i>	2	No exon skipping	E2+83G>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	276	79.31	N 35	TGGCACC GGATCAG	TGGCACCAGGATCAG	91
<i>MLH1</i>	3	No exon skipping	E3+85G>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	146	38.42	N 36	TACCTATGGCTTCG	TACCTATAGCTTCG	99
<i>MLH1</i>	4	No exon skipping	E4+44C>T	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	250	89.29	N 37	ATTACAACGAAAACA	ATTACAATGAAAACA	74
<i>MLH1</i>	8	No exon skipping	E8+59T>G	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	180	52.94	N 38	GACAATATTCGCTCC	GACAATAGTCGCTCC	89
<i>MLH1</i>	9	No exon skipping	E9+101C>T	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	12	2.75	N 39	CTTCTTACTCTTCAT	CTTCTTATCTTCAT	113
<i>MLH1</i>	11	No exon skipping	E11+76G>C	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	529	88.17	N 40	ACGAGGAGAGCATCC	ACGAGGACAGCATCC	154
<i>MLH1</i>	12	No exon skipping	E12+345 G>T	Lastella, P et al. (2004) J Med Genet. 41(6):e72.	979	66.69	N 41	CAGAGAAGAGAGGAC	CAGAGAATAGAGGAC	371
<i>MLH1</i>	13	No exon skipping	E13+12G>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	540	93.10	N 42	AGACATCGGGAAGAT	AGACATCAGGAAGAT	149
<i>MLH1</i>	14	No exon skipping	E14+58C>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	112	26.67	N 43	CAGTGGGCCTTGGCA	CAGTGGGACTTGGCA	109
<i>MLH1</i>	16	No exon skipping	E16+89T>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	553	85.87	N 44	GAAGGACTTGCTGAA	GAAGGACATGCTGAA	165
<i>MLH1</i>	17	No exon skipping	E17+63G>T	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	121	33.99	N 45	AGGGACTGCCTATCT	AGGGACTTCCTATCT	93
<i>MLH1</i>	18	No exon skipping	E18+52G>A	Tournier, I. et al. (2008) Hum. Mutat. 29(12):1412-24.	163	37.05	N 46	AGAATGCGCTATGTT	AGAATGCACTATGTT	114

Legend: ¹ID, identification number for cases of disease-causing exon skipping (S; test group) and for non-skipping (N) mutations (control group).

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

Supplementary Table 5 HOT-SKIP output for *CFTR* exon 12 mutations

Exon 12 inclusion (%) ¹	Exon 12 position	Base	PESS Count ²	FAS-ESS (hex 2) count	FAS-ESS (hex 3) count	IIE count	IIE sum	Trusted NI-ESS count	NI-ESS all	PESE count	RESCUE-ESE count	EIE count	EIE sum	Trusted NI-ESE count	NI-ESE all	ESS sum	ESE sum	ESS/ESE ratio
80	12	A	0	0	0	0	0	0	-0.2635	0	4	6	69.6107	4	4.2636	0	14	0
10	12	T	1	0	0	1	7.603	3	-3.2152	0	0	0	0	0	0.3799	5	0	5
95	12	G	0	0	0	0	0	0	0	0	3	5	57.6878	5	5	0	13	0
60	13	T	0	0	0	0	0	1	-0.9999	0	1	3	20.0724	1	2.6736	1	5	0.2
95	13	G	0	0	0	0	0	0	0	0	2	4	63.2879	4	4.9964	0	10	0
10	14	A	3	0	0	0	0	5	-4.9999	0	0	5	65.7805	0	0.632	8	5	1.6
10	14	C	0	0	0	0	0	0	-0.3026	0	2	2	14.0768	2	3.4767	0	6	0
5	14	T	1	0	0	0	0	3	-4.0897	0	0	3	23.4016	0	0	4	3	1.33
50	15	T	0	0	0	3	51.24	0	-0.574	0	0	0	0	0	1.5646	3	0	3
35	15	G	0	0	0	2	24.33	0	-0.2152	0	0	1	5.4886	1	2.0346	2	2	1
95	16	A	0	0	0	0	0	0	0	4	3	6	126.3512	6	6	0	19	0
95	16	C	0	0	0	0	0	0	0	1	1	2	19.7412	4	4.6168	0	8	0
90	16	G	0	0	0	0	0	0	0	0	2	2	26.3952	5	5.4927	0	9	0
5	17	A	0	0	0	0	0	0	-0.4099	0	0	2	17.4719	3	3.3977	0	5	0
45	17	C	0	0	0	0	0	0	0	3	1	6	58.4969	6	5.9973	0	16	0
5	18	T	0	0	0	6	112.9	0	-0.9958	0	1	0	0	1	1.2636	6	2	3
40	18	G	1	1	0	5	60.66	1	-2.1332	0	0	1	9.8008	1	1.3364	8	2	4
40	19	A	0	0	0	0	0	0	-0.3026	0	1	2	19.6334	3	3.6372	0	6	0
95	19	C	0	0	0	0	0	0	0	0	0	0	0	0	1.7983	0	0	0
95	19	G	0	0	0	1	17.90	0	-0.2152	1	0	0	0	1	2.4378	1	2	0.5
90	22	C	0	0	0	4	50.70	1	-1.165	0	0	0	0	2	3.0764	5	2	2.5
25	25	A	5	0	0	6	79.38	5	-5.5458	0	0	2	14.3023	0	0	16	2	8
95	28	C	0	0	0	1	10.69	2	-3.1782	0	0	2	12.0784	0	0	3	2	1.5
100	34	G	0	0	0	3	57.20	1	-1.7667	1	1	4	35.9767	4	4	4	10	0.4
100	37	T	0	1	0	2	23.09	1	-2.185	0	0	0	0	0	0.6804	4	0	4
90	40	A	0	0	0	1	9.112	0	0	0	0	5	39.0741	0	2.2617	1	5	0.2
5	40	C	0	0	0	0	0	0	0	0	0	0	0	1	1.571	0	1	0
100	40	G	0	0	0	1	14.84	0	0	0	0	0	0	0	0.9789	1	0	1
100	43	A	0	0	0	2	29.76	0	-1.0397	0	1	0	0	1	1.4219	2	2	1
90	43	C	0	0	0	2	38.66	0	-0.4563	1	0	0	0	0	0.9951	2	1	2
90	43	G	0	2	1	2	27.83	2	-3.0345	0	0	0	0	0	0.05	7	0	7
0	47	A	0	2	0	4	52.85	3	-4.034	0	0	1	6.9016	1	0.8854	9	2	4.5
5	48	A	0	0	0	3	65.76	0	-0.9088	0	0	3	23.3866	0	1.9339	3	3	1
5	48	C	0	0	0	3	45.12	2	-2.6413	0	0	0	0	0	0	5	0	5
10	49	T	1	0	0	4	69.62	3	-3.6678	0	0	2	36.6295	0	0	8	2	4
10	49	G	0	1	1	3	58.31	3	-3.4007	0	0	1	13.1954	0	0.2474	8	1	8
75	50	A	0	0	0	1	13.77	0	0	0	5	6	101.6249	5	5.664	1	16	0.06
10	50	G	0	0	0	1	16.14	0	0	1	0	6	109.1461	6	5.9468	1	13	0.08
95	51	T	0	0	0	1	15.39	1	-1.3821	0	0	1	6.6422	1	2.027	2	2	1
95	51	G	0	0	0	1	12.01	0	-0.3412	0	0	3	32.4598	1	1.7985	1	4	0.25
15	52	T	0	0	0	0	0	2	-2.1118	0	1	1	5.5892	1	1.4898	2	3	0.67
5	52	G	0	0	0	0	0	0	-0.6537	0	0	2	14.8965	0	0.5058	0	2	0

