Rules and tools to predict the splicing effects of exonic and intronic mutations

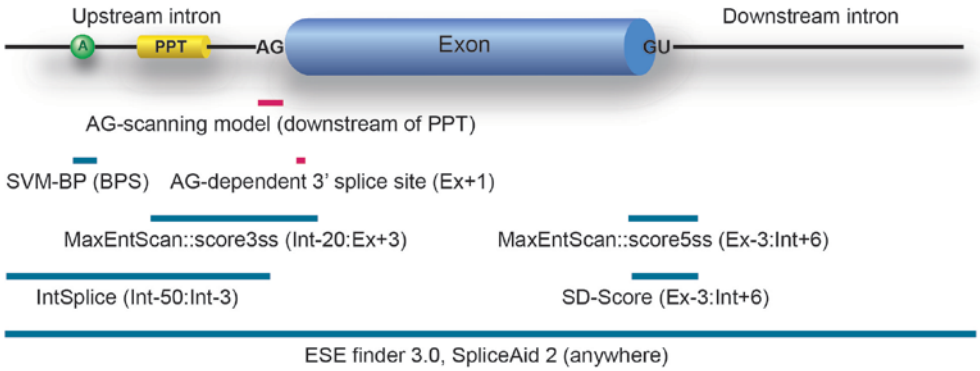
Ohno, Kinji*
Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan
ohnok@med.nagoya-u.ac.jp
No conflict of interest to declare
Takeda, Jun-ichi
Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan
No conflict of interest to declare
Masuda, Akio
Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan
No conflict of interest to declare

Abstract

Development of the next generation sequencing technologies has enabled detection of extensive arrays of germline and somatic single nucleotide variations (SNVs) in human diseases. The splicing effects of SNVs disrupting invariant intronic GT-AG dinucleotides can be readily predicted, but prediction of the splicing effects of the other SNVs is challenging. The splicing effects of SNVs generating cryptic AG or disrupting canonical AG, as well as copy number variations duplicating AG, at the 3' splice sites can be inferred from the AG-scanning model. Similarly, the splicing effects of SNVs affecting the first nucleotide "G" of an exon can be inferred from AG-dependence of the 3' splice site. The other SNVs require computational tools. Many different tools have been developed for SNVs affecting the 5' splice sites and auxiliary splicing *cis*-elements (exonic/intronic splicing enhancers/silencers, ESE/ESS/ISE/ISS). For SNVs affecting the 5' splice site, we prefer to use our SD-Score and MaxEntScan::score5ss. For SNVs affecting auxiliary splicing *cis*-elements, SpliceAid 2 is likely to be dependable in ~12 different tools. In contrast, only a single tool, the SVM-BP finder, is dedicated for prediction of the position of the branch point sequence (BPS). Similarly, the IntSplice, which we recently developed, is the only tool to predict the splicing effects of SNVs at intronic positions -50 to -3, where the BPS and the polypyrimidine tract (PPT) are located. Development of

efficient tools to predict aberrant splicing will facilitate our understanding of splicing pathomechanisms in human diseases.

Graphical/Visual Abstract and Caption



Representative rules (red) and tools (blue) to predict the splicing effects of exonic and intronic mutations. Ex, exonic position; Int, intronic position.

Introduction

We humans have evolved by acquiring tissue-specific and developmental stage-specific RNA metabolisms including alternative splicing and transcription of non-coding RNA, and not by acquiring novel coding genes. Among alternative RNA processing, alternative splicing is the most essential and the most extensively studied mechanisms to achieve proteome diversity. More than ~90% of transcripts are subject to alternative splicing,[1] which generate ~100,000 proteins from ~20,000 human genes.[2, 3] Tissue-specific and developmental stage-specific expressions of RNA-binding proteins (RBPs), which bind to specific *cis*-acting splicing elements on pre-mRNA, enable coordinated regulation of alternative and constitutive splicing.[4, 5] Single nucleotide variations (SNVs) that disrupt *cis*-acting splicing elements or comprise catalytic functions of *trans*-acting RBPs impair finely tuned alternative and constitutive splicing, and provoke pathological consequences.[6, 7]

Prediction of the splicing effects of SNVs affecting GT at the 5' splice sites and AG at the 3' splice sites is straightforward. On the contrary, prediction of the splicing effects of SNVs affecting the other nucleotides is challenging. Specific rules and tools to predict the splicing consequence of SNVs have been disclosed and developed, respectively, to help us understand the underlying aberrant splicing mechanisms. As many different tools with different algorithms are currently available, we indicate our favorite tools in each section in this review. Our preferences, however, are all based on

our personal experiences. Appropriateness of our preferences has not been extensively scrutinized or systemically validated.

Splicing machinery

Splicing of pre-mRNA is carried out in the nucleus by spliceosome, which is a macromolecular complex composed of RNA and proteins. Five small nuclear ribonucleoproteins (snRNPs) and multiple proteins (> 100) cooperate in spliceosome to catalyze the splicing reaction. Each snRNP is composed of a single uridine-rich small nuclear RNA (snRNA) and multiple proteins. Spliceosome-mediated splicing is achieved in two steps: (i) recognition of intron/exon boundary and (ii) catalysis of the transesterification reaction to excise out an intron followed by joining two exons. Recognition of intron/exon boundary is guided by essential splicing *cis*-elements close to either end of intron. These include the 5' splice site, the branch point sequence (BPS), the polypyrimidine tract (PPT), and the 3' splice site (Fig. 1a). In metazoans, especially in humans, these splicing *cis*-elements are highly degenerative, which weakens proper binding of essential *trans*-acting factors like U1snRNP, SF1, U2AF65, and U2AF35. As a consequence, multiple auxiliary *trans*-acting factors need to cooperate to form spliceosome and to favor folding of nuclear pre-mRNA to commit splicing. The assembly of spliceosome is further regulated by auxiliary splicing *cis*-elements either in a positive or negative manner (Fig. 1b). Positively modulating *cis*-elements are termed intronic/exonic splicing enhancers (ISEs/ESEs), whereas negatively modulating *cis*-elements are termed intronic/exonic splicing silencers (ISSs/ESSs). Most of these auxiliary *cis*-elements function by binding to the cognate RNA-binding proteins, whereas some *cis*-elements function by forming secondary structures. The majority of splicing *trans*-factors (RNA-binding proteins) for ESE are serine/arginine-rich (SR) proteins, which function as either an essential or regulatory factor.[8, 9] SR proteins can modulate several steps of spliceosome assembly through protein-protein and protein-RNA interaction.[10, 11] They possess one or two RNA-recognition motif (RRM) at the N-terminal end, and arginine and serine residues (RS domains) at the C-terminal end. The majority of splicing *trans*-factors (RNA-binding proteins) for splicing silencer elements (ISSs/ESSs) are heterogeneous nuclear ribonucleoproteins (hnRNPs).[12] Members of this family usually contain an RRM-type and KH-type RNA-binding domain along with other auxiliary domains to mediate protein-protein interaction. ISEs are less characterized compared to the other *cis*-elements. Recent analyses on ISEs suggest that hnRNP F, hnRNP H, NOVA1, NOVA2, FOX1, and FOX2 are candidate factors for ISEs.[13-16]

