

# LISTENING TO SILENCE AND UNDERSTANDING NONSENSE: EXONIC MUTATIONS THAT AFFECT SPLICING

Luca Cartegni\*, Shern L. Chew<sup>‡</sup> and Adrian R. Krainer\*

Point mutations in the coding regions of genes are commonly assumed to exert their effects by altering single amino acids in the encoded proteins. However, there is increasing evidence that many human disease genes harbour exonic mutations that affect pre-mRNA splicing.

Nonsense, missense and even translationally silent mutations can inactivate genes by inducing the splicing machinery to skip the mutant exons. Similarly, coding-region single-nucleotide polymorphisms might cause phenotypic variability by influencing splicing accuracy or efficiency. As the splicing mechanisms that depend on exonic signals are elucidated, new therapeutic approaches to treating certain genetic diseases can begin to be explored.

## PSEUDO-EXON

A pre-mRNA sequence that resembles an exon, both in its size and in the presence of flanking splice-site sequences, but that is never recognized as an exon by the splicing machinery (the spliceosome).

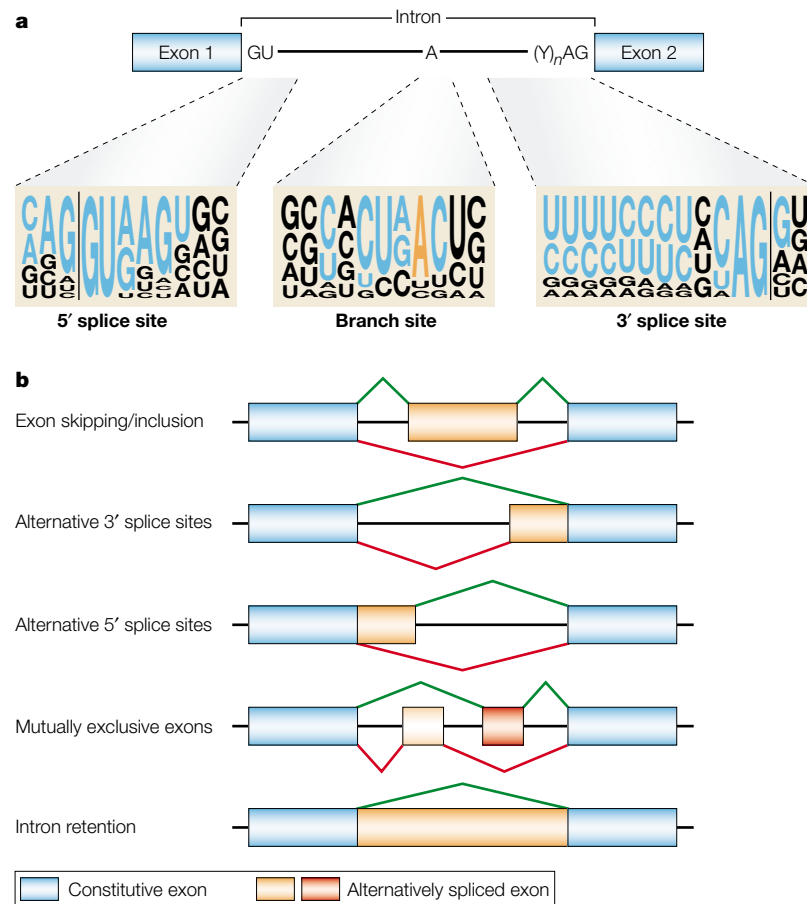
A typical mammalian gene is composed of several relatively short exons that are interrupted by much longer introns. To generate correct, mature mRNAs, the exons must be identified and joined together precisely and efficiently, in a process that requires the coordinated action of five small nuclear (sn)RNAs (U1, U2 and U4–U6) and more than 60 polypeptides<sup>1,2</sup>. The inaccurate recognition of exon–intron boundaries or the failure to remove an intron generates aberrant mRNAs that are either unstable or code for defective or deleterious protein isoforms.

Paradoxically, in higher eukaryotes, the requirement for accurate splicing is accompanied by exon–intron junctions that are defined by weakly conserved intronic *cis*-elements: the 5' splice site, 3' splice site and branch site<sup>1</sup> (FIG. 1a). These elements are necessary but are by no means sufficient to define exon–intron boundaries. Sequences that match the consensus splice-site signals as well as, or better than, natural splice sites are very common in introns, and they define a set of PSEUDO-EXONS that greatly outnumber genuine exons but that are normally not included in mature mRNAs<sup>3</sup>.

The production of several isoforms from the same transcription unit by various types of alternative splicing (FIG. 1b) is a very common event in mammalian cells, and makes the already formidable task of correctly identifying splice sites even more complex. Alternative splicing is responsible for much of the complexity of the proteome, which partly explains the unexpected finding that the human genome might consist of only 31,000–39,000 genes<sup>4,5</sup>. Because a single primary transcript can have several regions that each undergo alternative splicing, the resulting combinatorial effects of selecting different splice sites can be very pronounced, and genes that code for tens to hundreds of different isoforms are common<sup>6</sup>. In what is perhaps the most striking example of alternative splicing complexity known, the single pre-mRNA for a *Drosophila melanogaster* axon guidance receptor gene, *Down syndrome cell-adhesion molecule (Dscam)*, can be processed to generate potentially 38,016 different mature transcripts<sup>7</sup>. Even if only a subset of these isoforms is ever produced *in vivo*, this combinatorial use of alternative exons still represents an incredible source of diversity, especially given that the entire *Drosophila* genome consists of only ~14,000 genes<sup>8</sup>.

\*Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA.

<sup>‡</sup>Department of Endocrinology, St Bartholomew's Hospital, London EC1A 7BE, UK. Correspondence to A.R.K. e-mail: krainer@cshl.org  
DOI: 10.1038/nrg775



**Figure 1 | Classical splicing signals and modes of alternative splicing. a** | Conserved motifs at or near the intron ends. The nearly invariant GU and AG dinucleotides at the intron ends, the polypyrimidine tract (Y)<sub>n</sub>, preceding the 3' AG, and the A residue that serves as a branchpoint are shown in a two-exon pre-mRNA. The sequence motifs that surround these conserved nucleotides are shown below (adapted from REF 1). For each sequence motif, the size of a nucleotide at a given position is proportional to the frequency of that nucleotide at that position in an alignment of conserved sequences from 1,683 human introns<sup>1</sup>. Nucleotides that are part of the classical consensus motifs are shown in blue, except for the branch-point A, which is shown in orange. The vertical lines indicate the exon–intron boundaries. **b** | Five common modes of alternative splicing. In each case, one alternative splicing path is indicated in green, the other path in red. In the last example, the alternative pathway corresponds to no splicing. In complex pre-mRNAs, more than one of these modes of alternative splicing can apply to different regions of the transcript, and extra mRNA isoforms can be generated through the use of alternative promoters or polyadenylation sites.

**EXON DEFINITION**  
The recognition of a particular pre-mRNA segment as an exon by the spliceosome. It involves interactions between splice sites on either side of an internal exon or between a splice site and the 5' cap or the polyadenylation signal of a terminal exon.

**SNRNP**  
(Small nuclear ribonucleoprotein). A particle that is composed of a snRNA and several polypeptides.

The prevalence of alternative splicing seems to be much higher than earlier estimates had indicated. Recent analysis of reconstructed mRNAs that are derived from chromosome 22 indicates that ~60% of genes are represented by two or more transcripts<sup>4,9</sup>. Because only a subset of transcripts were sampled in this analysis, this value is still probably an underestimate, and the actual extent of alternative splicing is likely to be even greater.

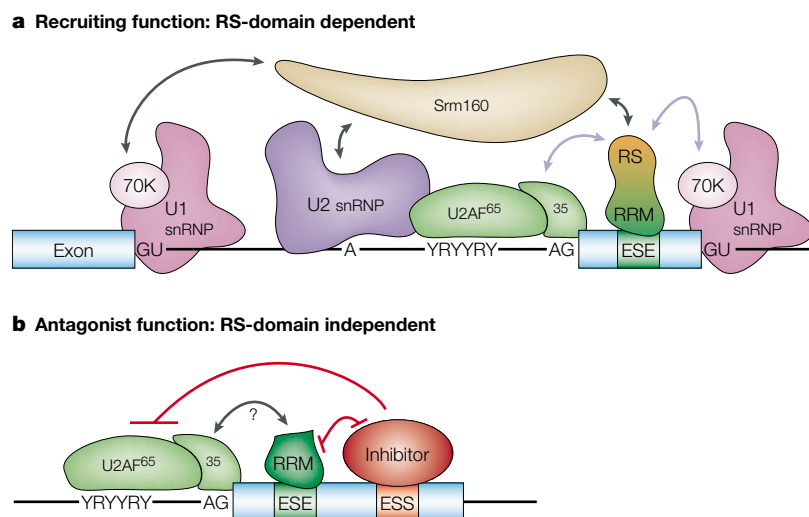
The need to regulate alternative splicing introduces an extra requirement for signals that must modulate splicing in a developmental and/or cell-type-specific fashion, and this complexity cannot be accommodated by the classical splicing signals (FIG. 1a). In this review, we discuss the other *cis*-acting elements that must be present in the coding sequences of genes to allow the

splicing machinery to distinguish between genuine and pseudo-exons and to modulate the selection of alternative splice sites (for more information about the splicing reaction, specific splicing factors and alternative splicing in general, see recent reviews in REFS 1,10–12). The mechanisms that are responsible for the recognition and function of these exonic signals are now being elucidated, and there is growing evidence that their disruption can be significant in the aetiology of several genetic diseases. In the light of these new findings, the predicted effect of nonsense, missense and silent mutations should be routinely evaluated to assess their possible consequences on pre-mRNA processing.

**Exon splicing signals: enhancers and silencers**  
Several examples of intronic and exonic *cis*-elements that are important for correct splice-site identification and that are distinct from the classical splicing signals have been described. These elements can act by stimulating (as do enhancers) or repressing (as do silencers) splicing, and they seem to be especially relevant for regulating alternative splicing. Exonic splicing enhancers (ESEs), in particular, appear to be very prevalent, and might be present in most, if not all, exons, including constitutive ones<sup>13,14</sup>.