Legend: ¹Exon inclusion data were published previously [Pagani et al., 2005; Pagani et al., 2003] . ²Auxiliary elements abbreviated as PESEs and PESSs [Zhang and Chasin, 2004], FAS-ESSs [Wang et al., 2004], IIEs [Zhang et al., 2008] and (trusted) NI-ESSs [Stadler et al., 2006] were as defined previously.

Supplementary Table 6 *BRCA1* exon 6 inclusion levels correlate with SF2/ASF ESE scores

Mutation	SF2/ASF	SF2/ASF (IgM-BRCA1)	SC35	SRp40	SRp55	%EI
224G	-58.3	-45.3	-40.4	-46.6	-41.8	72.5
224T	-63.9	-49.5	-38.8	-40.4	-41.3	54.0
224C	-55.3	-41.5	-38.6	-40.9	-43.2	85.2
225G	-55.7	-43.5	-40.1	-46.0	-43.2	78.3
225T	-63.4	-49.2	-39.3	-41.8	-43.5	63.6
225C	-55.3	-41.3	-37.3	-40.6	-42.8	72.5
226G	-56.5	-43.9	-41.4	-43.6	-43.2	75.2
226T	-63.8	-49.4	-38.2	-41.8	-43.5	76.5
226C	-57.5	-42.8	-37.0	-39.1	-42.8	71.4
227A	-63.3	-50.4	-40.7	-38.9	-42.7	66.8
227T	-67.2	-52.7	-39.9	-39.4	-43.2	73.8
227C	-60.9	-46.0	-38.7	-36.7	-42.6	79.1
228G	-52.6	-41.7	-41.7	-43.1	-42.7	85.4
228A	-56.0	-44.9	-41.6	-40.7	-42.5	82.8
228C	-53.6	-40.5	-39.6	-38.5	-42.3	81.9
229G	-56.5	-43.9	-40.9	-43.6	-43.2	85.9
229T	-63.8	-49.4	-40.0	-41.8	-43.5	81.2
229C	-57.5	-42.8	-38.8	-39.1	-42.8	79.3
230G	-58.9	-48.4	-43.0	-45.8	-43.4	55.3
230A	-62.3	-51.6	-42.9	-43.5	-43.1	57.8
230T	-66.2	-53.9	-42.1	-44.0	-43.6	65.5
231A	-63.3	-50.4	-40.7	-38.9	-42.7	71.0
231T	-67.2	-52.7	-39.9	-39.4	-43.2	38.8
231C	-60.9	-46.0	-38.7	-36.7	-42.6	66.4
232G	-56.5	-43.9	-40.9	-43.6	-43.2	76.9
232T	-63.8	-49.4	-40.0	-41.8	-43.5	74.6
232C	-57.5	-42.8	-38.8	-39.1	-42.8	85.0
233A	-63.3	-50.4	-40.7	-38.9	-42.7	66.8
233T	-67.2	-52.7	-39.9	-39.4	-43.2	64.5
233C	-60.9	-46.0	-38.7	-36.7	-42.6	77.5
234G	-56.5	-43.9	-40.9	-43.6	-43.2	64.7
234T	-63.8	-49.4	-40.0	-41.8	-43.5	78.1
234C	-57.5	-42.8	-38.8	-39.1	-42.8	70.4
235G	-52.6	-41.7	-41.7	-43.1	-42.7	66.1
235A	-56.0	-44.9	-41.6	-40.7	-42.5	79.5
235C	-53.6	-40.5	-39.6	-38.5	-42.3	81.5
WT	-59.9	-47.2	-40.8	-41.3	-43.0	69.8
r	0.52	0.57	0.18	0.06	0.04	

Legend: Predicted SR ESEs were calculated using the ESEfinder (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese finder.cgi?process=home>) as a sum of all scores at each position of the wild-type (WT) sequence and 36 mutated constructs. We used overlapping 24-mers as an input. r, correlation coefficient (values in bold are significant at the 0.01 level). EI%, per cent of exon inclusion.

