

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

Michela Raponi^{1,*}, Ellen Copson^{1,*}, Jana Kralovicova^{1,*}, Petr Divina², Diana Baralle¹,
Diana Eccles¹, Peter Johnson¹, 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 messenger RNA, but their overall fraction among disease-causing exonic substitutions is unknown. Here, we have systematically tested missense and silent mutations deposited in the *BRCA1* mutation databases of unclassified variants for their effects on exon inclusion in the mRNA experimentally. The introduction of 21 *BRCA1* variants in two minigene systems revealed a single example of exon 6 skipping upon mutation c.231G>T. Comprehensive mutagenesis of a 12-nt exon 6 segment overlapping this change showed that among 36 alterations this mutation gave the highest exon inclusion levels. The exon 6 inclusion levels significantly correlated with the profile of splicing enhancers/silencers, prompting the development of new online utilities, termed EX-SKIP and HOT-SKIP. The latter tool examines all possible mutations at each exon position and identifies ‘hot’ exonic substitutions with the highest ESS/ESE ratio that are likely to result in exon skipping while the former utility quickly determines which of the two or more alleles is more susceptible to this type of aberrant splicing. Using a set of disease-causing exonic substitutions that resulted in genetic disease we show that distribution of these changes was biased toward top-ranking HOT-SKIP positions. Together, these results identify a new splicing mutation in *BRCA1* exon 6 in an unstructured silencer element and facilitate prediction of exonic substitutions that reduce exon inclusion levels in mature transcripts.

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

INTRODUCTION

Single-nucleotide substitutions in exons 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 associated with diminished expression of canonical mRNAs and is usually dramatic in severe phenotypes. Nevertheless, single-nucleotide variants in the coding region may induce more subtle changes in the relative expression of exon-lacking or exon-containing transcripts [Cartegni et al., 2002; Cooper et al., 2009]. The exact proportion of such DNA alterations among disease-causing exonic substitutions is, however, unknown because RNA samples have not been examined in the majority of the cases and routine mutation screening has been largely based on DNA analysis.

Exonic substitutions often create or eliminate short elements that inhibit or activate exon inclusion, 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]. The latter gene 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 450 distinct *BRCA1* missense mutations have been identified in patients with a strong family history of breast cancer, 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 minigene systems. We have identified a single exon-skipping mutation c.231>G>T. Systematic mutagenesis of a 12-nt segment surrounding this alteration showed that this alteration gave the highest exon skipping levels among 36 minigene substitutions and revealed a regular pattern of exon inclusion across this motif. Because this mutation was predicted by a combination of ESS and ESE

elements, we have developed two online utilities termed EX-SKIP and HOT-SKIP that predict exon-skipping mutations *ab initio* and tested their performance using two datasets. These results reveal a new splicing regulatory element in *BRCA1* exon 6 and facilitate identification of exonic alterations that result in exon exclusion from mature transcripts.

MATERIAL AND METHODS

Selection of *BRCA1* variants

The Breast Cancer Information Core database (<http://research.nhgri.nih.gov/bic/>), 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 (Supplementary Table 1). The number of cases that sustained these mutations 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 that contained a high frequency of substitutions.

Construction of splicing reporters

OLIGO[®] online software (<http://www.oligo.net>) was used to design PCR primers to amplify three-exon *BRCA1* minigenes (Supplementary Table 2). Amplification was performed with anonymous genomic DNA, 1 μ l of each primer (10 pmol/ μ l) and a mixture of 0.25 μ l AmpliTaq Gold polymerase and 0.25 μ l *Pfu* DNA polymerase. PCR products were visualized on agarose gels, DNA fragments were extracted using GENECLAN[®] and subcloned into pCR3.1 (Invitrogen). Plasmid DNA was extracted

using Wizard[®] Plus SV Minipreps DNA Purification System (Promega). Site-directed mutagenesis was carried out using a megaprimer approach [Sarkar and Sommer, 1990] with mutagenic primers listed in Supplementary Table 3.

Cell culture and transfections

HEK293 cells 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 3×10^5 HEK293 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.

Analysis of spliced products

Total RNA was extracted using the Qiagen RNeasy Plus kit (Qiagen) according to the manufacturer's instructions. One μg 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].