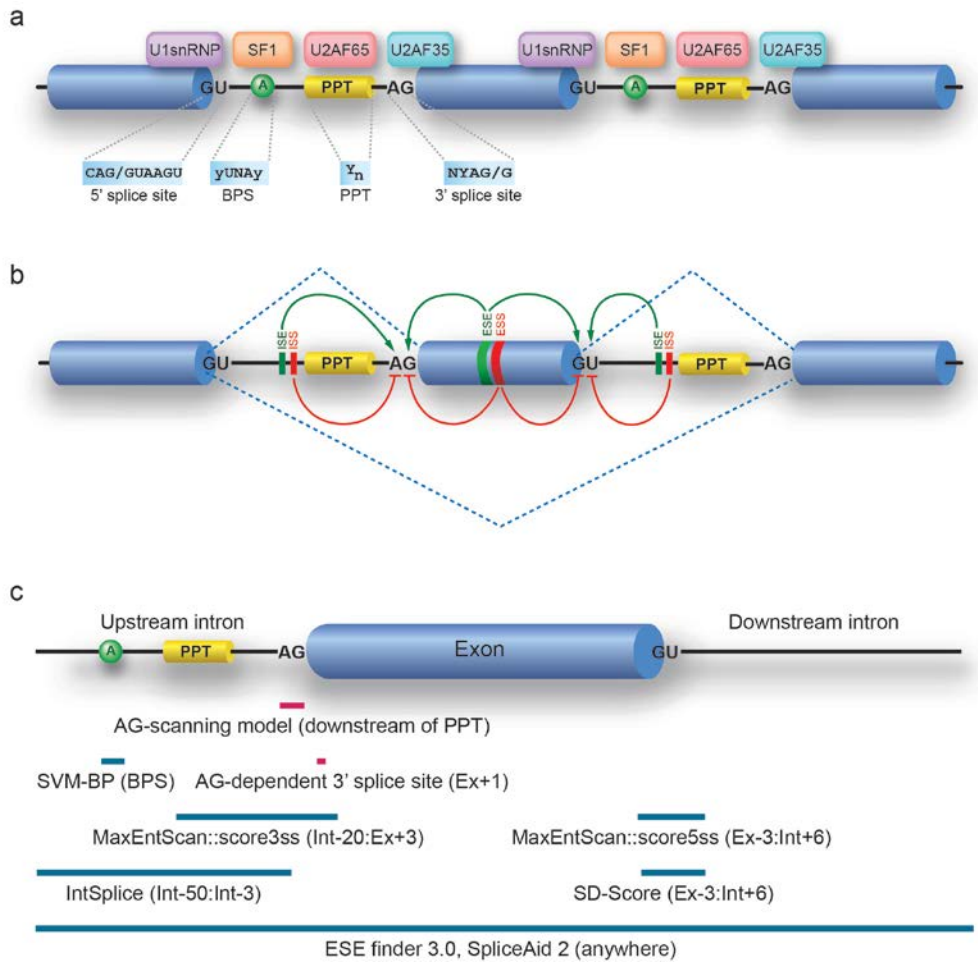


Figure 1. Essential and auxiliary splicing *cis*-elements. **(a)** Essential splicing *cis*-elements and the cognate splicing *trans*-factors to form spliceosome complex E. The 5' splice site (CAG/GUAAGU), the branch point sequence (BPS) (yUNAy), the polypyrimidine tract (PPT) (Y_n), and the 3' splice site (NYAG/G) are recognized by U1snRNP, SF1, U2AF65, and U2AF35, respectively, in the earliest step of spliceosome formation (E complex). Y = C/U and N = any nucleotide. **(b)** Auxiliary splicing *cis*-elements. Based on location and functional activity, these elements are categorized into intronic/exonic splicing enhancers (ISEs/ESEs) and intronic/exonic splicing silencers (ISSs/ESSs). Recognition of splice sites are promoted by enhancer elements and repressed by silencer elements. Alternative splicing is regulated by multiple auxiliary splicing *cis*-elements and not by a single element. Enhancers may antagonize the activity of silencers, and *vice versa*. Exon inclusion or skipping is finely regulated by the relative strength of these influential elements and the cognate splicing *trans*-factors. **(c)** Representative rules (red) and tools (blue) to predict the splicing effects of SNVs, which we prefer to use. Refer to the main text for the other tools. Int, intronic position; and Ex, exonic position.

Spliceosome assembly starts with the recognition of the 5' splice site by U1 snRNP, the BPS by SF1, the PPT by U2AF65, and the 3' terminal AG by U2AF35 (Fig. 1a). This initial assembly of spliceosome, which is termed E complex, is ATP-independent. Initial spliceosome assembly is usually formed across a single intron in a two-exon gene or short introns in multi-exon genes (termed intron-defined E complex), whereas it is formed across an exon flanked by long introns in higher metazoans (termed exon-defined E complex).[17-19] The ATP-independent E complex is then transformed to ATP-dependent spliceosome A complex, where SF1 is substituted for by U2 snRNP at the BPS. Subsequent recruitment of U4/U6-U5 snRNPs leads to the formation of B complex. Through extensive remodeling and conformational changes, an active spliceosome complex called C complex is formed by replacing U1 and U4 snRNPs, which subsequently catalyzes splicing. In most cases, splicing activators or repressors function by modulating the early spliceosome assembly at the stage of E complex or A complex. Therefore, the ultimate splicing consequence is determined by finely tuned coordination of *cis*-acting splicing elements and their cognate RNA-binding proteins. SNVs disrupting one of the *cis*-acting splicing elements can possibly cause aberrant splicing.[6, 7, 20]