**Enhancers.** Exonic enhancers are thought to serve as binding sites for specific serine/arginine-rich (SR) proteins<sup>10,15</sup>. These proteins are part of a growing family of structurally related and highly conserved splicing factors that are characterized by the presence of 1–2 RNA-recognition motifs (RRM) and by a distinctive carboxy-terminal domain that is highly enriched in Arg/Ser dipeptides (the RS domain)<sup>16</sup>. The RRM mediates sequence-specific binding to the RNA, and so determine substrate specificity, whereas the RS domain seems to be involved mainly in protein–protein interactions. SR proteins that are bound to ESEs can promote EXON DEFINITION<sup>17</sup> by directly recruiting the splicing machinery through their RS domain<sup>15,18,19</sup> and/or by antagonizing the action of nearby silencer elements<sup>20</sup> (FIG. 2).

These two models of splicing enhancement are not necessarily mutually exclusive, as they might reflect different requirements in the context of different exons. This idea is supported by the observation that ESE-dependent splicing requires the RS domain of an SR protein for the splicing of some substrates, which indicates that RS-domain-mediated protein–protein interactions can be essential in splicing<sup>21</sup>. The recruitment of the splicing factor U2AF (U2 snRNP auxiliary factor), either directly through RS–RS domain interactions or indirectly through SPLICING CO-ACTIVATORS, seems to be important, especially in those cases in which recognition of a weak pyrimidine tract is a rate-limiting step in the splicing reaction<sup>15,19</sup> (FIG. 2a). In other cases, the RS domain of an SR protein has been shown to be dispensable<sup>22</sup>, and the RRM is sufficient to promote *in vitro* splicing in S100 EXTRACT complementation assays. This result indicates that either an SR protein without the RS domain retains some protein–protein interaction capabilities, or that, in some situations, binding to the



**Figure 2 | Models of SR protein action in exonic-splicing-enhancer-dependent splicing.**  
**a** | RS-domain-dependent mechanism. An SR protein binds to an exonic splicing enhancer (ESE) through its RNA-recognition motifs (RRM) and contacts the splicing factor U2AF<sup>35</sup> (U2 auxiliary factor) and/or U1-70K at the adjacent splice sites through its RS domain. The U2AF splicing factor consists of two subunits (U2AF<sup>65</sup> and U2AF<sup>35</sup>), and the large subunit binds to the polypyrimidine (Y) tract, which here is interrupted by purines (R) and is therefore part of a weak 3' splice site. U2AF<sup>65</sup> also promotes binding of U2 snRNP to the branch site. U2AF<sup>35</sup> recognizes the 3' splice-site AG dinucleotide. The U1 snRNP particle binds to the upstream and downstream 5' splice sites through base pairing of the U1 snRNA; the 70K polypeptide of each U1 snRNP particle is shown. The three sets of splicing-factor–pre-mRNA interactions (U2AF–3' splice site, U1 snRNP–5' splice site and SR protein–ESE) are strengthened by the protein–protein interactions (blue arrows) that are mediated by the RS domain. For some ESE-dependent pre-mRNAs, indirect interactions (black arrows) are bridged by the splicing co-activator Smr160, which stimulates splicing of some ESE-dependent pre-mRNAs and also interacts with the U2 small nuclear ribonucleoprotein (snRNP)<sup>15</sup>. **b** | RS-domain-independent mechanism. Here, the main function of the SR protein that is bound to an ESE is to antagonize the negative effect on splicing of an inhibitory protein that is bound to a juxtaposed exonic splicing silencer (ESS). The SR protein is shown without its RS domain, although this domain is normally present and might still promote U2AF binding, or other domains might be involved in protein–protein interactions. Inhibitory interactions are shown (red), as is a putative stimulatory interaction (double-headed red arrow). These models are not mutually exclusive, and the splicing of some introns might involve a combination of these mechanisms.

ESE *per se* is sufficient to compete with negative regulatory factors, and therefore supports the antagonism model of enhancer function<sup>20,22</sup> (FIG. 2b). Finally, these two modes of SR protein/ESE-dependent function might sometimes be at work simultaneously. For example, the efficient splicing of an immunoglobulin M (*IgM*)<sub>MINIGENE</sub> pre-mRNA, which has both an enhancer and a silencer in the 3' exon<sup>20</sup>, requires an SR protein with its RS domain<sup>21</sup>. This finding indicates that a network of interactions with the splicing factor U2AF, and/or other factors, might be required to stabilize binding of the SR protein to the ESE, which would reinforce its ability to antagonize the silencer downstream.

Most early work on enhancers focused on purine-rich exonic elements, and most natural enhancers that have been reported tend to have a high purine content. Although this composition certainly reflects the actual preference of some splicing factors for purine-rich sites (TABLE 1), it probably also reflects the fact that compositionally simple sequences are easier to detect than more complex ones. High purine content by itself is not sufficient to define an ESE, as the precise sequence of the element, which might contain interspersed pyrimidines, is also important. In fact, only a few of several tested synthetic sequences of equal length and purine content can enhance the splicing of *IgM* pre-mRNA *in vitro*<sup>23</sup>. Also, several combined mutations in a cardiac troponin T minigene that do not change its purine content can disrupt enhancer activity<sup>24</sup>. Furthermore, eliminating purine-rich stretches in dispersed ESE elements in *IgM*<sup>25</sup> or avian sarcoma and leukemia virus (ASLV) pre-mRNAs does not abrogate enhancer activity *in vitro*.

The lack of a well-defined consensus sequence for purine-rich enhancers indicates that they might consist of numerous, functionally different classes, and/or that the factors involved recognize degenerate enhancer sequences. Clearly, the range of enhancer sequences that can be recognized by splicing factors is considerable (TABLE 1). Functional SELEX (systematic evolution of ligands by exponential enrichment) experiments (BOX 1) done *in vitro*<sup>27,28</sup> or *in vivo*<sup>29</sup> have also confirmed the existence of several types of ESE that include both purine-rich and non-purine-rich sequences, and have uncovered a new broad class of adenosine–cytosine-rich elements (ACEs)<sup>29</sup>. Because these studies were done in nuclear extracts or *in vivo* — that is, in the presence of a complex mixture of putative ESE-binding factors — neither well-defined consensus sequences nor the specific factors that are involved could be identified. However, by combining the SELEX approach with an S100 complementation assay, several sequence motifs that can act as enhancers and that are recognized by specific SR proteins have been determined<sup>13,30,31</sup>. Short (6–8-nucleotide), degenerate and partially overlapping ESE motifs were found using this approach (BOX 1), and this information has been used to predict computationally the occurrence of enhancers in exons<sup>13,31</sup>.

SR protein substrate specificity, as determined either by these functional SELEX studies or by identifying the specific SR proteins that bind to known

**Table 1 | RNA motifs recognized by human SR proteins**

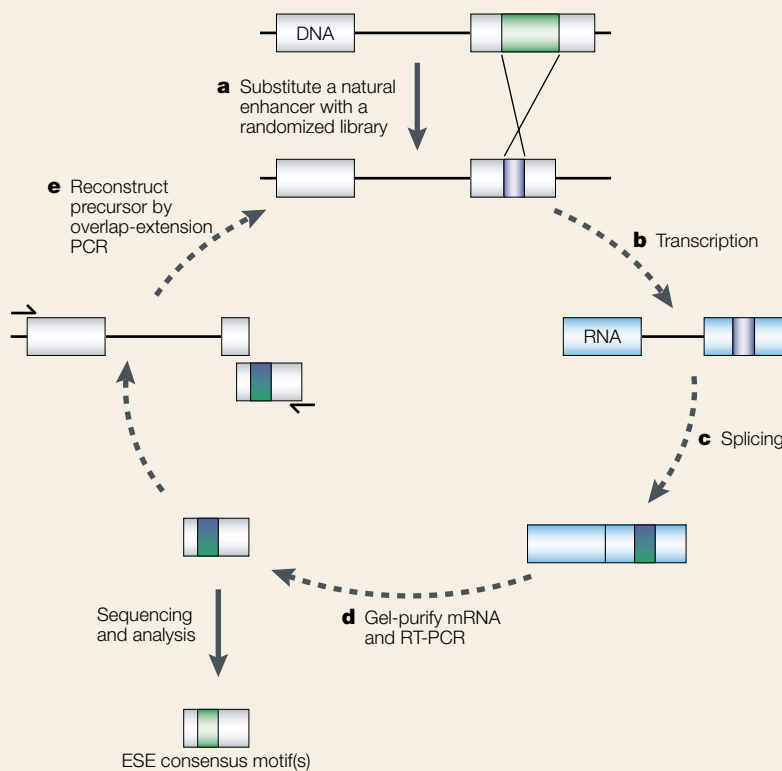
Protein	High-affinity binding site	Ref.	Functional ESE	Ref.
SRp20	WCWWC	112	GCUCCUCUCC	113
	CUCKUCY	14	CCUCGUCC	14
SC35	AGSAGAGUA	32	GRYYMCYR*	31
	GUUCGAGUA	32	UGCYGY	14
	UGUUCSAGWU	112		
	GWUWCCUGCUA	112		
	GGGUAUGCUG	112		
	GAGCAGUAGKS	112		
	AGGAGAU	112		
9G8	(GAC) <sub>n</sub>	112		
	ACGAGAGAY	112		
	WGGACRA	14		
SF2/ASF	RGAAGAAC	32	CRSMSGW*	13
	AGGACRRAGC	32		
SRp40	UGGGAGCRGUYRGUCUGY	114	YRCRKM*	13
SRp55			YYWCWSG*	13
TRA2β	(GAA) <sub>n</sub>	115		

\*These consensus sequences correspond to refined nucleotide-frequency matrices<sup>79</sup> derived from revised versions of the original data; representations of the matrices are shown in BOX 1. Nucleotide symbols used: M, A/C; R, A/G; W, A/U; Y, C/U; S, C/G; K, G/U. ESE, exonic splicing enhancer.

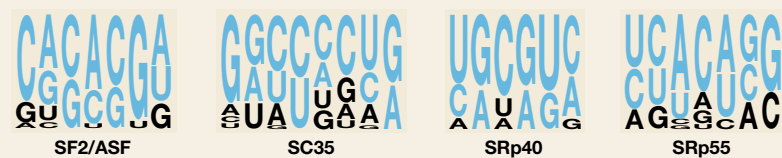
Box 1 | **Functional SELEX method**

To identify exonic splicing enhancer (ESE) motifs by functional *in vivo* or *in vitro* SELEX (systematic evolution of ligands by exponential enrichment<sup>105</sup>), a minigene is used that harbours ESE sequences that are required for the efficient splicing of its pre-mRNA. As shown in the accompanying figure, the natural enhancer (green box) is replaced by random sequences (blue) from an oligonucleotide library (a). The resulting pool of minigenes is then transfected into cultured cells, or is transcribed *in vitro*, to generate a pool of pre-mRNAs (b). Following *in vivo* or *in vitro* splicing (c), the pool of spliced mRNAs is gel purified and amplified by reverse-transcription (RT)-PCR (d). This pool of enhancer-enriched sequences is then used to reconstruct new minigene templates by OVERLAP-EXTENSION PCR<sup>106</sup> (e), to use in a new enrichment cycle. The iteration of this entire procedure yields a limited number of ‘winners’ — sequences that have good splicing enhancer activity<sup>27,107</sup>.