EX-SKIP and HOT-SKIP construction

Both tools were implemented as a common gateway interface script (written in Perl). Each utility perform a local search of the exonic sequence for putative ESEs/PESs (PESEs/PESs) [Zhang and Chasin, 2004], FAS-ESSs [Wang et al., 2004], RESCUE-ESEs [Fairbrother et al., 2002; Fairbrother et al., 2004], neighbourhood inference [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, their counts or their 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 (Supplementary Table 4). 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 exon 12 inclusion levels measured after transfection of the wild-type *CFTR* minigene and its 42 mutated versions, as published previously [Pagani et al., 2005; Pagani et al., 2003] (Supplementary Table 5).

RESULTS

Identification of a silent mutation in *BRCA1* that induced exon skipping

A search of the Breast Cancer Information Core (BIC) database for disease-associated exonic substitutions identified a total of 487 alterations. For further investigation, we selected substitutions in *BRCA1* exons 5, 6, 10 and 18 because they were flanked by relatively short introns and their length was also small (78, 89, 77 and 78 bp, respectively), thus facilitating DNA manipulation. We compiled a total of 38 non-synonymous and 2 synonymous substitutions (Supplementary Table 3).

Splicing of the wild-type reporter construct containing exons 5-7 (Fig. 1A) produced a mixture of correctly spliced products and exon skipping. Examination of RNA products of reporters mutated at each exonic variant (Fig. 1B) revealed that the c.231G>T mutation significantly increased exon skipping as compared to the wild-type reporter (Fig. 1C), while the remaining variants in exon 6 did not produce any splicing defects.

The wild-type exon 9-11 minigene did not induce exon skipping, nevertheless we mutated this construct at each variant because it gave retention of the second minigene intron (Fig. 2A-C). None of the mutants produced exon skipping, but mutation c.5138T>C improved splicing of the second minigene intron, perhaps as a result of a removal of predicted ESS octamer AAATGGGT [Zhang and Chasin, 2004] and overlapping FAS-ESS hexamers TAGTTA, GTAGTT, GGTAGT and GGGTAG [Wang et al., 2004] from the wildtype construct. Minigenes containing exons 5 and 18 as middle minigene exons were either not informative or their cloning was unsuccessful.

Taken together, the introduction of 21 *BRCA1* variants in two minigenes revealed a single (~5%) example of exon skipping upon c.231G>T mutation. This mutation was likely to markedly diminish expression of canonical *BRCA1* transcripts *in vivo* and may have thus contributed to the development of breast cancer in the reported case.

Characterization of a silencer element in *BRCA1* exon 6

The c.231G>T alteration 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 alteration 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 the contribution of each nucleotide, we mutated each position of a 12-nt region surrounding this mutation. Measurements of exon inclusion levels of the wild-type and 36 mutated constructs showed that the c.231G>T transversion gave the highest level of transcripts lacking exon 6 while only a few mutations reduced exon 6 skipping (Fig. 3A). This mutation creates a new TAG trinucleotide, which are highly enriched in splicing silencers [Wang et al., 2004; Vorechovsky, 2009]. In addition, exon skipping/inclusion levels across the mutated region showed 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 changes can be predicted. As expected, the average exon skipping tended to be highest for substitutions that introduced uracil in the pre-mRNA (Fig. 3B), consistent with a general enrichment of this nucleotide in splicing silencers [Wang et al., 2004; Vorechovsky, 2009]. Finally, although uracil is overrepresented in introns and in predicted splicing silencers, including FAS-ESSs [Wang et al., 2004] and PESSs [Zhang and Chasin, 2004], it is worth noting that X>T substitutions did not increase exon skipping at each position (Fig. 3A).

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], neighbourhood inference elements [Stadler et al., 2006] and EIE/IEEs [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 SR proteins. The highest correlation with enhancers was observed for the score density of the improved consensus of serine/arginine-rich (SR) protein SF2/ASF (Fig. 3C, Supplementary Tables 6,7). Together, ESEfinder predictions identified a total of 669 potential ESE motifs, with 34 potential ESEs within exons 5, 6, 10 and 18 (Supplementary Table 7), but putative ESEs for the remaining SR proteins did not significantly correlate with the exon inclusion. A significant correlation was also found for EIEs ($r=0.43$), followed by a simple sum of all enhancers ($r=0.37$, Supplementary Table 8). 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 ($r=0.25$; Supplementary Table 8).