SNVs affecting the 5' splice site

The consensus sequence of the 5' splice site is "CAG/GUAAGUAU", where "/" denotes an exon-intron boundary.[21, 22] This sequence is complementary to U1 snRNA. Several methods have been developed to evaluate the splicing strength of the 5' splice site including the Shapiro and Senapathy matrix,[23] the weight matrix model,[24] the information contents (R*i*)[25, 26] the first-order Markov model,[27] the maximum dependence decomposition model,[28] and the maximum entropy model (MaxEntScan)[29] These methods were developed to evaluate the strength of U1 snRNP-binding to the 5' splice site, and are not specifically designed for evaluating the effect of a mutation on aberrant splicing. These methods, however, have been successfully applied to predict the splicing effects of mutations at the 5' splice sites.[30-32] Additionally, algorithms to predict an activated cryptic 5' splice site due to a mutation at the 5' ss have been generated. An EST-based method, Cryptic Splice Finder,[33] uses the information contents at the 5' splice site.[25, 26] Calculation of free energy (ΔG) for stable binding of U1 snRNA to the 5' splice site also enables prediction of an activated cryptic 5' splice site.[34] Similarly, a structure-based method, the structure profiles and odds measure (SPO algorithm), was developed for identifying cryptic 5' splice sites.[35]

Dependence of these methods on prediction of binding of U1 snRNP to the 5' splice site causes the following problems. One is that we have the non-canonical 5' splice sites, where 0.9% of the 5' splice sites carry GC dinucleotide and 0.36% harbor the U12-type 5' splice sites with AT

dinucleotide.[36] Another problem is that base pairing is shifted by one nucleotide between the 5' splice site and U1 snRNA at 59 5' splice sites.[22] A subsequent report shows that as much as ~5% of the 5' splice sites are recognized by U1 snRNA with bulged nucleotides on either the 5' splice site or U1 snRNA.[37]

In order to predict a splicing consequence of a mutation at the 5' splice site, we have also analyzed splicing patterns of 31 mini-genes carrying naturally occurring or artificially introduced mutations at various nucleotides of the 5' splice site in cultured mammalian cells.[38] We proposed that a new parameter, the SD-score, efficiently predicts the splicing consequences of mutations at the 5' splice site. SD-score is a common logarithm of the frequency in the human genome of a specific 9-nucleotide sequence spanning 3 nucleotides at the 3' end of an exon and 6 nucleotides at the 5' end of an intron. The SD-score predicted the splicing consequences of the 31 mini-genes and 179 reported SNVs at the 5' splice site with the sensitivity of 97.1 % and specificity of 94.7 %. We simulated all possible SNVs at 189,249 5' splice sites in the human genome using this SD-score, and found that 37.8%, 88.8%, 96.8%, 82.8%, 95.0%, 96.5%, and 63.0% of mutations at exonic positions -3, -2, -1, and intronic position +3, +4, +5, +6, respectively, were predicted to cause aberrant splicing, which were all higher than we expected. Using the SD-score we indeed detected four mutations at the 5' splice site causing aberrant splicing, which had been overlooked (Fig. 2).[38] The SD-score web service program is available at http://www.med.nagoya-u.ac.jp/neurogenetics/SD_Score/sd_score.html. Among various parameters and tools at the 5' splice site, we prefer to use MaxEntScan::score5ss[29] and the SD-score[38] (Fig. 1c).

a

	Wt.Seq				Mt.Seq				Difference			
pos	Wt.Seq (9bases)	SD-Score	Ri	CV	Mt.Seq (9bases)	SD-Score	Ri	CV	ΔSD-Score	ΔRi	ΔCV	Prediction
4	CAGGTAAGT	-2.074	12.849	1.000	CAGGTACGT	-2.838	9.679	0.893	-0.764	-3.170	-0.107	Aberrant

b

Simulation of all possible mutations

pos.	-3	-2	-1	+1	+2	+3	+4	+5	+6
Seq.	C	A	G	G	T	A	A	G	T
A	Normal	-	Normal	-	-	-	-	Normal	Normal
C	-	Aberrant	Aberrant	-	-	Aberrant	Aberrant	Aberrant	Normal
G	Normal	-	Normal	-	-	Normal	Normal	-	Normal
T	Aberrant	Normal	Aberrant	-	-	Aberrant	Normal	Aberrant	-

Analysis of multiple 5' splice sites

Wt.Seq	SD-Score	Ri	CV	Mt.Seq	SD-Score	Ri	CV	ΔSD-Score	ΔRi	ΔCV	Prediction
CAGGTAAGT	-2.074	12.849	1.000	AAGGTAAGT	-2.059	12.736	0.995	0.015	-0.113	-0.005	Normal
CAGGTAAGT	-2.074	12.849	1.000	CAGGTAAGT	-2.074	12.849	1.000	0.000	0.000	0.000	Normal
CAGGTAAGT	-2.074	12.849	1.000	GAGGTAAGT	-2.247	11.886	0.970	-0.173	-0.963	-0.030	Normal
CAGGTAAGT	-2.074	12.849	1.000	TAGGTAAGT	-2.496	11.257	0.958	-0.422	-1.592	-0.042	Aberrant
CAGGTAAGT	-2.074	12.849	1.000	CAGGTAAGT	-2.074	12.849	1.000	0.000	0.000	0.000	Normal
CAGGTAAGT	-2.074	12.849	1.000	CCGGTAAGT	-2.895	10.310	0.909	-0.821	-2.539	-0.091	Aberrant
CAGGTAAGT	-2.074	12.849	1.000	CGGGTAAGT	-2.836	10.393	0.910	-0.762	-2.456	-0.090	Aberrant
CAGGTAAGT	-2.074	12.849	1.000	CTGGTAAGT	-2.272	10.662	0.914	-0.198	-2.187	-0.086	Normal

Figure 2. SD-Score web service program at http://www.med.nagoya-u.ac.jp/neurogenetics/SD_Score/sd_score.html. The program accepts wild-type and mutant 5' splice site sequences spanning 3 nucleotides at the 3' end of an exon and 6 nucleotides at the 5' end of an intron. **(a)** The program returns differences in the SD-Score (ΔSD-Score), information contents (Δ*Ri*), and position-specific scoring matrix (ΔCV), as well as the predicted splicing consequence. **(b)** Clicking on the wild-type 9-nt sequence in **(a)** shows simulation of all possible mutations.