To identify ESEs that are recognized by individual SR proteins, the splicing step was carried out in S100 extract complemented with one of four different SR proteins (SF2/ASF, SC35, SRp40 and SRp55)<sup>13,30,31</sup>. Transcripts were obtained from an immunoglobulin- $\mu$  (*IGHM*)-derived minigene, in which the natural enhancer was substituted with a pool of 20-nucleotide random sequences. After a few cycles of enrichment, spliced products were sequenced and aligned to derive a consensus motif. The frequencies of the individual nucleotides at each position were then used to calculate a score matrix, which can be used to predict the location of SR-protein-specific putative ESEs in exonic sequences<sup>79,88</sup> (TABLE 1).



The consensus motifs obtained with these four SR proteins are shown below; the height of each letter reflects the frequency of each nucleotide at a given position, after adjusting for background nucleotide composition<sup>79</sup>. At each position, the nucleotides are shown from top to bottom in order of decreasing frequency; blue letters indicate above-background frequencies.

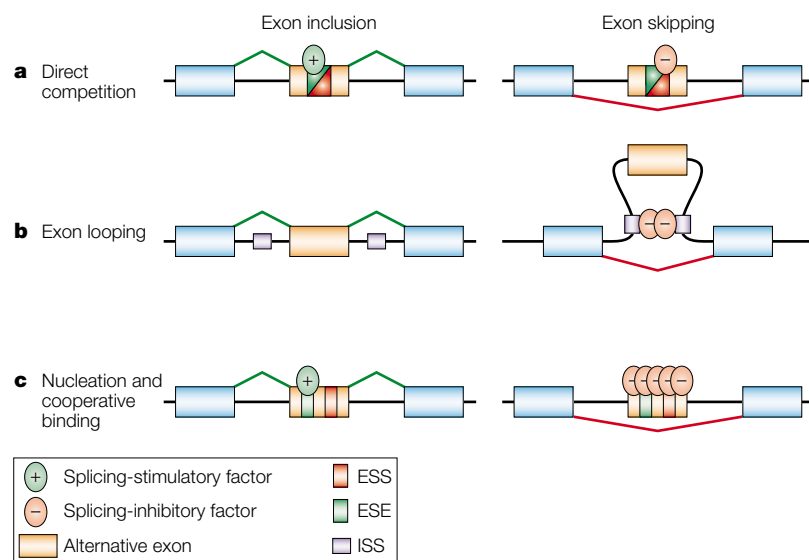


natural enhancers, often does not correspond to the optimal (highest-affinity) binding sites for these proteins (TABLE 1). This observation might reflect the fact that binding data are typically obtained in highly purified *in vitro* systems that do not fully reflect splicing conditions. For example, sequences with optimal binding sites for an SR protein might overlap with sequences that act as binding sites for other factors that positively or negatively influence splicing. Moreover, a higher binding affinity does not necessarily imply better functionality, which could explain why some of the binding sequences that come out as ‘winners’ in the conventional SELEX assay (based on selection for binding) do not work when tested as splicing enhancers *in vitro*<sup>32</sup>.

When ESE-prediction programs are used to analyse exonic sequences, putative ESEs are often found to be clustered, and the motifs tend to be enriched in regions with known natural enhancers<sup>13,31</sup>. Many of the longer, more loosely defined, natural enhancer sequences might therefore consist of clusters of several overlapping motifs, which are recognized by different proteins. This would explain the redundancy in the activity of SR proteins, the apparent lack of clearly identifiable consensus sequences and the requirement for several mutations to inactivate these dispersed elements.

**Silencers.** Exonic splicing silencers (ESSs) are less well characterized than ESEs, at present, but they are probably just as prevalent. About one-third of randomly selected short, human DNA fragments showed splicing inhibitory activity *in vivo* when inserted into the middle exon of a three-exon minigene<sup>33</sup>. This finding indicates that some exons might be maintained in a generalized ‘silenced’ state, which can be neutralized only by a combination of strong flanking splicing signals (that is, good matches to the consensus signals) and/or efficient enhancer elements. Most described silencers are intronic elements, but several ESS elements have also been reported<sup>20,34–39</sup>. Their mechanisms of action are still not fully understood. Silencers seem to work by interacting with negative regulators, which often belong to the heterogeneous nuclear ribonucleoprotein (hnRNP) family<sup>12,40,41</sup> — a class of diverse RNA-binding proteins that associate with nascent pre-mRNAs<sup>42,43</sup>. Similar to SR proteins, hnRNP proteins have a modular structure, which consists of one or more RNA-binding domains associated with an auxiliary domain that is often involved in protein–protein interactions. In particular, the hnRNP I protein (better known as polypyrimidine-tract-binding protein (PTB)) and proteins of the hnRNP A/B and hnRNP H families are the best-characterized mediators of silencing<sup>40,41,44</sup>.

Direct competition by inhibitory factors for overlapping enhancer-binding sites that are recognized by stimulatory factors could lead to silencing if the enhancer element is a crucial one (FIG. 3a). Alternatively, binding of PTB or hnRNP A1 to several sites around or within a silenced exon, followed by dimerization of the bound PTB proteins, might cause relevant portions of the pre-mRNA to loop out, which makes them unavailable for



**Figure 3 | Models of splicing silencing.** **a** | Silencing by direct competition. A splicing-stimulatory factor, such as an SR protein, binds to an exonic splicing enhancer (ESE) (+; left). A splicing-inhibitory factor, such as an heterogeneous nuclear ribonucleoprotein (hnRNP), binds to an exonic splicing silencer (ESS) (-; right). Because the ESE and the ESS overlap, the binding of the positive and negative factors is mutually exclusive. If the positive factor has a higher binding affinity or higher concentration than the negative one, the alternative exon is included; if vice versa, then it is excluded. **b** | Silencing by exon looping. A splicing inhibitory factor, such as an hnRNP protein, binds to duplicate intronic splicing silencer (ISS) elements present in the introns that flank an alternative exon. Dimerization of the bound protein brings the ISS elements into juxtaposition, causing the alternative exon to loop out and to be skipped by the splicing machinery. **c** | Silencing by nucleation and cooperative binding. The alternative exon harbours ESE and ESS elements at some distance from each other. In the absence of an inhibitory factor, the ESS element remains unoccupied, and an SR protein can bind to the ESE and stimulate exon inclusion (+; left). An inhibitory factor (-; right) initially binds to a high-affinity binding site in the ESS, and nucleates cooperative binding of additional inhibitory molecules, which polymerize along the exon and displace the ESE-bound SR protein or prevent its initial binding, resulting in exon skipping.

**SPLICING CO-ACTIVATOR**

A protein that mediates splicing enhancement without binding directly to the pre-mRNA.

**S100 EXTRACT**

A post-nuclear, post-ribosomal cellular fraction that can support *in vitro* splicing only when complemented with one or more SR proteins.

**MINIGENE**

A simplified laboratory version of a natural gene that lacks one or more of the gene's exons and introns, or portions of them.

**RT-PCR**

A type of PCR in which RNA is converted into single-stranded DNA, which is then amplified.

**OVERLAP-EXTENSION PCR**

A method that involves consecutive PCR reactions with overlapping primers, which is useful to recombine different DNA sequences *in vitro*.

splicing (see FIG. 3b and REFS 41,45). Another proposed silencing mechanism involves hnRNP A1 binding initially to a high-affinity site on the exon and then nucleating the cooperative assembly of inhibitory hnRNP complexes that coat the pre-mRNA. The polymerization of hnRNP A1 interferes directly with the initial steps of spliceosome assembly or antagonizes the action of nearby enhancers<sup>22</sup> (FIG. 3c).

The decision of whether to include an exon reflects the intrinsic strength of the flanking splice sites and the combinatorial effects of positive and negative elements (reviewed in REFS 40,46). The complexity and subtlety of this modulation is most apparent when alternative splicing is involved<sup>6,12</sup>. However, even seemingly simple decisions, such as the inclusion of a CONSTITUTIVE EXON, probably require elaborate crosstalk between several *cis*-acting elements and *trans*-acting factors. Mutations that disrupt only one of a few crucial elements in a given exon can therefore markedly affect the splicing pathway and the extent of exon inclusion.

**Splicing signals and point mutations**

Splicing signals are a frequent target of mutations in genetic diseases and cancer. In a widely cited survey, Krawczak and colleagues estimated that at least 15% of

point mutations that result in a human genetic disease cause RNA splicing defects<sup>47</sup>, a figure that is supported by the annotation of ~16,000 point mutations in the current **Human Gene Mutation Database**. Most splicing mutations that are considered in these surveys directly affect the standard consensus splicing signals (see FIG. 1a), and typically lead to skipping of the neighbouring exon. Less frequently, the mutations create an ectopic splice site or activate a CRYPTIC SPLICE SITE, thereby changing the overall splicing pattern of the mutant transcript. At present, most databases contain annotation data that are primarily or exclusively derived from genomic DNA analysis, and the effect of a mutation on the mRNA or on the encoded protein is usually predicted from the primary sequence, rather than by experimentally determining mRNA expression and splicing patterns. Therefore, point mutations that occur in introns and that affect the classical consensus splice-site signals are considered to be splicing mutations, whereas point mutations in the coding regions that do not create ectopic splice-site consensus sequences are usually scored as missense, nonsense or silent mutations.

Nonsense mutations are commonly assumed to produce truncated protein isoforms, whereas missense mutations are presumed to identify amino acids that are important for the structure or function of a protein. Translationally silent mutations are normally classified as allelic polymorphisms and are considered to be neutral. These assumptions might be correct in some cases, but when they are not supported by characterization at the mRNA level, they could be misleading, because mutations that affect sequences that are important for splicing modulation are likely to have a profound effect on the translated product. For example, if a missense mutation abrogates an ESE and causes exon skipping, the mutant protein, instead of having just a single amino-acid difference from the wild type, will carry a large internal deletion or, if the open reading frame (ORF) is not maintained, an entirely different and probably shorter carboxy-terminal domain. In either case, no inferences should be drawn about the functional importance of the amino acid that is encoded by the wild-type codon. In addition, premature termination codons (PTCs), whether they arise from nonsense or frameshift mutations or as a result of mutation-induced exon skipping, which in turn causes a frameshift, usually trigger nonsense-mediated mRNA decay (NMD, see BOX 2). This mRNA surveillance mechanism leads to a reduction in the abundance of PTC-harboring mRNAs and of the corresponding truncated protein.