Putative secondary structure

Of 36 minigene substitutions, the c.231G>T mutation gave the highest levels of exon skipping that were not matched by any other mutation. This could be due to an alteration of RNA secondary structure, which may convert the predicted single-stranded configuration into a base-paired structure and stabilize a terminal loop (Fig. 4A). In an attempt to address this hypothesis, we introduced an additional destabilizing mutation at position 238. However, exon inclusion levels for this mutant at both the 231G and 231T background were higher than for the wild-type reporter construct (Fig. 4B), which makes this explanation unlikely. Prediction of 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, suggesting that the c.231G>T mutation could primarily alter protein binding (data not shown).

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

Because combination of independently derived splicing regulatory motifs have 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 multiple 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 (lower case) on either side to account for adjacent 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 indicating the number of each ESSs/ESEs, their sum and ratio, which 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. Table data can be

sorted online by clicking on chosen elements or their combination in table headers. Typically, substitutions with the highest values of ESSs and lower values of ESEs are likely to represent ‘hot-spots’ for exon skipping that are highlighted in red in the exonic sequence. 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 calculated for each *BRCA1* exon is shown in Supplementary Table 9.

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 exon position for 33 exons that sustained a total of 37 alterations, each resulting in exon skipping and genetic disease (Supplementary Table 4). 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. Interestingly, 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. 5). 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 10 mutations that were in the third and fourth quartiles were located in the vicinity of top-ranking positions. For example, 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 within 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 first 5 exon-skipping mutations, just 6 nt from the disease-causing substitution at position 189 [Fackenthal et al., 2002]. In cryptic exons, identical positions were 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 cases, the ESS/ESE profile did not correctly predict mutations that are likely to reside within a shared secondary structure. This leads to prediction that the ESS/ESE profile may be more useful for less structured pre-mRNAs. In addition, it may not perform well for mutations very close to natural splice sites. For example, two of the 10 ‘non-performing’ mutations at position 7 in *MLH1* exon 3 [McVety et al., 2006] and position 5 in *GHI* exon 3 [Ryther et al., 2003]) were close to 3’ splice sites and the 3’YAG 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 U2 small nuclear ribonucleoprotein. The small subunit of this factor may bind up to 12 nt of the exon and has a preference for uridines [Wu et al., 1999]. Finally, 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. 5).

We also 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 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 highest correlation coefficient (-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 detail 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 10; (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 and probably unstructured ESS.

We believe that the first conclusion is not limited to the two *BRCA1* segments. Although nonsense, missense and translationally silent mutations can inactivate genes by inducing the splicing machinery to skip the mutant exons, most mutations to illustrate this point were in fact in the splice-site consensus (see 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] confirm 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 affect 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]. Thus, although the overall

fraction of pathogenic exonic substitutions that cause aberrant splicing is likely to be relatively low, they should always be considered, particularly in exons weakly included in mRNAs, such as *CFTR* exon 12, where almost a third of all point mutations induced exon skipping in a minigene system [Pagani et al., 2005] (Supplemental Table 5).

Remarkably, the ESE/ESS profile correctly predicted a pathogenic mutation over a considerable distance from the exonic position (Table 1) and suggested that a combination of independently derived elements has a better predictive and perhaps practical value. The new online tools are publicly available and should facilitate detection of DNA variants that alter splicing if carried out in conjunction with other tools that assess the strength of natural splice sites [Senapathy et al., 1990; Yeo and Burge, 2004] and branch points [Kol et al., 2005]. Similar algorithms have been recently applied to distinguish exon skipping and cryptic splice site activation *ab initio* [Divina et al., 2009].

Interestingly, most minigene substitutions altered exon inclusion levels that ranged between 0.39 and 0.86% (Fig. 3A). Disregarding the effect of the c.231G>T mutation, these changes were minor and translated only to ~1.6-fold difference between maximum and minimum exon skipping. Remarkably, the pattern of these alterations was not irregular, raising a question of the underlying mechanism. Apart from alterations of cross-exon protein interactions and/or RNA secondary/tertiary structure, chromatin and histone modifications have been recently shown to influence splicing [Andersson et al., 2009; Gama-Carvalho et al., 1997; Kolasinska-Zwierz et al., 2009; Kornblihtt et al., 2009; Schor et al., 2009; Tilgner et al., 2009]. Exon marking and nucleosome occupancy between exons and introns also show only ~1.5-fold differences and nucleosome occupancy improved exon recognition only to a limited degree [Spies et al., 2009].

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 pathogenic exon skipping are relatively infrequent. The newly developed online tools should facilitate detection of exonic alterations that reduce exon inclusion in mRNAs and represent the first step toward developing more sophisticated algorithms in the future.