SNVs affecting AG dinucleotide at the 3' splice site

The 3' splice site comprises the 3' end of an intron and the 5' end of an exon with a consensus sequence of "NYAG/G" (N, any nucleotide; Y, pyrimidine; R, purine; "/", intron-exon boundary).[39] A stretch of pyrimidines in the PPT immediately precedes "NYAG/G". U2AF35 binds to this sequence in an early step of spliceosome formation (E complex). Extensive mutational analyses of nucleotides surrounding the 3' splice site of *FAS* revealed that U2AF35-binding sequences are highly degenerative and auxiliary nucleotides are essential for binding of U2AF35.[40] A chromatin- and RNA-associated protein, DEK, facilitates the U2AF35-AG interaction and prevents binding of

U2AF65 to a pyrimidine tract not followed by AG.[41] HnRNPA1 also discriminates authentic 3' splice sites from pyrimidine-rich RNA segments by recognizing AG at the 3' end of an intron.[42] In later steps of splicing, the dinucleotide AG is scanned downstream of the branch point and the first AG is recognized as the 3' end of the intron, which is called the "AG-scanning model".[43] When the upstream cryptic 3' splice site is more than 24 nucleotides apart from the authentic 3' splice site, the authentic 3' splice site is used; less than 21 nucleotides, the cryptic 3' splice site is used; between 21 and 24 nucleotides, both cryptic and authentic 3' splice sites are used.[44, 45] Three exceptions against the AG-scanning model have been reported: (i) an AG dinucleotide less than 13 nucleotides downstream of the branch point is not recognized, likely due to steric hindrance of *trans*-acting factors; (ii) the first AG dinucleotide is hidden in a stable secondary structure; and (iii) two AG dinucleotides that are less than 12 nucleotide apart compete for being recognized.[46-48] Similarly, a decision tree optimized by the ID3 algorithm was developed for the AG-scanning model to an accuracy of 95.5%. This decision tree includes four steps: (i) whether the AG of interest is the first one after the branch point or not, (ii) whether the first AG is more than 7 nucleotides downstream the branch point or not, (iii) whether the distance between the first and next AG is more than 14 nucleotides or not, (iv) whether the AG decided is within an optimal range from the branch point (16-25 nucleotides).[49] These rules and algorithms are similar, but details are inconsistent, which is likely due to difference in the other splicing *cis*-elements in different genes.

The splicing effects of mutations at the 3' splice site can be inferred from the AG-scanning model (Fig. 1c). We[32] and others[50, 51] reported duplication of the 3' splice site spanning 16 to 69 nucleotides and including the "NYAG/G" motif. All these mutants gave rise to activation of an upstream copy of the 3' splice sites.[32, 50, 51] Similarly, we[52] and others[53] reported that a mutation generating *de novo* AG or disrupting the canonical AG also follows the AG-scanning model. On the contrary, a mutation of the *BTD* gene encoding biotinidase creates a downstream 3' splice site with a stronger PPT and causes usage of the downstream AG, which does not follow the AG-scanning model.[54]

SNVs affecting the first nucleotide "G" of exon

Introns with a long stretch of PPT strongly bind to U2AF65 and does not require binding of U2AF35, which is called the "AG-independent 3' splice site".[55-58] In contrast, introns with a short or degenerative PPT require both U2AF65 and U2AF35 for this substitution, which is called the "AG-dependent 3' splice site". Site-specific crosslinking revealed that the longest stretch of the U2AF35-binding site spans intronic position -3 and exonic position +12.[55] The SELEX motif for U2AF35 is AG/G (/, intron-exon boundary).[55]

Disease-causing mutations may affect the first nucleotide "G" of exon, but their effects on splicing have been rarely scrutinized[20]. Three such mutations in *FECH*,[59] *GH1*,[60] and *EYA1*[61] cause aberrant splicing, whereas two other mutations in *LPL*[62] and *HEXA*[63] have no effect on splicing. We demonstrated that an SNV at the first nucleotide "G" of exon causes aberrant splicing only at the AG-dependent 3' splice site.[64] When the length of a complete stretch of pyrimidines at the PPT is 10 to 15 nucleotides or more, an SNV at the first nucleotide "G" of exon is immune to splicing errors. We analyzed aberrant splicing in nine disease-causing mutations at the first nucleotide "G" of exon, which were reported to be missense mutations. We found that five of them, which had short PPTs, caused aberrant splicing.[64] AG-dependence of the 3' splice site efficiently predicts the splicing effects of an SNV at the first nucleotide "G" of exon (Fig. 1c). This rule, however, is applicable only when the wild-type nucleotide at the 5' end of an exon is "G".

SNVs affecting the branch point sequence (BPS)

Yeast carries a strictly conserved BPS of "UACUAAC".[65] In eukaryotes, however, the BPS had not been evaluated through global experimental verification. Previously reported consensus sequences of eukaryote BPS were either based on analyses of a few experimentally identified branch points or *in silico* search for sequences which resemble the yeast consensus BPS. We sequenced 367 lariat RT-PCR clones arising from 52 introns of 20 housekeeping genes, and identified that the human consensus BPS is "yUnAy" (y, pyrimidine; n, any nucleotide).[66] The fourth nucleotide "A" was the branch point (position +0) and was conserved in 92.3% of the clones. The "U" at position -2 was conserved in 74.6% of the clones. Screening of 1.2 billion total RNA reads for unconventional inverted alignment disclosed 861 reads from lariats arising from 759 U2-dependent introns and revealed that the consensus BPS is "URAY: (Y, pyrimidine).[49] Similar but novel approach, named "LaSSO", has been developed and identified that the human consensus BPS is "UNA".[67] Exoribonuclease (RNase R) digestion and targeted RNA-sequencing enabled detection of 59,359 human BPS in > 10,000 genes, which revealed two BPS motifs: (i) CUNAN (cysteine B-box) is associated with cytosine-rich PPTs, and with high GC% introns and downstream exons; and (ii) UUNAN (uracil B-box) is associated with uracil-rich PPTs, and with low GC% introns and downstream exons.[68]