Indeed, there is growing evidence that misclassification of mutations might commonly occur, and that the general extent of splicing mutations has been underestimated. Many reports have correlated specific point mutations in coding regions with the skipping of the exon that harbours the mutation. In two recent studies, several germ-line mutations in the neurofibromatosis type 1 (*NF1*) gene and the ataxia telangiectasia mutated (*ATM*) gene were systematically analysed at the DNA and RNA level in **neurofibromatosis type 1**

**CONSTITUTIVE EXON**  
An exon that is always included in the mature mRNA, even in different mRNA isoforms.

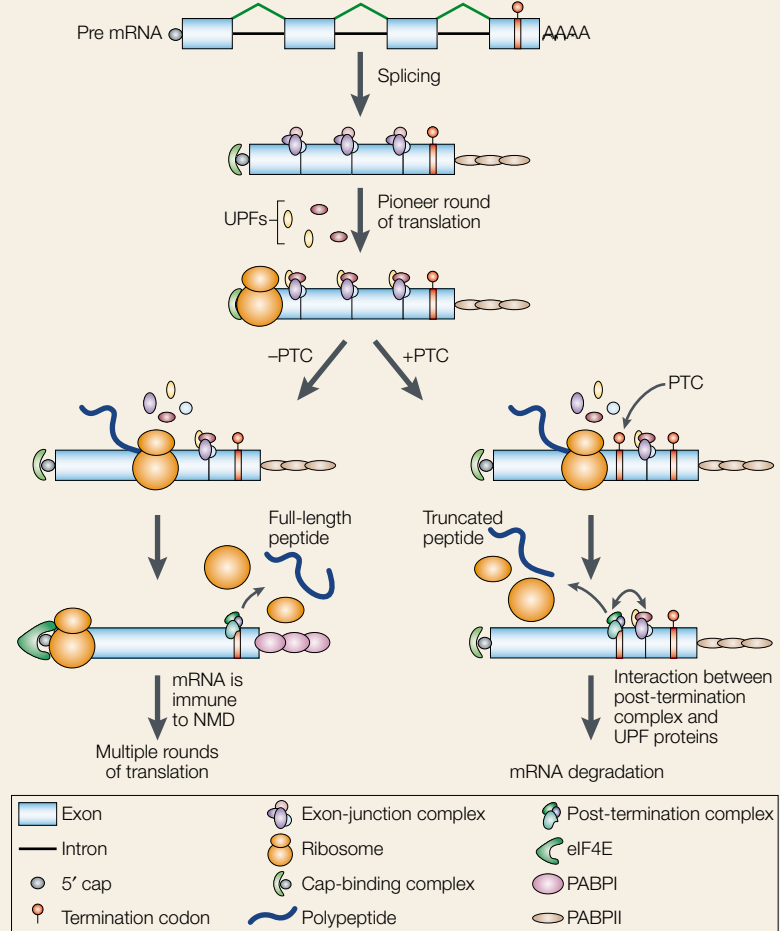
and **ataxia telangiectasia** patients that harbour these mutations<sup>48,49</sup>. In ~50% of these patients (26 out of 52 and 30 out of 62, respectively), the disease was found to be due to mutations that result in aberrant splicing, either by causing exon skipping or by activating new

cryptic splice sites. Of these mutations, 13 and 11% (7 out of 52 and 7 out of 62, respectively) would have been erroneously classified as frameshift, missense or nonsense mutations if the analysis had been limited to genomic sequence.

**Box 2 | Nonsense-mediated mRNA decay**

Nonsense-mediated mRNA decay (NMD) is an mRNA surveillance mechanism that has been described from yeast to humans and ensures mRNA quality by selectively targeting mRNAs that harbour premature termination codons (PTCs) for rapid degradation<sup>50,58,108</sup>. PTCs that are introduced as a consequence of DNA rearrangements, frameshifts or nonsense mutations, or are caused by errors during transcription or splicing, can lead to non-functional or deleterious proteins.

PTCs in higher eukaryotes are only recognized as such when they occur upstream of a 'boundary' on the spliced mRNA that is situated ~55 nucleotides before the last exon-exon junction<sup>108</sup>. As summarized in the accompanying figure, the prevalent view of the NMD mechanism is that the splicing process leaves a 'mark' ~20 nucleotides upstream of each exon-exon boundary, in the form of an exon-junction complex (EJC), which in turn provides an anchor for up-frameshift suppressor proteins (UPFs)<sup>108,109</sup>. During the first ('pioneer') round of translation of a normal mRNA, the stop codon is located downstream of the last mark, and all EJCs are displaced by elongating ribosomes<sup>110</sup>. During subsequent rounds of translation, the cap-binding complex is replaced by eIF4E (eukaryotic initiation factor 4E) and PABPII (poly(A)-binding protein II) is replaced by PABPI, new ribosomes no longer encounter EJCs, and the mRNA is immune to NMD. However, when a PTC is present, ribosomes stop and fail to displace the downstream EJCs from the transcript. Interactions between the marking factors and components of the post-termination complex trigger mRNA decay.



**Nonsense-associated altered splicing**

Several reported point mutations that are linked to exon skipping introduce a PTC. This phenomenon is known as nonsense-associated altered splicing (NAS, see BOX 3)<sup>50,51</sup>, and the role of reading-frame recognition in this process is particularly interesting. Does the presence of a PTC cause the aberrant splicing or is exon skipping a consequence of the fortuitous disruption of cis-elements that are important for splicing?

**Nuclear scanning.** At least in some cases, recognition of the translational reading frame seems to be involved in NAS. A nonsense mutation in exon 51 of the fibrillin 1 (*FBNI*) gene is associated with exon skipping and causes **Marfan syndrome**<sup>52</sup>. All three nonsense codons, but not a synonymous codon, induce exon skipping, and the splicing pattern is largely restored when the PTCs are placed out of frame by introducing a frameshift mutation upstream<sup>53</sup>. Similarly, in transcripts of a mouse parvovirus, PTCs inhibit splicing in a frame-dependent but not translation-dependent manner, as mutations at the translational start site do not always reverse NAS<sup>54,55</sup>. In addition, transcripts from the T-cell receptor  $\beta$  (*TRB*) or immunoglobulin  $\mu$  heavy-chain (*IGHM*) genes accumulate at the site of transcription only when they harbour PTCs, which indicates that the mRNA reading frame can exert intranuclear effects<sup>56</sup>. A proposed explanation for these observations is the existence of a nuclear mRNA surveillance system that is distinct from NMD and directly affects the splicing process. This putative process would somehow verify the integrity of an ORF and, when necessary, direct the splicing machinery to skip the offending exon (see BOX 3a). This nuclear-scanning model posits that translation-like machinery might exist in the nucleus, perhaps composed of functional nuclear ribosomes, which possibly explains the observation that many mRNAs that undergo NMD still seem to be associated with the nucleus<sup>50,57-60</sup>. Although the existence of a nuclear ORF 'scanner' remains controversial, the presence of translation initiation factors, such as eIF4E and eIF4G, and charged tRNAs in the nucleus, combined with evidence indicating that translation is coupled to transcription in both mammalian and slime-mould cell nuclei, supports this unconventional view<sup>61-64</sup>.

Arguably, a mechanism that promotes the skipping of PTC-containing exons might confer a selective advantage by allowing a protein with residual function to be made, instead of a completely inactive one. This situation is seen in the case of **Duchenne muscular dystrophy (DMD) and its milder variant, Becker muscular dystrophy (BMD)**, which are caused by mutations in the dystrophin gene (*DMD*)<sup>65</sup>. Most

CRYPTIC SPLICE SITE

A sequence that matches the 5' or 3' splice-site consensus, but is only used as a splice site in the context of a mutation elsewhere in the gene, such as one that inactivates or weakens an authentic splice site.

described *DMD* mutations eliminate one or more internal exons or alter proper splicing. In general, the less severe BMD phenotypes correlate with the maintenance of the reading frame<sup>66</sup>. Similarly, the skipping of PTC-containing internal exons, which preserves the *DMD* ORF, sometimes leads to BMD<sup>67,68</sup>. However, the production of a partially functional protein that results from the skipping of a constitutive exon is not a common event. More frequently, altered splicing

results in frameshifts, which in turn trigger NMD. If the reading frame were maintained, or if the PTC that was introduced by a frameshift were recognized as a normal termination codon because of its position, the surveillance function of NMD would be pre-empted, so increasing the chance of a dominant-negative phenotype.

Several other observations argue that NAS probably does not have a general role in mRNA surveillance. First,

Box 3 | Possible mechanisms of nonsense-associated altered splicing

In nonsense-associated altered splicing (NAS), exons that harbour premature termination codons (PTCs) are excluded from the mature transcript<sup>50</sup>. Four possible mechanisms for NAS are discussed below and in the accompanying figure.