Supplementary Table 7 ESEFinder predictions for each BIC substitution

Exon	Mutation	ESE	Exonic position	Motif	Wildtype score	Mutation score	Predicted Effect
5	c.135A>T	Nil					
5	c.139T>G	SRp55	5	TGCATG	3.763907	-	-
5	c.140G>T	SC35	4	TTTCATG	-	2.942525	+
		SRp55	5	TGCATG	3.763907	-	-
5	c.154C>T	SRp40	19	TTTCAAC	3.866663	3.554607	↓
5	c.181T>G	SRp40	41	TCACAGG	3.942668	6.324404	↑
		SF2/ASF	42	CACAGGG	3.138162	5.258191	↑
		SRp55	45	AGTGTC	2.697407	-	-
5	c.182G>A	SF2/ASF	42	CACAGTA	3.138162	3.864944	↑
		SRp55	45	AGTGTC	2.697407	-	-
5	c.189A>T	SC35	48	GTCCTTTA	2.865828	-	-
		SRp55	54	TATGTA	3.610283	-	-
5	c.190T>G	SRp55	54	TATGTA	3.610283	-	-
5	c.191G>A	SRp55	54	TATATA	3.610283	3.001623	↓
5	c.199G>T	SRp55	65	TATATA	-	3.610283	+
5	c.201T>G	Nil					
5	c.203T>A	Nil					
5	c.211A>G	SRp40	72	CCAAAGG	3.830855	4.398684	↑
6	c.216C>A	SF2/ASF	1	GAGACTA	-	2.645522	+
		SRp40	2	AGACTAC	-	2.797046	+
		SC35	3	GACTACAA	-	3.914543	+
6	c.230C>T	Nil					
6	c.231G>T	SRp40	15	GTACTAG	-	2.694455	+
6	c.259T>G	SC35	42	AGCTAGTG	-	3.081021	+
6	c.266T>C	Nil					
6	c.269T>C	SRp40	54	TCACTTG	-	4.056130	+
6	c.286G>A	SRPp40	73	TGACACA	2.815500	-	-
6	c.292G>C	SF2/AS	74	GACACAA	-	2.434044	+
		SRp40	75	ACACAAG	5.274785	4.706956	↓
10	c.612G>C	Nil					
10	c.637A>G	Nil					
10	c.641A>G	SF2/ASF	43	CAGGGGT	-	2.476925	+
10	c.661G>T	SRp40	63	ATTCTGC	2.827091	-	-
18	c.5085T>A	SC35	7	AGTTAGTG	-	2.512089	+
		SRp55	11	TGTGTG	2.820871	-	-
18	c.5086G>C	SC35	7	AGTTTCTG	-	3.140572	+
		SRp55	11	TGTGTG	2.820871	-	-
18	c.5089T>C	SRp55	11	TGTGTG	2.820871	-	-
18	c.5095C>T	SRp40	17	TGAATGG	3.688106	+	+
18	c.5096G>A	SC35	22	AGACACT	4.968561	3.216489	↓
18	c.5096G>T	SC35	22	TGACATG	4.968561	2.9188919	↓
		SRp40	22	TGACACT	-	2.815500	+
18	c.5113C>T	SRp40	36	TTTCTAG	4.151518	-	-
18	c.5117G>A	SC35	43	GAATTGCG	2.530293	-	-
18	c.5117G>C	SC35	43	GAATTGCG	2.530293	-	-
18	c.5123C>A	SRp40	47	TGCGGG	3.558695	-	-
18	c.5138T>C	SRp55	60	TGGGTA	2.708481	-	-
18	c.5143A>C	SC35	66	GTTGCTA	-	3.887473	+
18	c.5143A>T	SC35	66	GTTTGCTA	-	3.656632	+

Legend: Predicted effect of missense mutations within *BRCA1* exons 5, 6, 10 and 18 SR ESEs according to the ESEfinder algorithm (-, loss of ESE, +, gain of ESE, ↓, reduced score above threshold, ↑, increased score above threshold).

Supplementary Table 8 Correlation of *BRCA1* exon 6 inclusion levels with auxiliary splicing elements

Mutation	Input 15-mer	PESS (cnt)	FAS- ESS (hex2)	FAS- ESS (hex3)	IIE(cnt)	IIE(sum)	NI-ESS- trusted(cnt)	NI-ESS- all(sum)	PESE(cnt)	RESCUE- ESE(cnt)	EIE(cnt)	EIE(sum)	NI-ESE- trusted(cnt)	NI-ESE- all(sum)	ESS total	ESE total
224G	CTACAAGgAAGTACG	0	0	0	0	0	0	0	2	3	6	113.28	5	5.58	0	16
224T	CTACAAGaAAGTACG	0	0	0	0	0	0	-0.26	0	2	3	54.55	3	4.02	0	8
224C	CTACAAGcAAGTACG	0	0	0	0	0	0	0	0	2	6	102.64	5	5.63	0	13
225G	TACAAGAgAGTACGA	0	0	0	0	0	0	0	0	3	5	92.41	6	6.00	0	14
225T	TACAAGaAgAGTACGA	0	0	0	0	0	0	-0.6	0	3	4	71.50	4	4	0	11
225C	TACAAGAcAGTACGA	0	0	0	0	0	0	0	0	3	6	94.11	5	5.77	0	14
226G	ACAAGAAgGTACGAG	0	0	0	0	0	1	-1	1	4	4	109.68	4	4.64	1	13
226T	ACAAGAAiGTACGAG	0	0	0	1	13.40	1	-1.46	0	4	5	96.30	4	4	2	13
226C	ACAAGAAcGTACGAG	0	0	0	0	0	0	0	0	3	4	91.61	5	4.89	0	12
227A	CAAGAAAiACGAGA	0	0	0	0	0	1	-1	0	4	6	159.06	4	4	1	14
227T	CAAGAAiACGAGA	0	0	0	0	0	1	-0.89	0	2	6	142.98	3	3.67	1	11
227C	CAAGAAcACGAGA	0	0	0	0	0	0	0	5	5	6	149.85	5	5.55	0	21
228G	AAGAAAgACGAGAT	0	0	0	0	0	0	0	0	4	6	118.79	5	5.78	0	15
228A	AAGAAAcACGAGAT	0	0	0	0	0	0	0	3	5	6	166.33	6	6	0	20
228C	AAGAAAgACGAGAT	0	0	0	0	0	0	0	1	4	5	98.59	5	5.77	0	15
229G	AGAAAGTgCGAGATT	0	0	0	0	0	0	0	0	1	4	51.28	3	3.87	0	8
229T	AGAAAGTcCGAGATT	0	1	1	0	0	1	-1.21	0	1	2	35.44	2	2.53	2	5
229C	AGAAAGTAcCGAGATT	0	1	0	0	0	1	-1.15	0	1	2	35.44	2	2.50	2	5
230G	GAAAGTAgGAGATT	0	1	1	0	0	1	-1	2	0	1	11.70	4	4.37	2	7
230A	GAAAGTAiGAGATT	0	0	0	0	0	0	-0.26	0	0	1	11.70	1	2.28	0	2
230T	GAAAGTAAGATT	0	0	0	0	0	1	-1.64	0	0	5	53.90	1	1.77	1	6
231A	AAAGTACaAGATT	0	0	0	0	0	0	0	1	2	5	73.62	3	4.30	0	11
231T	AAAGTACiAGATT	0	1	0	0	0	1	-1.46	0	0	3	23.71	1	2.41	2	4
231C	AAAGTACgAGATT	0	0	0	0	0	0	0	0	0	4	52.27	2	3.68	0	6
232G	AAGTACGgATT	0	0	0	0	0	0	-0.22	0	0	4	33.83	1	1.95	0	5
232T	AAGTACGiATT	0	0	0	0	0	0	-0.42	0	0	4	38.65	2	2.59	0	6
232C	AAGTACGcATT	0	0	0	0	0	0	0	0	0	3	22.17	1	1.46	0	4
233A	AGTACGAiTTAGT	0	0	0	0	0	0	0	0	1	6	50.18	2	3.41	0	9
233T	AGTACGAiTTAGT	1	0	0	1	10.81	0	-0.22	0	0	2	15.89	0	1.62	2	2
233C	AGTACGAcTTAGT	0	0	0	0	0	0	0	0	0	6	49.15	3	4.24	0	9
234G	GTACGAGgTTAGTC	0	1	1	1	9.88	1	-1	0	0	3	34.55	1	2.83	3	4
234T	GTACGAGiTTAGTC	0	1	0	2	33.31	1	-1.64	0	0	3	34.70	0	1.55	4	3
234C	GTACGAGcTTAGTC	0	0	0	1	10.89	0	0	0	0	4	39.13	3	3.93	1	7
235G	TACGAGgTTAGTCA	0	0	0	0	0	0	-0.41	0	0	3	43.16	3	3.19	0	6
235A	TACGAGaTTAGTCA	0	0	0	0	0	0	0	0	0	6	79.24	3	4.18	0	9
235C	TACGAGAcTTAGTCA	0	0	0	0	0	0	0	0	0	5	58.59	3	3.89	0	8
WT	AAAGTACgAGATT	0	0	0	0	0	0	0	0	0	5	58.45	2	3.34	0	7
Correlation with exon 6 skipping/inclusion (r)		0.133	0.274	0.161	0.016	-0.023	0.249	-0.344	0.106	0.281	0.433	0.370	0.321	0.291	0.239	0.372