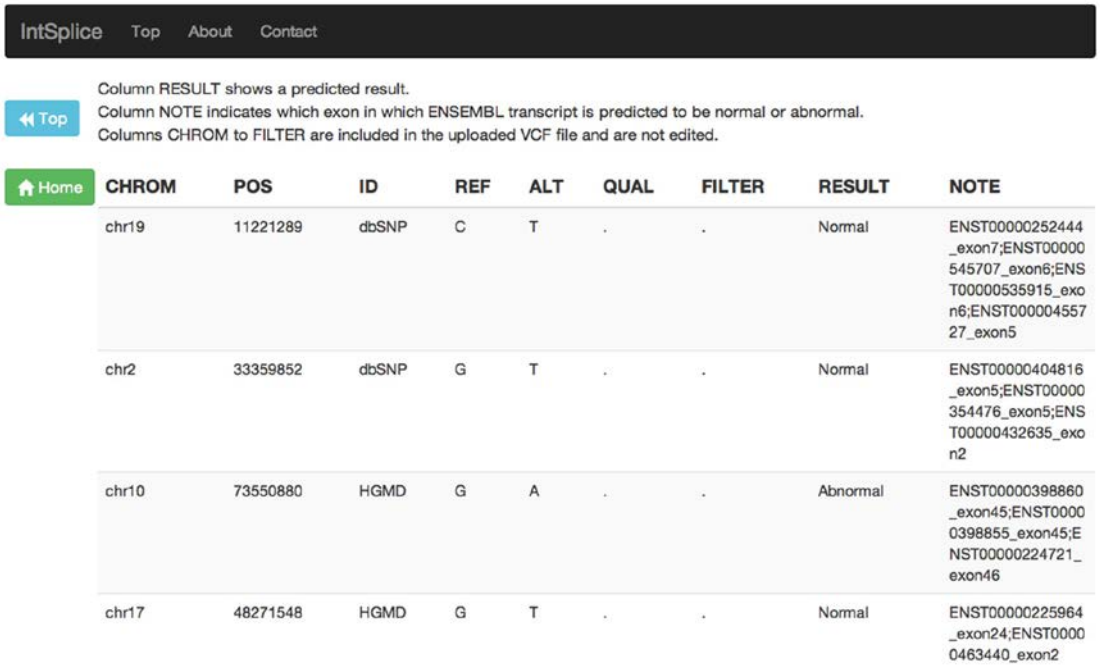
Although the extracted BPS motifs are slightly different between the studies, all BPS motifs carry U at position -2 and A at position +0. Indeed, previously identified BPS-disrupting mutations affect either U at position -2 or A at position 0.[20, 66] An SF1 affinity model, which exploits validated SF1 affinity data, efficiently predicts BPS in putative SF1-dependent introns,[69] although depletion of SF1 does not abrogate splicing in HeLa nuclear extract.[70] A computational tool to predict BPS, the

SVM-BP finder (http://regulatorygenomics.upf.edu/Software/SVM_BP/), had been developed (Fig. 1c).[71]

SNVs affecting the PPT and the BPS

Degeneracy of the human BPS supports the notion that the human BPS is likely to be recognized along with the downstream PPT where U2AF65 binds and additionally with the invariant AG dinucleotide at the 3' splice site where U2AF35 binds.[72, 73] In PPT, uridines are preferred over cytidines.[74, 75] In addition, PPT with 11 continuous uridines is highly competent and the position of such PPT is not critical.[75]

SNVs affecting the PPT possibly compromise splicing, but such SNVs that had been experimentally proven are as rare as those for the BPS.[20, 76] No efficient tool had been available to predict the splicing effect of an SNV affecting the PPT. In an effort to detect splicing SNVs affecting the PPT and the BPS, we developed a support vector machine (SVM) tool to predict splicing SNVs at intronic position -50 to -3.[77] Following an effect-size analysis of each intronic nucleotide on annotated alternative splicing, we extracted 110 features that potentially dictate the strength of the splicing signals. We generated SVM models using the 110 features to differentiate pathogenic mutations in the human gene mutation database (HGMD) and normal polymorphisms in the dbSNP database. We obtained models with a sensitivity of 0.800 ± 0.041 (mean and SD) and a specificity of 0.849 ± 0.021. We generated a web service program, IntSplice (http://www.med.nagoya-u.ac.jp/neurogenetics/IntSplice) (Fig. 3). IntSplice is applicable to any SNVs at intronic positions -50 to -3 (Fig. 1c), which may affect the BPS/PPT, *de novo* create AG dinucleotide (stated above), or disrupt intronic splicing enhancers/silencers (stated below). We applied IntSplice to a naturally occurring and nine artificial intronic mutations in *RAPSN* causing congenital myasthenic syndrome. IntSplice correctly predicted the splicing consequences of nine of the ten mutants at intronic positions -11 to -3.[77]



CHROM	POS	ID	REF	ALT	QUAL	FILTER	RESULT	NOTE
chr19	11221289	dbSNP	C	T	.	.	Normal	ENST00000252444_exon7;ENST00000545707_exon6;ENST00000535915_exon6;ENST00000455727_exon5
chr2	33359852	dbSNP	G	T	.	.	Normal	ENST00000404816_exon5;ENST00000354476_exon5;ENST00000432635_exon2
chr10	73550880	HGMD	G	A	.	.	Abnormal	ENST00000398860_exon45;ENST00000398855_exon45;ENST00000224721_exon46
chr17	48271548	HGMD	G	T	.	.	Normal	ENST00000225964_exon24;ENST00000463440_exon2

Figure 3. IntSplice web service program at http://www.med.nagoya-u.ac.jp/neurogenetics/IntSplice. The program accepts a VCF file, which includes seven features from CHROM to FILTER according to the VCF file format. Among these, CHROM, POS, REF, and ALT are required for IntSplice. Predicted results are shown in the 'RESULT' column. The rightmost 'NOTE' column indicates which exon in which ENSEMBL transcript is predicted to be abnormally or normally spliced.

SNVs affecting the exonic splicing enhancers/silencers (ESE/ESS) and intronic splicing enhancers/silencers (ISE/ISS)

Pathogenic missense mutations in *MSH1* and *MLH2* affect predicted ESE more frequently than normal SNPs. Comparison of the observed and expected frequencies of missense mutations at ESE predicts that at least 16-20% of missense mutation possibly affect splicing.[78] Similarly, comparison of the effects of mutations in the human gene mutation database (HGMD) and normal SNPs in the dbSNP database on ESE/ESS hexamers and subsequent validation of some *de novo* generated ESSs revealed that 7,154 out of 27,681 missense and nonsense mutations (26%) are likely to cause exon skipping.[79] Alternative inclusion of exon P3A of *CHRNA1* generates a nonfunctional AChR α1 subunit in human. We reported that skipping of exon P3A is achieved by binding of PTPB1 and hnRNP H to an ISS in the preceding intron 2, as well as by binding of hnRNP L to an ESS in exon P3A.[80-82] In congenital myasthenic syndrome, a mutation disrupting the ISS[80] and another mutation disrupting the ESS[82] cause exclusive inclusion of nonfunctional exon P3A. Similarly,

p.E415G in *COLQ* in congenital myasthenic syndrome[83] causes exon skipping by mutating GGAGG to GGGGG, which displaces splicing-enhancing SRSF1 and *de novo* gains binding of splicing-suppressing hnRNP H.[84] We identified that these mutations were splicing mutations, because functional assays of mutant proteins or cells expressing mutant proteins exhibited no abnormality. However, it is challenging to predict mutations that affect splicing *cis*-elements (ESE/ESS/ISE/ISS), because most of them are highly degenerative.