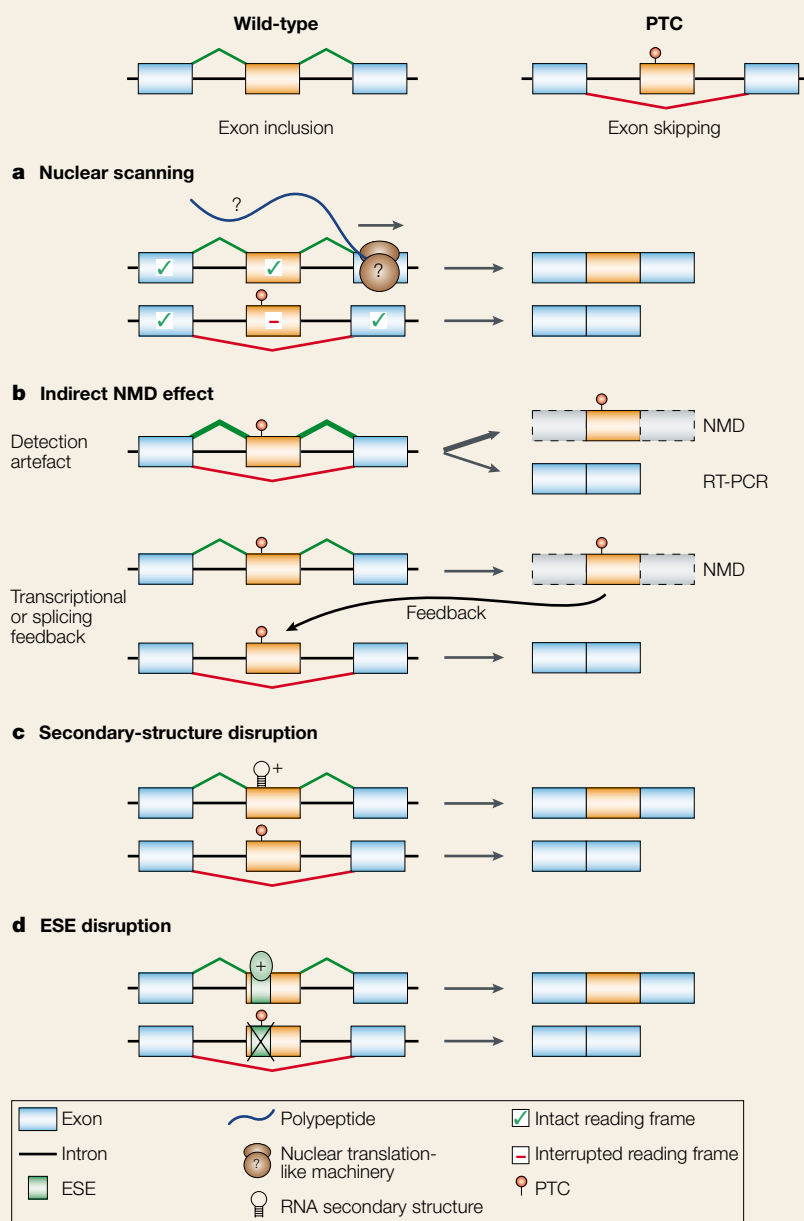
The top part of the figure shows a wild-type, constitutively spliced pre-mRNA, and a mutant pre-mRNA in which the middle exon harbours a PTC and is absent from the mature mRNA.

In the nuclear scanning model (a), a translation-like machinery in the nucleus scans the reading frame and surveys its integrity before splicing<sup>51,60</sup>. If the exons have intact reading frames, splicing ensues and all the exons are included. If an interrupted reading frame is sensed, the splicing machinery somehow skips the exon. This presumptive machinery might or might not synthesize a polypeptide, and would have to deal somehow with the usual interruption of the reading frame by introns.

In the indirect nonsense-mediated mRNA decay (NMD) model (b), NAS is an indirect consequence of NMD. In some cases, a small amount of exon-skipped mRNA is always generated, whether the pre-mRNA has a PTC or not. When there is a PTC, the full-length mRNA is degraded by conventional NMD; as a result, the apparent abundance of the exon-skipped isoform might be magnified, especially when mRNA is detected by reverse-transcription (RT)-PCR. This artefact accounts for some, but not all, instances of reported NAS<sup>73</sup>. The PTC in a spliced mRNA might also trigger a feedback signal, which results in the accumulation of exon-skipped mRNA. If the protein encoded by the full-length mRNA is involved in a negative-feedback loop to downregulate the transcription of its own gene, then a non-functional protein encoded by the exon-skipped mRNA might lead to more transcripts being generated; the PTC-containing ones would be degraded by NMD, and the exon-skipped ones would accumulate. A similar outcome is expected if the full-length protein promotes the inclusion of that exon in its own pre-mRNA. Alternatively, triggering of NMD might alter the local concentration of factors that normally bind to the included exon, which somehow causes the PTC-harboured exon to be skipped.

The secondary-structure disruption model applies to pre-mRNAs in which local RNA secondary structure is required to promote exon inclusion (c). If a PTC disrupts the hairpin, exon inclusion is no longer favoured<sup>73,74</sup>. This mechanism does not require that the mutation be a nonsense mutation.

In the exonic splicing enhancer (ESE)-disruption model (d), the PTC fortuitously disrupts the recognition motif for an RNA-binding protein that enhances splicing, such as an SR protein, and exon inclusion is no longer favoured<sup>67,79</sup>. A similar disruption might be caused by certain missense or translationally silent mutations.



it is not clear how common NAS is, as most PTCs do not induce exon skipping, but instead elicit NMD<sup>51</sup>. In addition, in many cases in which splicing is affected, the exon that is involved seems to be weakly defined. For example, a silent mutation in exon 51 of *FBNI*, at a different position from the previously described PTC, also causes exon skipping and Marfan syndrome<sup>69</sup>, and the previously mentioned PTCs in the mouse parvovirus transcript occur at a single position close to an ESE<sup>55</sup>, which implies that the effects are exon specific, rather than being dependent on nonsense mutations. Furthermore, both NAS-triggering and non-triggering natural PTCs have been shown to coexist within the same exons of human *NF1* and hypoxanthine phosphoribosyltransferase 1 (*HPRT1*) genes<sup>70,71</sup>, which indicates that the presence of a PTC *per se* is not sufficient to induce exon skipping.

**An indirect nonsense-mediated mRNA decay effect?** The reported upregulation of many alternatively spliced products in which PTC-containing exons have been skipped could turn out to be an indirect effect of NMD, rather than being an independent consequence of the PTC (see BOX 3b). In several cases, the correlation between PTCs and exon skipping has been shown to be an artefact of the reverse-transcription (RT)-PCR detection method<sup>72,73</sup>. Because of the presence of the PTC, the mRNA that carries the mutated exon is targeted by NMD and degraded. Consequently, background levels of exon-skipped mRNA species, which would normally go undetected because of competition from abundant, wild-type mRNA, can now be more efficiently amplified by RT-PCR.

Alternatively, the increased levels of exon-skipped mRNA might arise indirectly as a result of NMD combined with a transcriptional feedback regulation (see BOX 3b). If a gene is normally autoregulated by a feedback loop to generate a steady-state level of transcripts, the presence of a PTC would result in increased transcription by causing mRNA decay and a reduction in the encoded protein. As a consequence, any alternatively spliced transcript that no longer harbours the PTC would become upregulated. The feedback target might be the splicing event itself, which could theoretically be activated by a specific limiting factor(s) acting *in trans*.

An NMD-dependent *trans*-acting mechanism could also explain why, paradoxically, a PTC can inhibit the removal of an upstream intron, even though the sensing of whether the PTC is in frame should require that the intron first be spliced out, which indicates that the initial PTC recognition might be a post-splicing event<sup>56</sup>.

**The cis connection.** A very different explanation for NAS is that PTCs are fortuitously involved in altered splicing, and the observed effects are caused by the disruption of *cis*-elements that are important for correct splicing<sup>73</sup> (see BOX 3c,d). RNA folding can be important in promoting splicing, as illustrated by the fibronectin (*FNI*) gene transcript, which requires a

correct secondary structure for the splicing of an alternative exon<sup>74</sup>. Alteration of secondary structure in the Fanconi anaemia complementation group C (*FANCC*) gene transcript has been proposed to explain skipping of exon 6 — which is associated with a nonsense mutation in exon 6 itself<sup>75</sup> — and with nonsense and missense mutations in exon 5 (REF. 76). A strong correlation between exon skipping and the disruption of RNA secondary structure by nonsense and missense mutations has also been shown for *HPRT1* exons 2, 4 and 8 and *FBNI* exon 51 (REF. 71).

Alternatively, the *cis*-elements that are affected by nonsense mutations can be splicing enhancers. A putative role for ESE disruption in PTC-associated exon skipping has been previously proposed<sup>73,77</sup> and, in some cases, direct evidence for this mechanism has been provided. The E1211X nonsense mutation in exon 27 of *DMD* introduces a uracil into a purine-rich element within the exon<sup>67</sup>. The wild-type, but not the mutant, element is able to promote *in vitro* splicing in a heterologous context, which indicates that the cause of exon skipping might be the ESE disruption. Similarly, the E1694X nonsense mutation in exon 18 of the breast cancer susceptibility gene (*BRCA1*) results in exon skipping<sup>78</sup>. An analysis of exon 18 sequence, using the ESE-score matrices that are described in BOX 1, has indicated that the E1694X nonsense mutation abrogates a putative ESE that is recognized by the SR protein SF2/ASF. Furthermore, mutational analysis has shown that the inclusion of this exon correlates with the maintenance of a high-score ESE motif, regardless of the nature of the mutations (nonsense or missense)<sup>79</sup>. Significantly, when the ESE motif analysis was extended to 50 other mutations that cause exon skipping in various human disease genes, more than half of the single-base substitutions reduced or eliminated at least one ESE high-score motif<sup>79</sup>. Together, these findings indicate that interfering with the function of exonic *cis*-elements that are involved in splicing might be a common mechanism for inappropriate exon skipping.

A definitive answer about the mechanisms that are responsible for NAS will not be available until more systematic studies of several nonsense, missense and silent mutations are carried out. Because exon skipping can be advantageous in some cases, it is possible that specific mechanisms have evolved to exploit NMD when the occurrence of nonsense codons is a likely event. For example, in the *TRB*<sup>80</sup> and *IGHM*<sup>81</sup> genes, programmed DNA rearrangements that juxtapose variable and constant regions introduce PTCs in approximately two-thirds of the recombination events. However, most of the available evidence indicates that the correlation between the alteration of splicing patterns and the presence of PTCs is largely coincidental. Most of these mutations are likely to affect either secondary structures that are essential for correct splicing or the integrity of an ESE (BOX 3c,d). Clearly, both of these mechanisms are independent of the reading frame and so should apply equally to nonsense, missense and silent mutations.