Legend: References to auxiliary splicing elements are in the legend to Supplementary Table S5. Cnt, count. Splicing silencers are highlighted in grey.

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

Supplementary Table 9 Top ten positions in *BRCA1* exons that have the highest ESS/ESE ratios

Exon	Position	Base	PESs(cnt)	FAS-ESS(hex2)	FAS-ESS(hex3)	III(cnt)	III(um)	NI-ESS-trusted(cnt)	NI-ESS-all(um)	PES(cnt)	RESCUE-ESE(cnt)	EIE(cnt)	EIE(um)	NI-ESE-trusted(cnt)	NI-ESE-all(um)	ESS(total)	ESE(total)	ESS(only)/ESE(only)
1	41	G	0	5	3	6	104.1312	6	-5.9074	0	0	0	0.0000	0	0.0000	20	0	20.00
1	40	G	0	4	3	6	108.1341	5	-5.2873	0	0	0	0.0000	0	0.0000	18	0	18.00
1	41	T	0	3	3	6	138.0529	6	-5.9772	0	0	0	0.0000	0	0.0000	18	0	18.00
1	42	T	0	4	2	6	96.6484	6	-5.9074	0	0	0	0.0000	0	0.0000	18	0	18.00
1	37	G	0	3	3	6	100.7085	3	-3.9964	0	0	0	0.0000	0	0.8139	15	0	15.00
1	43	T	0	3	2	5	77.1529	5	-5.2101	0	0	0	0.0000	0	0.0000	15	0	15.00
1	42	G	0	3	3	5	78.5736	3	-4.8677	1	0	0	0.0000	0	0.0000	14	1	14.00
1	36	T	1	2	2	5	101.9593	3	-3.6639	0	0	0	0.0000	0	0.0000	13	0	13.00
1	34	G	0	4	1	4	70.5812	3	-3.3027	0	0	1	6.4767	0	0.6496	12	1	12.00
1	45	G	0	2	2	5	95.3031	3	-3.4469	0	0	1	6.0839	0	0.2913	12	1	12.00
2	24	G	3	3	2	5	84.6901	6	-5.9992	0	0	1	6.1053	0	0.0000	19	1	19.00
2	87	G	0	5	4	4	50.8820	5	-5.0000	0	0	0	0.0000	0	0.0000	18	0	18.00
2	24	T	3	1	1	5	121.7954	6	-5.9999	0	0	0	0.0000	0	0.0000	16	0	16.00
2	61	T	6	0	0	4	65.2914	6	-6.0000	0	0	0	0.0000	0	0.0000	16	0	16.00
2	23	T	4	0	0	4	71.3012	6	-6.0000	0	0	0	0.0000	0	0.0000	14	0	14.00
2	69	T	4	0	0	4	58.9190	6	-6.0000	0	0	1	8.1962	0	0.0000	14	1	14.00
2	87	T	0	3	2	5	99.2004	4	-4.8639	0	0	1	10.1204	0	0.0000	14	1	14.00
2	3	T	0	2	1	4	106.2723	5	-5.7668	0	0	1	6.4022	0	0.0000	12	1	12.00
2	30	T	2	0	0	6	138.6834	4	-4.4219	0	0	0	0.0000	0	0.0000	12	0	12.00
2	62	T	2	0	0	5	90.3386	5	-5.6131	0	0	0	0.0000	0	0.0000	12	0	12.00
3	11	T	0	3	3	4	68.3002	4	-4.3364	0	0	1	8.6707	0	0.6351	14	1	14.00
3	47	G	1	3	2	4	64.4232	4	-4.3920	0	0	0	0.0000	0	0.0000	14	0	14.00
3	6	G	0	3	1	6	96.7677	3	-3.9768	0	0	0	0.0000	0	0.0000	13	0	13.00
3	5	T	0	2	1	6	88.5776	3	-3.6508	0	0	0	0.0000	1	1.0000	12	1	12.00
3	47	T	0	0	0	4	87.6495	6	-5.7488	0	0	0	0.0000	0	0.0000	10	0	10.00
3	48	T	0	0	0	5	98.3617	5	-5.0401	0	0	0	0.0000	0	0.0000	10	0	10.00
3	46	A	0	0	0	3	68.4692	6	-5.8634	0	0	0	0.0000	0	0.0000	9	0	9.00
3	46	T	1	0	0	4	127.7230	4	-4.2153	0	0	0	0.0000	0	0.1120	9	0	9.00
3	48	A	3	0	0	3	44.3670	3	-4.1261	0	0	1	5.2486	0	0.1120	9	1	9.00
3	49	T	0	0	0	5	98.3617	4	-4.1547	0	0	0	0.0000	0	0.2913	9	0	9.00
4	56	G	0	5	5	2	19.6733	5	-5.3746	0	0	0	0.0000	0	0.0000	17	0	17.00
4	50	T	2	1	1	6	116.4913	5	-5.6670	0	0	0	0.0000	0	0.0000	15	0	15.00
4	51	T	2	0	0	6	148.7456	5	-5.2275	0	0	0	0.0000	0	0.0000	13	0	13.00
4	6	T	2	0	0	5	87.9461	5	-5.5439	0	0	0	0.0000	0	0.0000	12	0	12.00
4	48	T	1	3	0	3	43.7527	5	-4.9968	0	0	0	0.0000	0	0.0000	12	0	12.00
4	39	T	3	1	1	4	57.3263	3	-3.5439	0	0	1	7.4277	0	0.7985	12	1	12.00
4	55	T	0	0	0	6	183.2497	5	-5.0616	0	0	0	0.0000	0	0.0000	11	0	11.00
4	66	G	2	1	1	1	11.6777	6	-5.7951	0	0	1	10.1713	0	0.0000	11	1	11.00
4	57	T	2	0	0	3	80.6450	5	-4.8483	0	0	0	0.0000	0	0.2153	10	0	10.00
4	66	T	2	0	0	3	48.6613	5	-5.4401	0	0	1	5.5915	0	0.0000	10	1	10.00
5	63	T	3	1	0	6	220.5992	6	-6.0000	0	0	0	0.0000	0	0.0000	16	0	16.00
5	60	T	4	0	0	6	144.8614	5	-5.5205	0	0	0	0.0000	0	0.0000	15	0	15.00
5	62	T	0	2	2	6	169.2435	5	-5.7816	0	0	0	0.0000	0	0.0000	15	0	15.00