A variety of tools to predict splicing *cis*-elements have been developed: ESE finder 3.0,[85] FAS-ESS,[86] PESXs,[87, 88] RESCUE-ESE,[89] ESRsearch,[90] Human Splicing Finder 3,[91] CRYP-SKIP,[92] SFmap,[93]SpliceAid,[94] Spliceman,[95] and SpliceAid2,[96] and RegRNA 2.0.[97] Features of these tools are summarized in Table 1. According to our personal experiences, we prefer to use SpliceAid2 to predict splicing *cis*-elements and the cognate RNA-binding. For SRSF1, we suppose that ESE finder 3.0 is also dependable (Fig. 1c). As stated in the introduction, our preferences are all based on our personal experiences and the validity of our preferences has not been systemically scrutinized.

Table 1. Web service programs to predict splicing *cis*-elements (ESE/ESS/ISE/ISS)

Tool	URL	Features
ESE finder 3.0[85]	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home	Functional SELEX motifs
FAS-ESS[86]	http://genes.mit.edu/fas-ess/	Functional screening of 6-nt motifs for ESS
PESXs[87, 88]	http://cubio.biology.columbia.edu/pesx/pesx/	Extensive search for 8-nt motifs in noncoding exons
RESCUE-ESE[89]	http://genes.mit.edu/burgelab/rescue-ese/	Extensive search for 6-nt motifs in exons with weak splice sites

ESRsearch[90]	http://ibis.tau.ac.il/ssat/ESR.htm	Based on conservation of wobble positions between human and mouse exons
Human Splicing Finder 3[91]	http://www.umd.be/HSF3/	Integration of previously reported matrices of splicing *cis*-elements
CRYP-SKIP[92]	http://cryp-skip.img.cas.cz	Based on a dataset of 250 skipped exons and 204 activated cryptic splice sites due to disease-causing mutations
SFmap[93]	http://sfmap.technion.ac.il	A tool for motif analysis and prediction of splicing factors
SpliceAid[94]	http://www.introni.it/splicing.html	A database of experimentally proven RNA target motifs bound by splicing proteins in human
Spliceman[95]	http://fairbrother.biomed.brown.edu/spliceman/	Positional distribution of splicing elements

SpliceAid2[96]	http://www.introni.it/spliceaid.html	A database of tissue-specific human splicing factors and RNA target motifs
RegRNA 2.0[97]	http://regrna.mbc.nctu.edu.tw/php/browse.php	Computational integration of previously published motif databases

Conclusion

Aberrant splicing due to mutations affecting invariant GT-AG dinucleotides has been well known, since the emergence of nucleotide sequencing in human diseases. Accurate prediction of the splicing effects of SNVs affecting the other nucleotides at the 5' and 3' splice sites, as well as the other essential splicing *cis*-elements (BPS and PPT), is still challenging. SNVs may affect auxiliary splicing *cis*-elements (ESE/ESS/ISE/ISS), prediction of which, however, is also challenging. A large number of tools have been developed to decode hidden splicing code in human diseases. Development of novel tools will still proceed, but it is also essential to consider the possibility of aberrant splicing for any SNVs.

Acknowledgements

Works done in the author' laboratory were supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT); Ministry of Health, Labour and Welfare (MHLW); and Japan Agency for Medical Research and Development (AMED) of Japan.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008, 456:470-476.
2. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* 2010, 11:345-355.
3. Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* 2013, 10:133-139.
4. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 2003, 72:291-336.
5. Shi Y, Manley JL. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev* 2015, 29:889-897.
6. Rahman MA, Nasrin F, Masuda A, Ohno K. Decoding abnormal splicing code in human diseases. *J Investig Genomics* 2015, 2:00016.
7. Rahman MA, Ohno K. Splicing aberrations in congenital myasthenic syndromes. *J Investig Genomics* 2015, 2:00038.
8. Graveley BR. Sorting out the complexity of SR protein functions. *RNA* 2000, 6:1197-1211.
9. Lin S, Fu XD. SR proteins and related factors in alternative splicing. *Adv Exp Med Biol* 2007, 623:107-122.
10. Shen H, Green MR. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell* 2004, 16:363-373.
11. Shen H, Kan JL, Green MR. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell* 2004, 13:367-376.
12. Martinez-Contreras R, Cloutier P, Shkreta L, Fisette JF, Revil T, Chabot B. hnRNP proteins and splicing control. *Adv Exp Med Biol* 2007, 623:123-147.
13. Wang Y, Ma M, Xiao X, Wang Z. Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat Struct Mol Biol* 2012, 19:1044-1052.

14. Mauger DM, Lin C, Garcia-Blanco MA. hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. *Mol Cell Biol* 2008, 28:5403-5419.
15. Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. An RNA map predicting Nova-dependent splicing regulation. *Nature* 2006, 444:580-586.
16. Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* 2009, 16:130-137.
17. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 2009, 136:701-718.
18. House AE, Lynch KW. Regulation of alternative splicing: more than just the ABCs. *J Biol Chem* 2008, 283:1217-1221.
19. Schneider M, Will CL, Anokhina M, Tazi J, Urlaub H, Luhrmann R. Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol Cell* 2010, 38:223-235.
20. Ohe K, Masuda A, Ohno K. Intronic and exonic nucleotide variations that affect RNA splicing in humans. In: *Genomics I – Humans, Animals and Plants*: iConcept Press; 2012.
21. Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. Are snRNPs involved in splicing? *Nature* 1980, 283:220-224.
22. Roca X, Krainer AR. Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. *Nat Struct Mol Biol* 2009, 16:176-182.
23. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987, 15:7155-7174.
24. Staden R. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res* 1984, 12:505-519.
25. Rogan PK, Faux BM, Schneider TD. Information analysis of human splice site mutations. *Hum Mutat* 1998, 12:153-171.
26. Nalla VK, Rogan PK. Automated splicing mutation analysis by information theory. *Hum Mutat* 2005, 25:334-342.
27. Salzberg SL. A method for identifying splice sites and translational start sites in eukaryotic mRNA. *Comput Appl Biosci* 1997, 13:365-376.

28. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997, 268:78-94.
29. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004, 11:377-394.
30. Eng L, Coutinho G, Nahas S, Yeo G, Tanouye R, Babaei M, Dork T, Burge C, Gatti RA. Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Hum Mutat* 2004, 23:67-76.
31. Ohno K, Brengman J, Engel A. How does an A-to-G splice donor site mutation at position+3 result in aberrant splicing? A lesson learned from a mutation in the COLQ gene. *Am J Hum Genet* 1999, 65:A80-A80.
32. Ohno K, Tsujino A, Shen XM, Milone M, Engel AG. Spectrum of splicing errors caused by CHRNE mutations affecting introns and intron/exon boundaries. *J Med Genet* 2005, 42:e53.
33. Kapustin Y, Chan E, Sarkar R, Wong F, Vorechovsky I, Winston RM, Tatusova T, Dibb NJ. Cryptic splice sites and split genes. *Nucleic Acids Res* 2011, 39:5837-5844.
34. Roca X, Sachidanandam R, Krainer AR. Determinants of the inherent strength of human 5' splice sites. *RNA* 2005, 11:683-698.
35. Tsai KN, Wang D. Identification of activated cryptic 5' splice sites using structure profiles and odds measure. *Nucleic Acids Res* 2012, 40:e73.
36. Stein A, Weber G, Wahl MC, Jahn R. Helical extension of the neuronal SNARE complex into the membrane. *Nature* 2009, 460:525-528.
37. Roca X, Akerman M, Gaus H, Berdeja A, Bennett CF, Krainer AR. Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev* 2012, 26:1098-1109.
38. Sahashi K, Masuda A, Matsuura T, Shinmi J, Zhang Z, Takeshima Y, Matsuo M, Sobue G, Ohno K. In vitro and in silico analysis reveals an efficient algorithm to predict the splicing consequences of mutations at the 5' splice sites. *Nucleic Acids Res* 2007, 35:5995-6003.
39. Zhang MQ. Statistical features of human exons and their flanking regions. *Hum Mol Genet* 1998, 7:919-932.
40. Corrionero A, Raker VA, Izquierdo JM, Valcarcel J. Strict 3' splice site sequence requirements for U2 snRNP recruitment after U2AF binding underlie a genetic defect leading to autoimmune disease. *RNA* 2011, 17:401-411.

41. Soares LM, Zanier K, Mackereth C, Sattler M, Valcarcel J. Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. *Science* 2006, 312:1961-1965.
42. Tavanez JP, Madl T, Kooshapur H, Sattler M, Valcarcel J. hnRNP A1 proofreads 3' splice site recognition by U2AF. *Mol Cell* 2012, 45:314-329.
43. Smith CW, Porro EB, Patton JG, Nadal-Ginard B. Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature* 1989, 342:243-247.
44. Wieringa B, Hofer E, Weissmann C. A minimal intron length but no specific internal sequence is required for splicing the large rabbit beta-globin intron. *Cell* 1984, 37:915-925.
45. Chua K, Reed R. An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. *Mol Cell Biol* 2001, 21:1509-1514.
46. Blasband AJ, Rogers KT, Chen XR, Azizkhan JC, Lee DC. Characterization of the rat transforming growth factor alpha gene and identification of promoter sequences. *Mol Cell Biol* 1990, 10:2111-2121.
47. Smith CWJ, Chu TT, Nadalginard B. Scanning and Competition between Ags Are Involved in 3' Splice-Site Selection in Mammalian Introns. *Mol Cell Biol* 1993, 13:4939-4952.
48. Chua K, Reed R. The RNA splicing factor hSlu7 is required for correct 3 ' splice-site choice. *Nature* 1999, 402:207-210.
49. Taggart AJ, DeSimone AM, Shih JS, Filloux ME, Fairbrother WG. Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. *Nat Struct Mol Biol* 2012, 19:719-721.
50. Dlott B, d'Azzo A, Quon DV, Neufeld EF. Two mutations produce intron insertion in mRNA and elongated beta-subunit of human beta-hexosaminidase. *J Biol Chem* 1990, 265:17921-17927.
51. Bianchi P, Zanella A, Alloisio N, Barosi G, Bredi E, Pelissero G, Zappa M, Vercellati C, Baronciani L, Delaunay J, et al. A variant of the EPB3 gene of the anti-Lepore type in hereditary spherocytosis. *Br J Haematol* 1997, 98:283-288.
52. Selcen D, Shen XM, Milone M, Brengman J, Ohno K, Deymeer F, Finkel R, Rowin J, Engel AG. GFPT1-myasthenia: clinical, structural, and electrophysiologic heterogeneity. *Neurology* 2013, 81:370-378.

53. Sabbagh A, Pasmant E, Imbard A, Luscan A, Soares M, Blanche H, Laurendeau I, Ferkal S, Vidaud M, Pinson S, et al. NF1 molecular characterization and neurofibromatosis type I genotype-phenotype correlation: the French experience. *Hum Mutat* 2013, 34:1510-1518.
54. Pomponio RJ, Reynolds TR, Mandel H, Admoni O, Melone PD, Buck GA, Wolf B. Profound biotinidase deficiency caused by a point mutation that creates a downstream cryptic 3' splice acceptor site within an exon of the human biotinidase gene. *Hum Mol Genet* 1997, 6:739-745.
55. Wu S, Romfo CM, Nilsen TW, Green MR. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 1999, 402:832-835.
56. Guth S, Martinez C, Gaur RK, Valcarcel J. Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). *Mol Cell Biol* 1999, 19:8263-8271.
57. Guth S, Tange TO, Kellenberger E, Valcarcel J. Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol Cell Biol* 2001, 21:7673-7681.
58. Pacheco TR, Coelho MB, Desterro JM, Mollet I, Carmo-Fonseca M. In vivo requirement of the small subunit of U2AF for recognition of a weak 3' splice site. *Mol Cell Biol* 2006, 26:8183-8190.
59. Wang X, Poh-Fitzpatrick M, Chen T, Malavade K, Carriero D, Piomelli S. Systematic screening for RNA with skipped exons--splicing mutations of the ferrochelatase gene. *Biochim Biophys Acta* 1995, 1271:358-362.
60. Takahashi I, Takahashi T, Komatsu M, Sato T, Takada G. An exonic mutation of the GH-1 gene causing familial isolated growth hormone deficiency type II. *Clin Genet* 2002, 61:222-225.
61. Okada M, Fujimaru R, Morimoto N, Satomura K, Kaku Y, Tsuzuki K, Nozu K, Okuyama T, Iijima K. EYA1 and SIX1 gene mutations in Japanese patients with branchio-oto-renal (BOR) syndrome and related conditions. *Pediatr Nephrol* 2006, 21:475-481.
62. Ikeda Y, Takagi A, Nakata Y, Sera Y, Hyoudou S, Hamamoto K, Nishi Y, Yamamoto A. Novel compound heterozygous mutations for lipoprotein lipase deficiency. A G-to-T transversion at the first position of exon 5 causing G154V missense mutation and a 5' splice site mutation of intron 8. *J Lipid Res* 2001, 42:1072-1081.
63. Petroulakis E, Cao Z, Clarke JT, Mahuran DJ, Lee G, Triggs-Raine B. W474C amino acid substitution affects early processing of the alpha-subunit of beta-hexosaminidase