ACKNOWLEDGEMENTS

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

REFERENCES

- Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. 2009. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res* 19:1732-1741.
- 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.
- 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. *European Journal of Human Genetics*:published online 14 January 2009.
- 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.
- Gama-Carvalho M, Krauss RD, Chiang L, Valcarcel J, Green MR, Carmo-Fonseca M. 1997. Targeting of U2AF65 to sites of active splicing in the nucleus. *J Cell Biol* 137:975-987.
- 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.
- 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.
- 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.
- Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genetics* 41:376-381.
- Kornblihtt AR, Schor IE, Allo M, Blencowe BJ. 2009. When chromatin meets splicing. *Nat Struct Mol Biol* 16:902-903.
- Kralovicova J, HOUNGINOU-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.
- 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.
- 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 U S A* 101:7287-7292.
- 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.
- Ozsahin H, Arredondo-Vega FX, Santisteban I, Fuhrer H, Tuchschnid 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 U S A* 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.
- 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.
- 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.
- Schor IE, Rascovan N, Pelisch F, Allo M, Kornblihtt AR. 2009. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc Natl Acad Sci U S A* 106:4325-4330.
- 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.
- Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J, Guigo R. 2009. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*.
- 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. 2009. Transposable elements in disease-associated cryptic exons. *Human Genetics*:Published Oct 10.
- 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 U S A* 105:5797-5802.
- Zhang XH, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* 18:1241-1250.

FIGURES AND TABLES

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	AGTAC ^G ₂₃₁ AGATT	0	0	13	0	6	1	19	1	0.053
Mutant	AGTAC ^T ₂₃₁ AGATT	0	1	2	9	1	1	3	11	3.667

FIGURES

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

Legend: **A**, 5-7 minigene reporter construct. Exons are shown as boxes (numbered 5-7), introns as lines. Normal and aberrantly spliced products are shown as dotted lines above and below the pre-mRNA, respectively. Transposable 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 the overabundance of sequence motifs between human and mouse orthologous exons [Goren et al., 2006]. **C**, RNA products of the *BRCA1* 5-7 minigene (schematically shown to the right) visualized by reverse-transcription PCR and separated by PAGE. Minigene mutations (top) correspond to panel B.

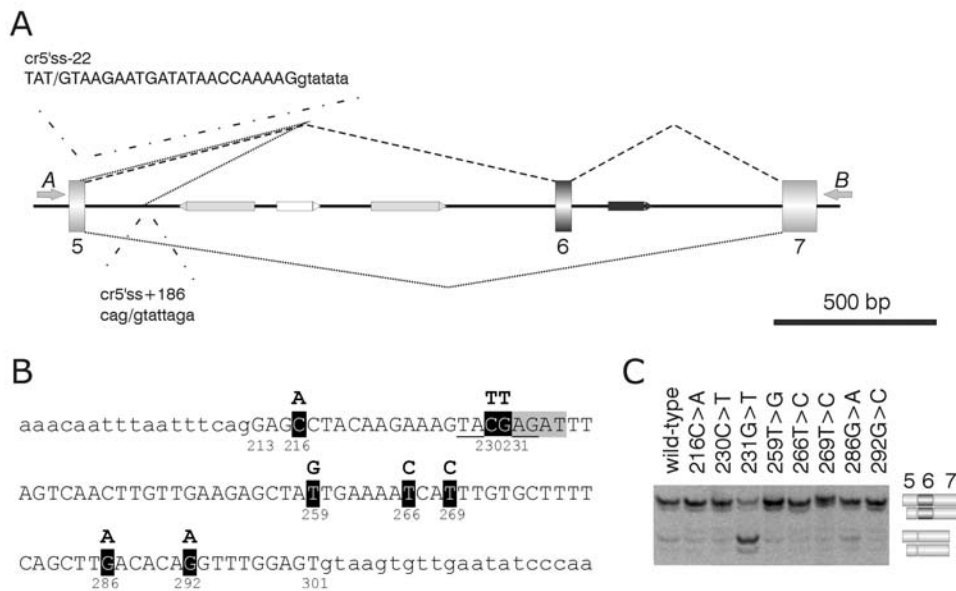


Figure 2 Splicing pattern of the 9-11 reporter construct

Legend: **A**, Exons are shown as boxes (numbered 9-11), introns as lines. Normal and aberrantly spliced products are shown as dotted lines above and below the pre-mRNA, respectively. Transposable elements in introns are schematically shown as grey (Alu), white (L1) and black (MIR3) rectangles; their orientation is denoted by arrowheads. Cloning primers are shown as grey arrows. **B**, exon 10 sequence with single-nucleotide substitutions; **C**, RNA products of the 9-11 minigene (schematically shown to the right) visualized by reverse-transcription PCR and separated by PAGE. Minigene mutations (top) correspond to panel B.