5	86	T	0	3	2	6	118.6753	4	-5.4137	0	0	0	0.0000	0	0.0000	15	0	15.00
5	63	G	0	2	1	6	138.8923	4	-4.3579	0	0	1	5.9780	0	0.0000	13	1	13.00
5	84	G	0	3	0	6	98.4990	4	-4.2229	0	0	0	0.0000	0	0.0000	13	0	13.00
5	81	G	0	3	3	3	46.1823	3	-3.1974	0	0	0	0.0000	0	0.7840	12	0	12.00
5	81	T	0	3	1	4	86.3805	4	-3.9968	0	0	0	0.0000	0	0.0500	12	0	12.00
5	49	T	4	0	0	1	10.3777	6	-6.0000	0	0	1	6.1513	0	0.0000	11	1	11.00
5	85	A	0	2	2	5	50.4868	2	-2.9451	0	0	1	6.9035	0	0.4219	11	1	11.00
6	16	G	1	4	1	4	81.2169	5	-5.3027	0	0	0	0.0000	0	0.0000	15	0	15.00
6	81	T	0	4	3	4	73.9876	4	-4.4219	0	0	0	0.0000	0	0.2636	15	0	15.00
6	16	T	3	0	0	5	122.2375	6	-5.8483	0	0	0	0.0000	0	0.0000	14	0	14.00
6	67	T	1	2	0	5	138.5992	5	-5.3019	0	0	0	0.0000	0	0.0000	13	0	13.00
6	67	G	0	2	2	4	99.6882	4	-4.6640	0	0	0	0.0000	0	0.1489	12	0	12.00
6	17	T	1	0	0	4	69.9199	6	-5.8482	0	0	0	0.0000	0	0.0000	11	0	11.00
6	18	T	1	0	0	5	80.6321	5	-5.1395	0	0	0	0.0000	0	0.0000	11	0	11.00
6	22	T	1	0	0	4	96.6372	6	-5.8483	0	0	1	8.2076	0	0.0000	11	1	11.00
6	133	G	1	3	3	1	9.6621	3	-3.6948	0	0	0	0.0000	0	0.1120	11	0	11.00
6	17	G	0	3	1	2	34.2750	4	-4.8668	0	0	0	0.0000	0	0.0000	10	0	10.00
7	15	T	0	3	2	4	45.7771	4	-4.8232	0	0	0	0.0000	0	0.0000	13	0	13.00
7	89	T	0	1	1	4	84.1810	3	-3.5755	0	0	0	0.0000	0	0.7668	9	0	9.00
7	100	T	1	0	0	3	50.5510	5	-5.0416	0	0	0	0.0000	0	0.0000	9	0	9.00
7	102	T	0	1	1	6	126.9902	1	-2.8082	0	0	0	0.0000	0	0.0000	9	0	9.00
7	91	T	0	1	0	3	51.7880	4	-4.4369	0	0	1	8.0362	0	0.0000	8	1	8.00
7	101	T	0	0	0	6	129.8127	2	-4.1551	0	0	1	6.4022	0	0.2153	8	1	8.00
7	19	T	0	2	0	2	19.3848	3	-4.0038	0	0	0	0.0000	0	0.2913	7	0	7.00
7	16	T	0	0	0	6	110.2691	0	-0.7985	0	0	0	0.0000	0	0.7051	6	0	6.00
7	29	T	1	0	0	2	29.6602	3	-3.7358	0	0	0	0.0000	0	0.1974	6	0	6.00
7	93	A	0	0	0	3	37.3820	3	-4.5406	0	0	0	0.0000	0	0.0000	6	0	6.00
8	7	T	1	2	0	6	138.1069	5	-5.0000	0	0	0	0.0000	0	0.4522	14	0	14.00
8	2	T	0	3	3	3	34.0855	3	-3.8793	0	0	0	0.0000	0	0.3746	12	0	12.00
8	10	T	0	0	0	6	158.6042	4	-5.4161	0	0	0	0.0000	0	0.0000	10	0	10.00
8	28	G	0	3	2	2	20.5800	3	-4.2716	0	0	0	0.0000	0	0.0000	10	0	10.00
8	2	G	0	2	2	3	30.5550	2	-2.5426	0	0	0	0.0000	0	0.7818	9	0	9.00
8	40	T	0	0	0	5	114.6853	3	-3.8118	0	0	0	0.0000	0	0.2153	8	0	8.00
8	40	G	0	1	0	5	91.9272	2	-2.1785	0	0	0	0.0000	1	1.0588	8	1	8.00
8	6	T	1	0	0	2	23.4689	4	-4.4193	0	0	1	7.8747	0	0.7668	7	1	7.00
8	27	C	0	2	0	1	9.5850	4	-4.3364	0	0	0	0.0000	0	0.1974	7	0	7.00
8	10	G	0	2	2	6	123.5278	3	-4.1851	1	0	1	5.8763	0	0.0000	13	2	6.50
9	2	T	2	5	4	6	108.5106	5	-5.3746	0	0	0	0.0000	0	0.0000	22	0	22.00
9	3	G	0	5	5	6	97.4951	5	-5.4136	0	0	0	0.0000	0	0.0000	21	0	21.00
9	63	T	1	1	0	6	130.0020	6	-5.6623	0	0	0	0.0000	0	0.0000	14	0	14.00
9	63	G	0	2	2	6	103.5836	4	-5.0319	0	0	0	0.0000	0	0.0000	14	0	14.00
9	62	T	1	1	1	5	118.7999	5	-5.1974	0	0	0	0.0000	0	0.0000	13	0	13.00
9	61	A	0	3	2	4	44.3605	3	-4.3321	0	0	0	0.0000	0	0.4355	12	0	12.00
9	19	T	1	0	0	5	95.2741	5	-5.1472	0	0	1	5.8523	0	0.0000	11	1	11.00
9	7	T	0	2	2	4	71.5893	2	-2.4930	0	0	0	0.0000	1	1.7453	10	1	10.00
9	57	T	0	1	1	4	69.2467	4	-4.2693	0	0	0	0.0000	0	0.4219	10	0	10.00