Table 2 | Missense and silent mutations associated with altered splicing

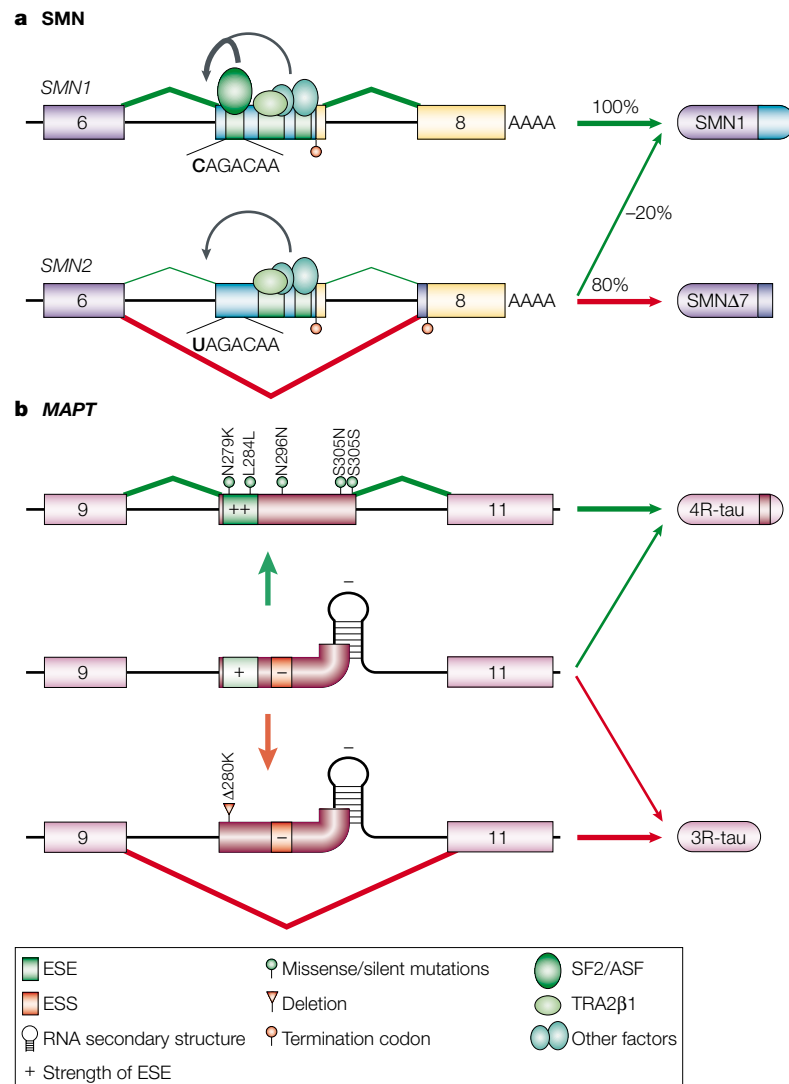
Gene	Mutation	Exon	Ref.	Gene	Mutation	Exon	Ref.
<b>Missense mutations</b>				<b>Silent mutations</b>			
<i>ADA</i>	A215T	7	116	<i>PDHA1</i>	A175T	6	129
<i>ATM</i>	E2032K	44	49	<i>PMM2</i>	E139K	5	130
<i>ATP7A</i>	G1302R	4	117	<i>RHAG</i>	G380V	9	131
<i>BRCA1</i>	E1694K	18	78	<b>Silent mutations</b>			
<i>CFTR</i>	G58E	9	118	<i>APC</i>	R623R	14	132
	D565G	12	119	<i>AR</i>	S888S	8	133
<i>F8</i>	R1997W	19	120	<i>ATM</i>	S706S <sup>‡</sup>	16	49
<i>FAH</i>	Q279R*	9	121		S1135S <sup>‡</sup>	26	49
<i>FBN2</i>	D1114H <sup>‡</sup>	25	122	<i>CYP27A1</i>	G112G	2	134
<i>FECH</i>	A155P <sup>‡</sup>	4	123	<i>FAH</i>	N232N	8	135
<i>HEXB</i>	P404L	11	124	<i>FBN1</i>	I2118I	51	69
<i>HMBS</i>	E29L <sup>‡</sup>	3	125	<i>HEXA</i>	L187L <sup>‡</sup>	5	136
<i>HPRT1</i>	G40V	2	73	<i>HMBS</i>	R28R	3	137
	R48H	3	73	<i>HPRT1</i>	F199F	8	73
	A161E	6	73	<i>ITGB3</i>	T420T	9	138
	G180E	8	73	<i>LIPA</i>	Q277Q <sup>‡</sup>	8	139
	G180V	8	73	<i>MAPT</i>	L284L <sup>§</sup>	10	89
	E182K	8	73		N296N <sup>§</sup>	10	89
	P184L	8	73		S305S <sup>‡§</sup>	10	89
	D194Y	8	73	<i>MLH1</i>	S577S <sup>‡</sup>	16	140
	E197K	8	73	<i>NF1</i>	K354K	8	141
	E197V	8	73	<i>PAH</i>	V399V	11	142
	D201V	8	73	<i>PDHA1</i>	G185G <sup>‡</sup>	6	143
<i>IL2RG</i>	R285Q <sup>‡</sup>	6	126	<i>PKLR</i>	A423A	9	144
<i>IVD</i>	R21C	2	127	<i>PTPRC</i>	P48P	4	145
	R21P	2	127	<i>PTS</i>	E81E <sup>‡</sup>	4	146
	D20N	2	127	<i>RET</i>	I647I	11	147
<i>MAPT</i>	N279K <sup>§</sup>	10	89	<i>SMN1</i>	F280F	7	84
	S305N <sup>*§</sup>	10	89	<i>TNFRSF5</i>	T136T	5	148
<i>MLH1</i>	R659P	17	128	<i>UROD</i>	E314E	9	149
	R659L	17	128				

\*Mutations in the penultimate nucleotide of the exon; †mutations in the last nucleotide of the exon; §mutations that increase exon inclusion. *ADA*, adenosine deaminase; *APC*, adenomatosis polyposis coli; *AR*, androgen receptor; *ATM*, ataxia telangiectasia mutated; *ATP7A*, ATPase, Cu<sup>2+</sup> transporting,  $\alpha$ -polypeptide; *BRCA1*, breast cancer 1, early onset; *CFTR*, cystic fibrosis transmembrane conductance regulator; *CYP27A1*, sterol-27-hydroxylase; *F8*, coagulation factor VIII, procoagulant component; *FAH*, fumarylacetoacetate hydrolase; *FBN1*, fibrillin 1; *FBN2*, fibrillin 2; *FECH*, ferrochelatase; *HEXA*, hexosaminidase A,  $\alpha$ -polypeptide; *HEXB*, hexosaminidase B,  $\beta$ -polypeptide; *HMBS*, hydroxymethylbilan synthase; *HPRT1*, hypoxanthine phosphoribosyltransferase 1; *IL2RG*, interleukin 2 receptor- $\gamma$ ; *ITGB3*, integrin- $\beta$ 3; *IVD*, isovaleryl coenzyme A dehydrogenase; *LIPA*, lipase A; *MAPT*, microtubule-associated protein tau; *MLH1*, mutL homologue 1; *PAH*, phenylalanine hydroxylase; *PDHA1*, pyruvate dehydrogenase; *PKLR*, pyruvate kinase, liver and red blood cells; *PMM2*, phosphomannomutase 2; *PTPRC*, protein-tyrosine phosphatase receptor type C; *PTS*, 6-pyruvoyltetrahydropterin synthase; *RET*, rearranged during transfection protooncogene; *RHAG*, Rhesus blood group-associated glycoprotein; *SMN1*, survival of motor neuron 1; *TNFRSF5*, tumour-necrosis factor receptor superfamily, member 5 (*CD40*); *UROD*, uroporphyrinogen decarboxylase.

**Missense or silent mutations and exon skipping**

Exon skipping that is associated with point mutations other than those of the nonsense type has been frequently observed (TABLE 2). The fact that even mutations that are predicted to be translationally silent cause exon skipping is particularly significant, because these mutations must act at the RNA level (except for possible subtle effects on translation efficiency owing to codon preferences). These mutations probably alter *cis*-elements that are important for correct splicing. Furthermore, it is very likely that such mutations are generally under-reported, because they

might be incorrectly assumed to be neutral polymorphisms that do not merit further characterization. A correct assessment of the relative frequency of different types of mutation that affect splicing will require more systematic studies. However, the common occurrence of both missense and silent mutations that cause exon skipping, in addition to the numerous examples of NAS<sup>73</sup>, indicates the possible existence of a single mechanism as the cause of the splicing defect, regardless of the type of exonic mutation. A shared mechanism would be inconsistent with NAS models that propose that the reading frame



**Figure 4 | Examples of splicing alterations caused by exonic mutations.** **a** | Splicing of pre-mRNAs from the human spinal muscular atrophy (SMA) survival of motor neuron 1 (*SMN1*) and *SMN2* genes. In *SMN1*, exon 7 is a constitutive exon, which harbours the normal termination codon. In *SMN2*, exon 7 is alternatively spliced: 80% of the spliced mRNAs skip exon 7, such that translation terminates at exon 8. The resulting SMNΔ7 protein has a different carboxy terminus, which renders it unstable. Twenty per cent of the *SMN2*-spliced mRNAs include exon 7, which gives rise to full-length, functional SMN protein. A single, translationally silent C → T transition at position +6 of exon 7 is responsible for these splicing differences. This nucleotide change inactivates an SF2/ASF-dependent exonic splicing enhancer (ESE), which begins at position +6. Other ESE elements<sup>86</sup> in exon 7 are probably responsible for the basal levels of exon 7 inclusion, and are recognized by different factors, including the SR-like protein TRA2β1 (REF. 111). So, SMA patients can produce functional SMN protein but at levels that are insufficient for normal function, which results in the progressive degeneration of spinal-cord motor neurons. **b** | Splicing of pre-mRNA from the human *MAPT* gene, which encodes tau and is involved in fronto-temporal dementia and parkinsonism associated with chromosome 17 (FTDP17). In the normal pre-mRNA, exon 10 harbours both ESE and exonic splicing silencer (ESS) elements (middle). A long tau protein isoform (4R-tau) is made when the alternative exon 10 (dark purple) is included (top), and a short isoform (3R-tau) is made when this exon is skipped (bottom). Disrupting the tightly regulated 4R-tau:3R-tau ratio in either direction can give rise to FTDP17. A secondary structure element that involves exon 10 and intron 10 sequences sequesters the 5' splice site, which prevents its recognition by the U1 small nuclear ribonucleoprotein (snRNP). Natural missense or silent alleles that promote too much exon inclusion are shown (top): N279K and L284L improve the enhancing activity of the multi-partite ESE; N296N disrupts the ESS; and S305N and S305S disrupt the inhibitory secondary structure. A small in-frame deletion allele, Δ280K (bottom), which abrogates the ESE, results in the skipping of exon 10.

undergoes surveillance before splicing (BOX 3). Moreover, the single-mechanism view is supported by the analysis of the spectrum of mutations in specific genes. For example, in the extensive database of human *TP53* (which encodes p53) tumour-associated point mutations (see link to the [IARC TP53 Mutation Database](#), which contains more than 15,000 entries<sup>82</sup>), there are many alleles for which the only mutation is a synonymous one (715 cases), and the total number of silent mutations is only slightly below that of nonsense mutations (5.6 versus 8.2% of all exonic point mutations). Furthermore, the nucleotides at which silent and nonsense mutations occur are frequently the same, which indicates that the same mechanism could be responsible for the effects of many of these *TP53* mutations.