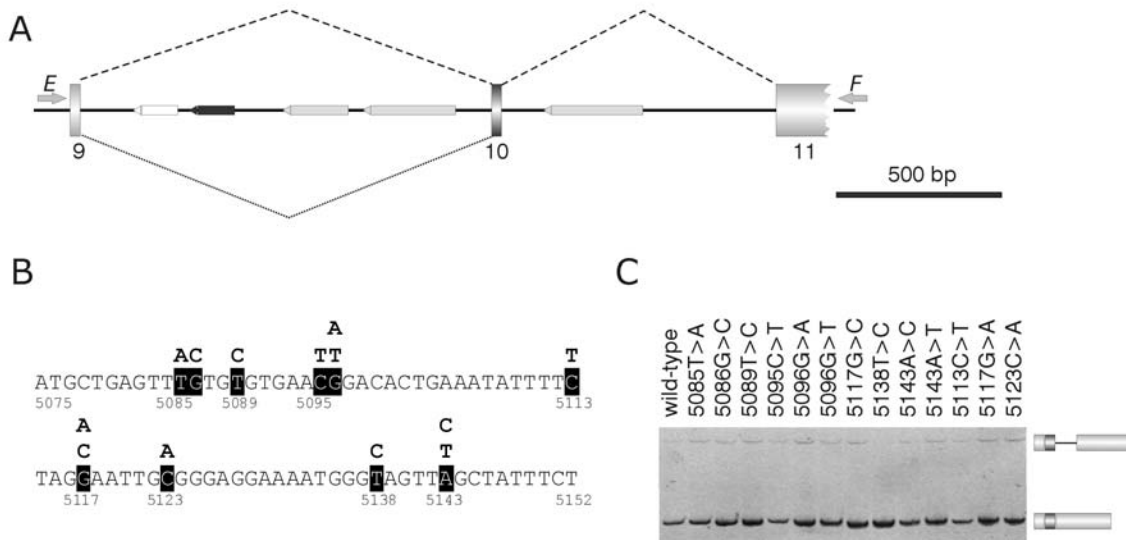


Figure 3 Systematic mutagenesis of a new splicing regulatory motif in *BRCA1* exon 6

Legend: **A**, exon 6 inclusion/skipping levels of the wild-type (WT) and 36 minigene mutants. Designation of substitutions (horizontal axis) corresponds to the numbering shown in Fig. 1B. The BIC-derived transversion is boxed. Errors 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**, Improved SF2/ASF ESE scores showed the highest correlation with exon inclusion levels. Full data are available in Supplementary Table 6.