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

9	61	T	0	2	0	5	129.3002	2	-3.6720	0	0	0	0.0000	0	0.1974	9	0	9.00
10	2591	G	1	6	6	5	74.3023	6	-6.0000	0	0	0	0.0000	0	0.0000	24	0	24.00
10	2590	G	2	6	5	4	50.2910	6	-6.0000	0	0	0	0.0000	0	0.0000	23	0	23.00
10	2592	G	1	6	5	5	97.7258	6	-6.0000	0	0	0	0.0000	0	0.0000	23	0	23.00
10	2184	G	0	6	6	4	74.6988	6	-6.0000	0	0	0	0.0000	0	0.0000	22	0	22.00
10	2593	G	0	5	5	6	96.2493	6	-5.8854	0	0	0	0.0000	0	0.0000	22	0	22.00
10	124	T	2	5	2	6	209.2728	6	-6.0000	0	0	0	0.0000	0	0.0000	21	0	21.00
10	2585	T	0	6	5	4	54.3328	6	-6.0000	0	0	0	0.0000	0	0.0000	21	0	21.00
10	2589	G	2	6	4	3	32.3961	6	-6.0000	0	0	0	0.0000	0	0.0000	21	0	21.00
10	2593	T	0	5	4	6	116.9061	6	-5.8854	0	0	0	0.0000	0	0.0000	21	0	21.00
10	2766	T	0	5	5	6	142.6675	5	-5.7818	0	0	0	0.0000	0	0.0000	21	0	21.00
11	18	G	0	3	3	5	83.3793	3	-4.0894	0	0	0	0.0000	0	0.5251	14	0	14.00
11	17	G	0	3	1	6	110.0782	3	-3.7006	0	0	1	9.1294	0	0.0000	13	1	13.00
11	16	G	0	3	1	5	95.8874	3	-3.5517	0	0	0	0.0000	1	1.0000	12	1	12.00
11	21	T	0	2	1	5	63.7477	3	-3.7890	0	0	0	0.0000	0	0.0000	11	0	11.00
11	52	T	0	3	2	2	17.8734	3	-4.4458	0	0	0	0.0000	0	0.2449	10	0	10.00
11	57	T	0	3	2	1	8.8434	3	-3.6554	0	0	0	0.0000	0	0.0000	9	0	9.00
11	74	T	1	0	0	6	85.1176	2	-2.6637	0	0	0	0.0000	0	0.3364	9	0	9.00
11	13	T	0	1	1	4	76.5599	2	-2.6241	0	0	0	0.0000	1	1.0000	8	1	8.00
11	17	T	0	0	0	6	100.5347	2	-2.7681	0	0	0	0.0000	0	0.2449	8	0	8.00
11	57	G	0	2	2	1	7.9465	3	-3.4988	0	0	0	0.0000	0	0.0000	8	0	8.00
12	4	G	0	3	3	3	31.9948	3	-3.3364	0	0	0	0.0000	1	1.1651	12	1	12.00
12	82	G	0	2	2	5	54.2518	3	-2.8136	0	0	0	0.0000	0	0.5024	12	0	12.00
12	66	T	0	1	0	6	145.0319	4	-4.5502	0	0	0	0.0000	1	1.0000	11	1	11.00
12	67	T	0	2	1	5	102.4150	3	-3.5499	0	0	1	7.4277	0	0.3364	11	1	11.00
12	64	T	0	2	1	4	114.9747	3	-2.9780	1	0	0	0.0000	0	0.6659	10	1	10.00
12	125	T	0	0	0	5	104.1831	4	-3.8051	0	0	0	0.0000	0	0.2913	9	0	9.00
12	8	G	0	2	2	2	20.2503	2	-3.0124	0	0	0	0.0000	0	0.1989	8	0	8.00
12	44	T	0	2	1	1	9.1281	4	-4.3209	0	0	0	0.0000	0	0.3027	8	0	8.00
12	82	T	0	1	1	4	54.9358	2	-2.6351	0	0	0	0.0000	0	0.1489	8	0	8.00
12	92	T	0	0	0	4	64.2369	4	-4.1601	0	0	0	0.0000	0	0.1974	8	0	8.00
13	76	G	0	6	4	6	125.8801	6	-6.0000	0	0	0	0.0000	0	0.0000	22	0	22.00
13	75	T	0	4	2	4	72.9330	6	-5.8482	0	0	0	0.0000	0	0.0000	16	0	16.00
13	28	T	0	2	2	4	61.7693	4	-4.9711	0	0	1	9.0639	0	0.0000	12	1	12.00
13	88	T	0	3	3	3	35.0210	3	-3.1120	0	0	1	6.3108	0	0.9667	12	1	12.00
13	8	T	2	0	0	4	53.4228	5	-4.9780	0	0	1	9.3056	0	0.2153	11	1	11.00
13	61	T	1	0	0	6	160.4328	4	-4.0987	0	0	0	0.0000	0	0.3746	11	0	11.00
13	80	T	0	2	1	4	62.6099	4	-3.9066	0	0	0	0.0000	1	1.4621	11	1	11.00
13	88	G	0	3	3	2	23.8012	3	-3.4355	0	0	0	0.0000	0	1.0695	11	0	11.00
13	5	T	0	1	0	4	70.0482	5	-4.9467	0	0	0	0.0000	0	0.3027	10	0	10.00
13	78	T	0	1	1	6	99.7300	2	-2.3256	0	0	0	0.0000	0	0.3302	10	0	10.00
14	117	G	0	5	4	6	96.6501	5	-3.3364	0	0	1	5.3360	0	0.0000	20	1	20.00
14	55	T	0	4	4	6	105.5948	4	-5.1182	0	0	0	0.0000	0	0.0000	18	0	18.00
14	120	G	0	4	4	6	105.0963	4	-4.3799	0	0	1	6.6391	0	0.2913	18	1	18.00
14	38	G	1	5	4	2	26.1256	5	-5.4621	0	0	0	0.0000	0	0.0000	17	0	17.00
14	107	T	0	5	3	3	92.7321	6	-5.8483	0	0	0	0.0000	0	0.0000	17	0	17.00