Important mechanistic insights into how silent mutations can cause exon skipping come from studies of proximal spinal muscular atrophy (SMA), a paediatric neurodegenerative disorder caused by the loss of both functional copies of the survival of motor neuron 1 (*SMN1*) gene<sup>83</sup>. An almost identical paralogue of the *SMN1* gene, *SMN2*, is present in both normal and affected individuals, but although the two genes potentially encode indistinguishable proteins, *SMN2* is only partially functional. A crucial, translationally silent, single-nucleotide C → T difference between *SMN1* and *SMN2* at position +6 of exon 7 (C6T) results in the very inefficient inclusion of exon 7 in *SMN2* mRNA<sup>84,85</sup> (FIG. 4a). This low level of exon inclusion results in the production of an SMN protein that is indistinguishable from that encoded by *SMN1*. However, the predominant *SMN2*-spliced mRNA arises from the skipping of exon 7, and the resulting aberrant protein, SMNΔ7, is unstable and apparently inactive<sup>86</sup>. Full-length *SMN2*-derived SMN protein is therefore present in limiting amounts, particularly in motor neurons, and only partially complements the *SMN1* defect, giving rise to the SMA phenotype and accounting for the disease-modifier role of *SMN2* (REF. 87).

An analysis of exon 7 in the *SMN* genes has recently shown that a predicted ESE is disrupted in *SMN2* (REF. 88). Indeed, *in vivo* and *in vitro* experiments using *SMN1*-derived minigenes have shown that the inclusion of exon 7 in *SMN1* depends on the SR protein SF2/ASF directly contacting its cognate ESE motif. In particular, a second-site suppressor mutation in the disrupted *SMN2* ESE motif, which was designed to restore a good match to the SF2/ASF consensus motif (see BOX 1 figure), fully re-establishes exon 7 inclusion<sup>88</sup>. This result emphasizes the importance of developing computational tools that can accurately predict ESE or ESS elements using consensus recognition sequences for different splicing factors, which will allow neutral polymorphisms to be distinguished from mutations that lead to improper splicing.

Although loss-of-function mutations are likely to be more common than gain-of-function mutations, effects on pre-mRNA processing can also result from mutations that create or strengthen ESEs, ESSs or RNA secondary structures that exert positive or negative effects on splicing. An example that nicely illustrates

how several aspects of splicing regulation can be affected by silent and missense mutations involves **FTDP17** (frontotemporal dementia and parkinsonism associated with chromosome 17). This is an autosomal-dominant disorder that is related to **Alzheimer disease** and is caused by mutations in the microtubule-associated protein tau (*MAPT*) gene<sup>89</sup>.

*MAPT* encodes up to six developmentally regulated microtubule-associated protein isoforms, which vary in the number of microtubule-binding motifs they contain. *MAPT* mutations that are associated with FTDP17 tend to cluster in or around alternative exon 10, which encodes one of the microtubule-binding motifs and which can be alternatively spliced to give isoforms that contain three or four repeats (the 3R-tau and 4R-tau isoforms, respectively). Whereas some mutations directly affect the function of the microtubule-binding domain, most mutations increase the ratio of 4R-tau to 3R-tau by promoting exon 10 inclusion, which eventually leads to the formation, in neuronal or glial cells, of filamentous tau aggregates that are responsible for the pathology of the disorder<sup>89</sup>. The effect of these mutations involves several mechanisms. There is a cluster of exonic and intronic mutations that disrupt the formation of an RNA stem-loop structure at the 3' end of exon 10 (REFS 90,91). This secondary structure normally modulates alternative splicing by restricting the accessibility of the downstream 5' splice site to U1 snRNP (FIG. 4b). Other *cis*-elements participate in the fine-tuning of exon 10 inclusion. Accumulation of exon-10-containing isoforms can also result from mutations that either strengthen a multipartite ESE at the 5' end of the exon, or weaken an ESS immediately downstream<sup>92</sup> (FIG. 4b). In addition, a small in-frame deletion that abrogates the ESE and causes constitutive exon skipping has also been described in FTDP17 patients<sup>93</sup>. So, tight regulation of the proper ratio of 3R-tau to 4R-tau is very important for normal development of the nervous system, and even subtle variations in the efficiency of exon 10 inclusion can lead to significant pathology. Interestingly, many of the exonic point mutations that affect *MAPT* exon 10 usage are translationally silent (FIG. 4b; TABLE 2), and it is possible that there are other polymorphic alleles that do not cause a detectable phenotype, but either predispose or provide resistance to the disorder.

#### Evolutionary and human genetics implications

Silent mutations can be very informative, not only when they are present and give rise to an observable splicing defect, as in the above examples, but also when they are conspicuously absent. Their absence might, in fact, reflect a PURIFYING SELECTIVE PRESSURE against silent mutations in certain regions in the coding sequence of a gene<sup>94</sup>. When comparing the coding sequences of homologous genes from different mammals, it is assumed that no selection is at work for or against silent mutations (in mammals, CODON-USAGE BIAS is not thought to be important in selection, owing to small population sizes and a limited number of generations). However, a

phylogenetic analysis of *BRCA1* showed a strikingly low extent of SYNONYMOUS DIVERGENCE in a small subregion of the gene that spans exons 10 and 11, which could not be explained by codon-usage bias or other current models<sup>94</sup>. Negative selection against the silent mutations must therefore be at work. A possible explanation for this type of phenomenon is that the constraint comes from splicing regulation requirements<sup>95</sup>. In support of this view, this region of *BRCA1* undergoes alternative splicing and comprises two putative ESE motifs. These observations indicate that one or more functional elements might be needed to modulate the use of the various splice sites in this region of the gene<sup>95</sup>. Furthermore, synonymous divergence between mammalian species is generally higher in constitutive than in alternative exons<sup>96</sup>, possibly reflecting a stronger selective pressure to maintain regulatory elements that are required for alternative splicing. More generally, the presence of crucial sequence elements that are involved in splicing fidelity in the coding regions of genes is indicative of a further constraint on the evolution of exon sequences, in addition to the selective pressures due to codon-usage bias, and protein structure and function.

Just as missense and silent mutations can affect pre-mRNA splicing, coding single-nucleotide polymorphisms (cSNPs) can be expected to do the same. The only difference is that, by definition, SNPs occur at a >1% frequency in a population<sup>97</sup>. A recent systematic survey of cSNPs in 106 human genes estimated their frequency to be 1 in 346 bp (~50:50 synonymous:nonsynonymous changes), predicting a total of ~400,000 cSNPs in the whole genome<sup>98</sup>. The density of cSNPs that have been so far mapped in the whole human genome is ~1 per 1.08 kb (REF. 99). Considering the average length of human exons (~145 bp) and the fact that SNP mapping is still in progress, it is reasonable to assume that approximately every other exon might contain one or more cSNPs. Missense cSNPs have been proposed to account for a significant fraction of disease variation and to affect gene function in general<sup>98</sup>. However, because individual exons can contain several positive and negative *cis*-elements, any cSNP in a sequence that is important for splicing modulation might subtly influence the efficiency and tissue or stage specificity of alternative splicing. Such effects could contribute to phenotypic variability, to the variable PENETRANCE of mutations elsewhere in the gene and to cell-type-specific differences in gene expression. Moreover, the incremental accumulation of cSNPs might have had a role in the evolution of new gene functions by affecting the extent or accuracy of splicing events.

#### Perspectives

The growing appreciation of the potential effects of single-nucleotide changes in coding regions on the extent and accuracy of pre-mRNA splicing is expected to have a significant impact on the diagnosis and treatment of genetic diseases. The correct classification of mutations, in terms of their actual mechanism of gene inactivation, is essential for understanding structure–function relationships in the corresponding protein, for assessing the

#### PURIFYING SELECTIVE PRESSURE

Negative selection that tends to eliminate divergent mutations and maintain the current gene sequence.

#### CODON-USAGE BIAS

Species-specific frequencies of the alternative codons that specify the same amino acid.

#### SYNONYMOUS DIVERGENCE

The accumulation of silent mutations due to neutral evolution.

#### cSNPs

Single-nucleotide polymorphisms that are present in the coding regions of a gene.

#### PENETRANCE

The proportion of genotypically mutant organisms that show the mutant phenotype.

phenotypic risk in individuals with familial disease predispositions and for devising new therapies.

There is considerable promise in developing methods to correct splicing defects that are caused by particular mutations. Antisense strategies have been used to prevent the use of inappropriate splice sites or to restore a competitive advantage to splice sites that had been weakened by mutation<sup>100,101</sup>. Although drugs that can selectively modulate the activity of the splicing machinery are not yet available, ongoing high-throughput

screens have started to identify compounds that can somehow promote the inclusion of exon 7 in *SMN2* (see FIG. 4a), an approach that might eventually help all SMA patients<sup>102–104</sup>. Similar approaches might be possible to treat other genetic diseases that are caused by specific silent or missense alleles that cause exon skipping. The more we understand the mechanisms that define exons and modulate their inclusion, the sooner that effective compounds or new therapeutic strategies can be developed.