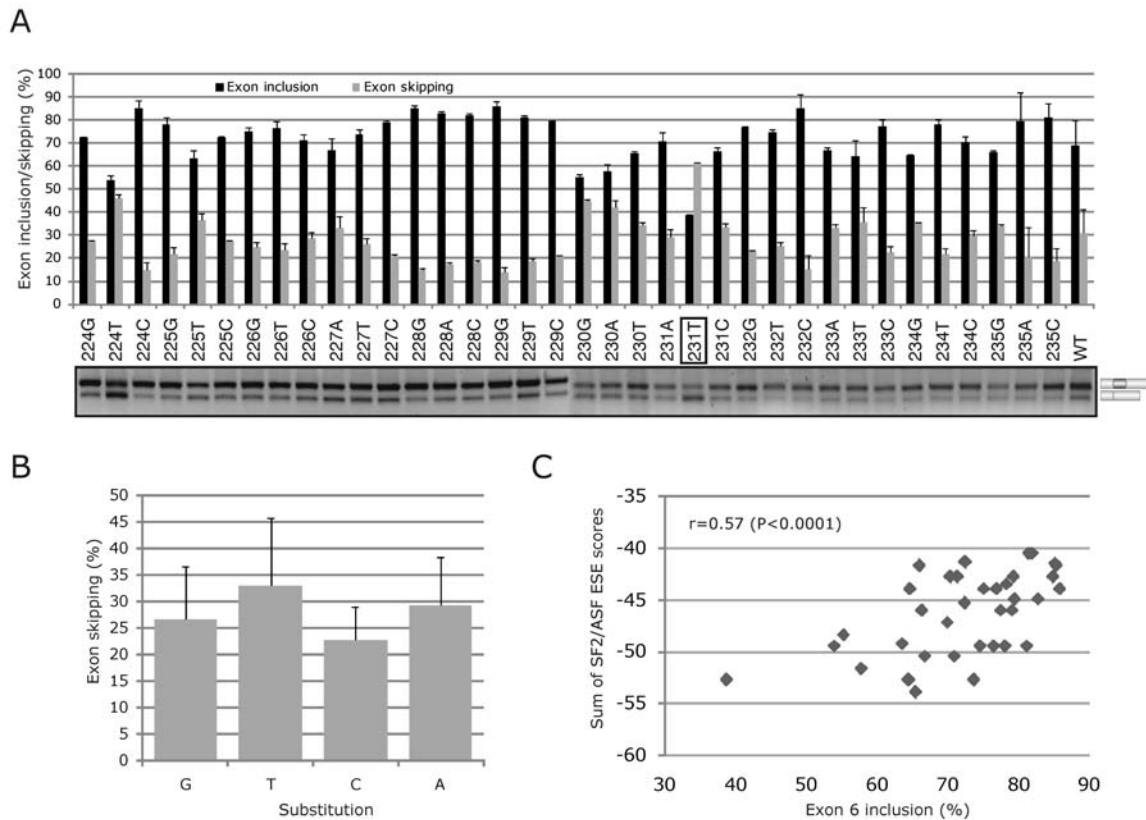


Figure 4 Predicted RNA secondary structure of the c.231G/T alleles

Legend: **A**, predicted secondary structure of the wild type (WT) and mutated (positions 231 and 238) pre-mRNA. Mutations are denoted by arrows. Secondary structure predictions were carried out using the algorithm implemented in the RNAStructure program (v. 4.6) [Mathews et al., 2004]. **B**, The splicing pattern of four mutated reporter constructs upon transient transfections in HEK293 cells. EI, exon inclusion.

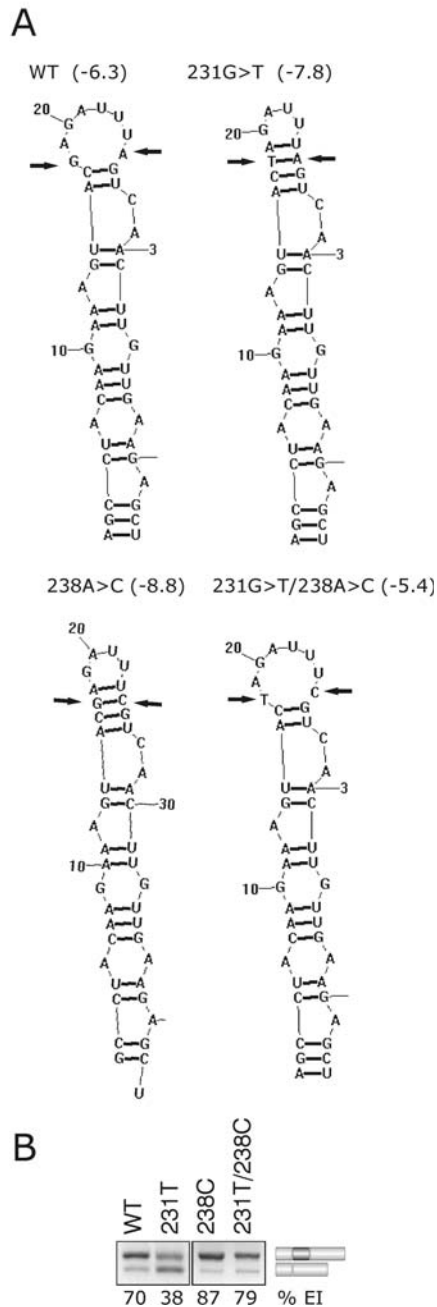


Figure 5 Prediction of exon-skipping substitutions by HOT-SKIP

Legend: Each point represents a disease-causing mutation that resulted in exon skipping (full details are in Supplementary Table 4). HOT-SKIP percentile was calculated as $(a-4)*400/n$, where n is exon length in nucleotides and a is the rank of the mutation among all possible exon substitutions as determined by sorting the ESS/ESE ratios.