14	105	G	0	5	3	3	36.0103	5	-5.1974	0	0	0	0.0000	0	0.0000	16	0	16.00
14	104	T	0	3	3	3	30.1142	6	-5.7100	0	0	0	0.0000	0	0.0000	15	0	15.00
14	109	A	0	3	3	4	50.5979	5	-4.9780	0	0	0	0.0000	0	0.4355	15	0	15.00
14	110	T	0	3	2	6	161.6334	4	-5.2099	0	0	0	0.0000	0	0.0000	15	0	15.00
14	113	C	2	1	1	6	68.2555	5	-5.0000	0	0	0	0.0000	0	0.0000	15	0	15.00
15	283	G	0	5	4	5	88.5264	6	-5.8854	0	0	0	0.0000	0	0.0000	20	0	20.00
15	199	T	0	2	2	6	91.0882	6	-5.9992	0	0	0	0.0000	0	0.0000	16	0	16.00
15	58	T	0	3	2	6	83.3104	4	-4.3103	1	0	0	0.0000	0	0.4522	15	1	15.00
15	286	G	1	3	2	6	101.7566	3	-3.8732	0	0	0	0.0000	0	0.0000	15	0	15.00
15	205	T	4	1	0	3	38.1007	6	-6.0000	0	0	0	0.0000	0	0.0000	14	0	14.00
15	199	G	0	2	1	6	93.3233	4	-5.2289	0	0	0	0.0000	0	0.0000	13	0	13.00
15	287	T	0	2	0	6	111.6955	5	-5.2430	0	0	0	0.0000	0	0.0000	13	0	13.00
15	93	T	0	2	1	5	86.1417	4	-4.4298	0	0	0	0.0000	0	0.0000	12	0	12.00
15	284	A	0	2	2	4	51.7832	4	-5.1032	0	0	0	0.0000	0	0.0000	12	0	12.00
15	289	T	0	1	1	6	120.6807	4	-4.0932	0	0	0	0.0000	0	0.4621	12	0	12.00
16	76	T	3	1	0	6	206.9996	6	-6.0000	0	0	1	10.2122	0	0.0000	16	1	16.00
16	73	T	3	1	1	5	137.6425	5	-5.2693	0	0	0	0.0000	0	0.0000	15	0	15.00
16	75	A	2	3	2	3	40.7681	5	-5.9337	0	0	0	0.0000	0	0.0000	15	0	15.00
16	76	G	0	2	1	6	132.8496	4	-5.1032	0	0	0	0.0000	0	0.0000	13	0	13.00
16	76	A	3	0	0	4	63.7653	6	-5.9074	0	0	1	13.7667	0	0.0000	13	1	13.00
16	49	G	3	0	0	3	40.1040	6	-5.8402	0	0	1	5.8562	0	0.0000	12	1	12.00
16	74	T	0	2	1	6	127.9846	3	-4.7672	0	0	0	0.0000	0	0.0000	12	0	12.00
16	75	T	0	2	1	6	137.1580	3	-4.7672	0	0	0	0.0000	0	0.0000	12	0	12.00
16	77	T	0	2	1	5	111.9201	4	-5.4911	0	0	1	12.2039	0	0.0000	12	1	12.00
16	49	T	3	0	0	3	46.0814	5	-5.6351	0	0	1	7.7980	0	0.0000	11	1	11.00
17	13	G	0	5	3	6	129.2447	6	-5.9468	0	0	0	0.0000	0	0.0000	20	0	20.00
17	14	T	4	3	2	6	209.0031	5	-4.9919	0	0	0	0.0000	0	0.0000	20	0	20.00
17	63	T	3	3	3	4	44.5842	6	-6.0000	0	0	0	0.0000	0	0.0000	19	0	19.00
17	64	T	1	4	4	4	46.1980	6	-5.9919	0	0	0	0.0000	0	0.0000	19	0	19.00
17	65	A	0	5	5	3	35.3238	6	-5.9919	0	0	0	0.0000	0	0.0000	19	0	19.00
17	65	T	1	4	2	6	123.4893	6	-5.8475	0	0	0	0.0000	0	0.0000	19	0	19.00
17	71	T	3	3	3	4	74.0062	6	-5.9973	0	0	0	0.0000	0	0.0000	19	0	19.00
17	49	G	0	4	4	6	104.4848	4	-4.6351	0	0	1	5.3811	0	0.1651	18	1	18.00
17	63	G	1	4	3	4	46.1980	6	-5.9919	0	0	1	6.1003	0	0.0000	18	1	18.00
17	39	T	4	2	1	5	142.3472	5	-5.4930	0	0	0	0.0000	0	0.0000	17	0	17.00
18	2	G	0	4	4	4	43.6522	4	-4.4930	0	0	0	0.0000	1	0.8854	16	1	16.00
18	13	T	1	1	1	4	85.6632	5	-4.9468	0	0	0	0.0000	0	0.6351	12	0	12.00
18	15	T	0	0	0	5	93.6233	5	-5.2275	0	0	0	0.0000	0	0.0000	10	0	10.00
18	33	T	2	0	0	5	93.1546	2	-2.3364	0	0	0	0.0000	1	1.5850	9	1	9.00
18	13	A	1	1	0	2	26.0076	4	-4.1974	0	0	1	5.6530	0	0.6640	8	1	8.00
18	35	T	2	0	0	2	21.5173	4	-4.7468	0	0	1	8.2257	0	0.0000	8	1	8.00
18	15	G	0	0	0	4	69.9898	3	-3.9689	0	0	0	0.0000	0	0.0000	7	0	7.00
18	3	T	0	3	2	4	49.5520	3	-3.1651	1	0	0	0.0000	1	1.6351	12	2	6.00
18	6	T	0	1	0	3	38.4873	1	-1.0000	0	0	0	0.0000	0	0.6440	5	0	5.00
18	2	T	0	2	1	3	39.5495	3	-3.4902	0	0	2	19.4620	0	0.6351	9	2	4.50
19	26	G	1	4	2	4	54.2491	4	-4.6579	0	0	1	9.8008	0	0.0000	15	1	15.00