A and is associated with subacute G(M2) gangliosidosis. *Hum Mutat* 1998, 11:432-442.

64. Fu Y, Masuda A, Ito M, Shinmi J, Ohno K. AG-dependent 3'-splice sites are predisposed to aberrant splicing due to a mutation at the first nucleotide of an exon. *Nucleic Acids Res* 2011, 39:4396-4404.
65. Langford CJ, Gallwitz D. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. *Cell* 1983, 33:519-527.
66. Gao K, Masuda A, Matsuura T, Ohno K. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res* 2008, 36:2257-2267.
67. Bitton DA, Rallis C, Jeffares DC, Smith GC, Chen YY, Codlin S, Marguerat S, Bahler J. LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq. *Genome Res* 2014, 24:1169-1179.
68. Mercer TR, Clark MB, Andersen SB, Brunck ME, Haerty W, Crawford J, Taft RJ, Nielsen LK, Dinger ME, Mattick JS. Genome-wide discovery of human splicing branchpoints. *Genome Res* 2015, 25:290-303.
69. Pastuszak AW, Joachimiak MP, Blanchette M, Rio DC, Brenner SE, Frankel AD. An SF1 affinity model to identify branch point sequences in human introns. *Nucleic Acids Res* 2011, 39:2344-2356.
70. Tanackovic G, Kramer A. Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo. *Mol Biol Cell* 2005, 16:1366-1377.
71. Corvelo A, Hallegger M, Smith CW, Eyras E. Genome-wide association between branch point properties and alternative splicing. *PLoS Comput Biol* 2010, 6:e1001016.
72. Zorio DA, Blumenthal T. Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature* 1999, 402:835-838.
73. Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* 1999, 402:838-841.
74. Mullen MP, Smith CW, Patton JG, Nadal-Ginard B. Alpha-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes Dev* 1991, 5:642-655.
75. Coolidge CJ, Seely RJ, Patton JG. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res* 1997, 25:888-896.
76. Lewandowska MA. The missing puzzle piece: splicing mutations. *Int J Clin Exp Pathol* 2013, 6:2675-2682.

77. Shibata A, Okuno T, Rahman MA, Azuma Y, Takeda J, Masuda A, Selcen D, Engel AG, Ohno K. IntSplice: prediction of the splicing consequences of intronic single-nucleotide variations in the human genome. *J Hum Genet* 2016, 61:633-640.
78. Gorlov IP, Gorlova OY, Frazier ML, Amos CI. Missense mutations in hMLH1 and hMSH2 are associated with exonic splicing enhancers. *Am J Hum Genet* 2003, 73:1157-1161.
79. Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res* 2011, 21:1563-1571.
80. Masuda A, Shen XM, Ito M, Matsuura T, Engel AG, Ohno K. hnRNP H enhances skipping of a nonfunctional exon P3A in CHRNA1 and a mutation disrupting its binding causes congenital myasthenic syndrome. *Hum Mol Genet* 2008, 17:4022-4035.
81. Bian Y, Masuda A, Matsuura T, Ito M, Okushin K, Engel AG, Ohno K. Tannic acid facilitates expression of the polypyrimidine tract binding protein and alleviates deleterious inclusion of CHRNA1 exon P3A due to an hnRNP H-disrupting mutation in congenital myasthenic syndrome. *Hum Mol Genet* 2009, 18:1229-1237.
82. Rahman MA, Masuda A, Ohe K, Ito M, Hutchinson DO, Mayeda A, Engel AG, Ohno K. HnRNP L and hnRNP LL antagonistically modulate PTB-mediated splicing suppression of CHRNA1 pre-mRNA. *Sci Rep* 2013, 3:2931.
83. Kimbell LM, Ohno K, Engel AG, Rotundo RL. C-terminal and heparin-binding domains of collagenic tail subunit are both essential for anchoring acetylcholinesterase at the synapse. *J Biol Chem* 2004, 279:10997-11005.
84. Rahman MA, Azuma Y, Nasrin F, Takeda J, Nazim M, Bin Ahsan K, Masuda A, Engel AG, Ohno K. SRSF1 and hnRNP H antagonistically regulate splicing of COLQ exon 16 in a congenital myasthenic syndrome. *Sci Rep* 2015, 5:13208.
85. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003, 31:3568-3571.
86. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. *Cell* 2004, 119:831-845.
87. Zhang XH, Chasin LA. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* 2004, 18:1241-1250.
88. Zhang XH, Kangsamaksin T, Chao MS, Banerjee JK, Chasin LA. Exon inclusion is dependent on predictable exonic splicing enhancers. *Mol Cell Biol* 2005, 25:7323-7332.

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

90. Goren A, Ram O, Amit M, Keren H, Lev-Maor G, Vig I, Pupko T, Ast G. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol Cell* 2006, 22:769-781.

91. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009, 37:e67.

92. Divina P, Kvitkovicova A, Buratti E, Vorechovsky I. Ab initio prediction of mutation-induced cryptic splice-site activation and exon skipping. *Eur J Hum Genet* 2009, 17:759-765.

93. Paz I, Akerman M, Dror I, Kosti I, Mandel-Gutfreund Y. SFmap: a web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Res* 2010, 38:W281-285.

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

95. Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A* 2011, 108:11093-11098.

96. Piva F, Giulietti M, Burini AB, Principato G. SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Hum Mutat* 2012, 33:81-85.

97. Chang TH, Huang HY, Hsu JB, Weng SL, Horng JT, Huang HD. An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics* 2013, 14 Suppl 2:S4.