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

19	5	T	2	1	0	6	178.5894	4	-4.9089	0	0	0	0.0000	0	0.0000	13	0	13.00
19	5	G	0	2	1	4	92.1606	4	-4.4678	0	0	0	0.0000	0	0.0000	11	0	11.00
19	21	G	1	2	2	4	47.2445	2	-3.1220	0	0	0	0.0000	0	0.8272	11	0	11.00
19	30	T	0	2	0	3	42.7251	6	-5.8402	0	0	0	0.0000	0	0.0000	11	0	11.00
19	10	T	0	0	0	3	74.5780	5	-5.1587	0	0	0	0.0000	0	0.0000	8	0	8.00
19	11	T	0	1	0	4	107.0449	3	-3.8443	0	0	0	0.0000	0	0.1120	8	0	8.00
19	28	T	0	0	0	4	83.2884	4	-4.4667	0	0	0	0.0000	0	0.0000	8	0	8.00
19	7	T	0	0	0	6	128.3531	1	-3.4725	0	0	1	6.2089	0	0.6351	7	1	7.00
19	10	A	2	0	0	2	27.3763	3	-3.2913	0	0	1	7.9553	0	0.3027	7	1	7.00
20	6	T	0	3	3	5	52.2851	4	-5.0386	0	0	0	0.0000	0	0.0000	15	0	15.00
20	13	T	0	4	4	3	43.6268	4	-4.5459	0	0	1	7.4277	0	0.3364	15	1	15.00
20	13	G	0	4	4	3	40.0267	4	-4.3103	0	0	0	0.0000	1	1.0000	15	1	15.00
20	27	T	0	3	1	6	124.0563	5	-5.7668	0	0	1	11.9438	0	0.0000	15	1	15.00
20	26	T	1	3	0	3	63.6071	5	-5.7818	0	0	1	12.1273	0	0.0000	12	1	12.00
20	21	T	0	2	1	4	115.0444	4	-4.4487	0	0	1	15.8357	0	0.6640	11	1	11.00
20	29	T	1	1	1	6	129.7393	1	-2.7838	0	0	0	0.0000	0	0.0000	10	0	10.00
20	7	T	0	1	1	6	89.8675	1	-2.3634	0	0	0	0.0000	0	0.2913	9	0	9.00
20	8	T	0	1	1	4	41.6597	3	-3.6819	0	0	0	0.0000	0	0.4219	9	0	9.00
20	13	A	0	2	2	3	30.4406	2	-2.7132	0	0	0	0.0000	0	0.6272	9	0	9.00
21	15	G	0	3	2	5	98.9773	4	-5.0080	0	0	1	6.1003	0	0.0000	14	1	14.00
21	34	T	1	2	0	6	136.4068	5	-5.3096	0	0	0	0.0000	0	0.0000	14	0	14.00
21	34	G	0	3	2	6	96.9836	3	-4.6631	0	0	1	8.0120	0	0.0000	14	1	14.00
21	15	T	1	1	0	5	82.1792	4	-5.2925	0	0	0	0.0000	0	0.0000	11	0	11.00
21	40	G	0	2	1	6	89.3671	2	-3.6411	0	0	0	0.0000	0	0.0000	11	0	11.00
21	30	A	0	2	1	4	70.3991	2	-4.1183	0	0	0	0.0000	0	0.0000	9	0	9.00
21	31	T	1	0	0	6	161.7842	2	-3.8790	0	0	0	0.0000	0	0.0000	9	0	9.00
21	32	G	0	1	1	6	92.7101	1	-2.2546	0	0	0	0.0000	0	0.0000	9	0	9.00
21	33	T	0	1	0	6	89.8983	2	-3.7074	0	0	0	0.0000	0	0.0000	9	0	9.00
21	37	T	0	0	0	6	173.8509	3	-4.1146	0	0	1	5.9780	0	0.0000	9	1	9.00
22	17	G	0	4	2	6	107.6740	4	-4.6351	0	0	1	5.3811	0	0.1651	16	1	16.00
22	19	T	2	1	1	6	204.3045	6	-5.9919	0	0	0	0.0000	0	0.0000	16	0	16.00
22	20	G	0	4	2	6	125.7077	4	-4.3799	0	0	1	7.7102	0	0.3027	16	1	16.00
22	18	A	0	3	2	6	98.8402	4	-5.3492	0	0	0	0.0000	0	0.0000	15	0	15.00
22	18	T	0	3	1	6	171.4680	4	-5.5336	0	0	0	0.0000	0	0.0000	14	0	14.00
22	22	T	0	2	1	6	122.9822	4	-4.1767	0	0	0	0.0000	1	0.8854	13	1	13.00
22	4	T	0	3	2	3	36.7787	4	-4.5425	0	0	0	0.0000	0	0.2913	12	0	12.00
22	16	T	0	1	1	5	112.9953	5	-5.0212	0	0	0	0.0000	0	0.0000	12	0	12.00
22	54	T	2	1	0	3	47.3199	5	-5.5425	0	0	0	0.0000	0	0.0000	11	0	11.00
22	19	G	0	1	0	6	126.2256	3	-3.8567	0	0	1	5.9780	0	0.0000	10	1	10.00
23	916	G	0	6	5	6	123.6430	6	-6.0000	0	0	0	0.0000	0	0.0000	23	0	23.00
23	46	G	0	6	4	6	114.1818	6	-6.0000	0	0	0	0.0000	0	0.0000	22	0	22.00
23	1272	G	2	4	4	6	154.3724	6	-5.9780	0	0	0	0.0000	0	0.0000	22	0	22.00
23	847	G	0	5	5	5	100.0172	6	-5.8854	0	0	0	0.0000	0	0.0000	21	0	21.00
23	1045	T	4	4	2	6	129.9822	5	-5.5459	0	0	0	0.0000	0	0.0000	21	0	21.00
23	1147	G	0	5	5	6	99.9365	5	-5.6351	0	0	0	0.0000	0	0.0000	21	0	21.00
23	354	G	0	5	3	6	114.7952	6	-6.0000	0	0	0	0.0000	0	0.0000	20	0	20.00

23	913	G	0	5	4	6	942765	5	-5.5503	0	0	0	0.0000	0	0.0000	20	0	20.00
23	1146	T	5	2	1	6	2726192	6	-6.0000	0	0	0	0.0000	0	0.0000	20	0	20.00
23	1322	T	5	2	1	6	2146496	6	-5.8854	0	0	0	0.0000	0	0.0000	20	0	20.00

For Peer Review