- Burge, C. B., Tuschl, T. & Sharp, P. A. In *The RNA World II* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 525–560 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
- Will, C. L. & Lührmann, R. Spliceosomal U snRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301 (2001).
- Sun, H. & Chasin, L. A. Multiple splicing defects in an intronic false exon. *Mol. Cell. Biol.* **20**, 6414–6425 (2000). **This study provides a systematic sequence analysis of the numerous pseudo-exons in *HPRT*, and investigates some of the features that distinguish authentic exons from pseudo-exons.**
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Graveley, B. R. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**, 100–107 (2001).
- Schmucker, D. *et al.* *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–684 (2000). **The combinatorial use of alternative exons allows this single gene in *Drosophila* to express 38,016 potential isoforms.**
- Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nature Genet.* **30**, 13–19 (2002).
- Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA* **6**, 1197–1211 (2000).
- Hastings, M. L. & Krainer, A. R. Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* **13**, 302–309 (2001).
- López, A. J. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* **32**, 279–305 (1998).
- Liu, H. X., Zhang, M. & Krainer, A. R. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* **12**, 1998–2012 (1998).
- Schaal, T. D. & Maniatis, T. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol. Cell. Biol.* **19**, 261–273 (1999).
- Blencowe, B. J. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* **25**, 106–110 (2000).
- Birney, E., Kumar, S. & Krainer, A. R. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**, 5803–5816 (1993).
- Robberson, B. L., Cote, G. J. & Berget, S. M. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**, 84–94 (1990). **This paper proposed the concept that the unit of recognition during the early stages of spliceosome assembly is the exon, rather than the intron, such that there is communication between the 3' and 5' splice sites that flank an exon.**
- Zuo, P. & Maniatis, T. The splicing factor U2AF35 mediates critical protein–protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* **10**, 1356–1368 (1996).
- Graveley, B. R., Hertel, K. J. & Maniatis, T. The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA* **7**, 806–818 (2001).
- Kan, J. L. & Green, M. R. Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. *Genes Dev.* **13**, 462–471 (1999).
- Zhu, J. & Krainer, A. R. Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes Dev.* **14**, 3166–3178 (2000).
- Zhu, J., Mayeda, A. & Krainer, A. R. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol. Cell* **8**, 1351–1361 (2001). **This paper showed that a factor that is bound to an exonic splicing silencer can repress exon recognition, depending on the factors that are available to recognize the splicing enhancers elsewhere in the exon.**
- Tanaka, K., Watakabe, A. & Shimura, Y. Polypurine sequences within a downstream exon function as a splicing enhancer. *Mol. Cell. Biol.* **14**, 1347–1354 (1994). **The first systematic analysis of a purine-rich exonic splicing enhancer, which showed that the purine-rich base composition is not sufficient to promote splicing.**
- Ramchatesingh, J., Zahler, A. M., Neugebauer, K. M., Roth, M. B. & Cooper, T. A. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol. Cell. Biol.* **15**, 4898–4907 (1995).
- Watakabe, A., Tanaka, K. & Shimura, Y. The role of exon sequences in splice site selection. *Genes Dev.* **7**, 407–418 (1993).
- Staknis, D. & Reed, R. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell. Biol.* **14**, 7670–7682 (1994).
- Tian, H. & Kole, R. Selection of novel exon recognition elements from a pool of random sequences. *Mol. Cell. Biol.* **15**, 6291–6298 (1995). **This paper reports the development of the functional SELEX approach to identify sequences that can enhance splicing. Enhancer sequences were found to be very diverse and not always purine rich.**
- Woerfel, G. & Bindereif, A. *In vitro* selection of exonic splicing enhancer sequences: identification of novel CD44 enhancers. *Nucleic Acids Res.* **29**, 3204–3211 (2001).
- Coulter, L. R., Landree, M. A. & Cooper, T. A. Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.* **17**, 2143–2150 (1997).
- Schaal, T. D. & Maniatis, T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.* **19**, 1705–1719 (1999).
- Liu, H. X., Chew, S. L., Cartegni, L., Zhang, M. Q. & Krainer, A. R. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol. Cell. Biol.* **20**, 1063–1071 (2000).
- Tacke, R. & Manley, J. L. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* **14**, 3540–3551 (1995).
- Fairbrother, W. G. & Chasin, L. A. Human genomic sequences that inhibit splicing. *Mol. Cell. Biol.* **20**, 6816–6825 (2000).
- Caputi, M. *et al.* A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. *Nucleic Acids Res.* **22**, 1018–1022 (1994).
- Amendt, B. A., Si, Z. H. & Stoltzfus, C. M. Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. *Mol. Cell. Biol.* **15**, 4606–4615 (1995).
- Staffa, A. & Cochrane, A. Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. *Mol. Cell. Biol.* **15**, 4597–4605 (1995).
- Zheng, Z. M., Huynen, M. & Baker, C. C. A pyrimidine-rich exonic splicing suppressor binds multiple RNA splicing factors and inhibits spliceosome assembly. *Proc. Natl. Acad. Sci. USA* **95**, 14088–14093 (1998).
- Del Gatto-Konczak, F., Olive, M., Gesnel, M. C. & Breathnach, R. hnRNP A1 recruited to an exon *in vivo* can function as an exon splicing silencer. *Mol. Cell. Biol.* **19**, 251–260 (1999).
- Chew, S. L., Baginsky, L. & Eperon, I. C. An exonic splicing silencer in the testes-specific DNA ligase III $\beta$  exon. *Nucleic Acids Res.* **28**, 402–410 (2000).
- Smith, C. W. & Valcárcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).
- Wagner, E. J. & Garcia-Blanco, M. A. Polypyrimidine tract binding protein antagonizes exon definition. *Mol. Cell. Biol.* **21**, 3281–3288 (2001).
- Dreyfuss, G., Matunis, M. J., Piñol-Roma, S. & Burd, C. G. hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289–321 (1993).
- Krecic, A. M. & Swanson, M. S. hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* **11**, 363–371 (1999).
- Chen, C. D., Kobayashi, R. & Helfman, D. M. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat  $\beta$ -tropomyosin gene. *Genes Dev.* **13**, 593–606 (1999).
- Blanchette, M. & Chabot, B. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* **18**, 1939–1952 (1999).
- Chabot, B. Directing alternative splicing: cast and scenarios. *Trends Genet.* **12**, 472–478 (1996).
- Krawczak, M., Reiss, J. & Cooper, D. N. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* **90**, 41–54 (1992). **This thorough survey of human mutations shows that ~15% of point mutations that are associated with human genetic diseases affect the conserved signals at the ends of introns and result in aberrant splicing patterns.**
- Ars, E. *et al.* Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum. Mol. Genet.* **9**, 237–247 (2000).
- Teracka, S. N. *et al.* Splicing defects in the ataxia-telangiectasia gene, *ATM*: underlying mutations and consequences. *Am. J. Hum. Genet.* **64**, 1617–1631 (1999). **In references 48 and 49, many mutations in *NF1* and *ATM*, respectively, were systematically analysed at the gene, mRNA and protein level. The authors show that ~50% of the mutations, including several in the coding sequence, result in aberrant splicing patterns.**
- Hentze, M. W. & Kulozik, A. E. A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* **96**, 307–310 (1999).
- Mendell, J. T. & Dietz, H. C. When the message goes awry: disease-producing mutations that influence mRNA content and performance. *Cell* **107**, 411–414 (2001).
- Dietz, H. C. *et al.* The skipping of constitutive exons *in vivo* induced by nonsense mutations. *Science* **259**, 680–683 (1993). **This study provided the first evidence that certain nonsense mutations can affect the fidelity of pre-mRNA splicing, and cause mutation-containing exons to be skipped.**
- Dietz, H. C. & Kendzior, R. J. Jr. Maintenance of an open reading frame as an additional level of scrutiny during splice site selection. *Nature Genet.* **8**, 183–188 (1994).
- Gersappe, A., Burger, L. & Pintel, D. J. A premature termination codon in either exon of minute virus of mice P4 promoter-generated pre-mRNA can inhibit nuclear splicing of the intervening intron in an open reading frame-dependent manner. *J. Biol. Chem.* **274**, 22452–22458 (1999).

55. Gersappe, A. & Pintel, D. J. A premature termination codon interferes with the nuclear function of an exon splicing enhancer in an open reading frame-dependent manner. *Mol. Cell. Biol.* **19**, 1640–1650 (1999).
56. Muhlemann, O. *et al.* Precursor RNAs harboring nonsense codons accumulate near the site of transcription. *Mol. Cell* **8**, 33–43 (2001).
57. Urlaub, G., Mitchell, P. J., Ciudad, C. J. & Chasin, L. A. Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. *Mol. Cell. Biol.* **9**, 2868–2880 (1989).
58. Maquat, L. E. When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. *RNA* **1**, 453–465 (1995).
59. Li, S. & Wilkinson, M. F. Nonsense surveillance in lymphocytes? *Immunity* **8**, 135–141 (1998).
60. Brogna, S. Pre-mRNA processing: insights from nonsense. *Curr. Biol.* **11**, R838–R841 (2001).
61. Etchison, D. & Etchison, J. R. Monoclonal antibody-aided characterization of cellular p220 in uninfected and poliovirus-infected HeLa cells: subcellular distribution and identification of conformers. *J. Virol.* **61**, 2702–2710 (1987).
62. Dostie, J., Lejbkowitz, F. & Sonenberg, N. Nuclear eukaryotic initiation factor 4E (eIF4E) colocalizes with splicing factors in speckles. *J. Cell Biol.* **148**, 239–247 (2000).
63. Mangiarotti, G. Coupling of transcription and translation in *Dicotylellum discoideum* nuclei. *Biochemistry* **38**, 3996–4000 (1999).
64. Iborra, F. J., Jackson, D. A. & Cook, P. R. Coupled transcription and translation within nuclei of mammalian cells. *Science* **293**, 1139–1142 (2001).
- This recent paper reports evidence that transcription-coupled translation might take place in the nucleus of mammalian cells.**
65. Ahn, A. H. & Kunkel, L. M. The structural and functional diversity of dystrophin. *Nature Genet.* **3**, 283–291 (1993).
66. Koenig, M. *et al.* The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. *Am. J. Hum. Genet.* **45**, 498–506 (1989).
67. Shiga, N. *et al.* Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. *J. Clin. Invest.* **100**, 2204–2210 (1997).
68. Ginjaar, I. B. *et al.* Dystrophin nonsense mutation induces different levels of exon 29 skipping and leads to variable phenotypes within one BMD family. *Eur. J. Hum. Genet.* **8**, 793–796 (2000).
69. Liu, W., Qian, C. & Francke, U. Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. *Nature Genet.* **16**, 328–329 (1997).
70. Hoffmeyer, S. *et al.* Nearby stop codons in exons of the neurofibromatosis type 1 gene are disparate splice effectors. *Am. J. Hum. Genet.* **62**, 269–277 (1998).
71. Tu, M., Tong, W., Perkins, R. & Valentine, C. R. Predicted changes in pre-mRNA secondary structure vary in their association with exon skipping for mutations in exons 2, 4, and 8 of the *Hprt* gene and exon 51 of the fibrillin gene. *Mutat. Res.* **432**, 15–32 (2000).
72. Valentine, C. R. & Heflich, R. H. The association of nonsense mutation with exon-skipping in *hprt* mRNA of Chinese hamster ovary cells results from an artifact of RT-PCR. *RNA* **3**, 660–676 (1997).
73. Valentine, C. R. The association of nonsense codons with exon skipping. *Mutat. Res.* **411**, 87–117 (1998).
- A thorough survey of nonsense (and missense) mutations that are associated with skipping of exons in human and rodent genes, which argues that most cases of nonsense-associated exon skipping are due to interference with exon definition or to non-quantitative RNA measurements.**
74. Muro, A. F. *et al.* Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol. Cell. Biol.* **19**, 2657–2671 (1999).
75. Gibson, R. A., Hajianpour, A., Murer-Orlando, M., Buchwald, M. & Mathew, C. G. A nonsense mutation and exon skipping in the Fanconi anaemia group C gene. *Hum. Mol. Genet.* **2**, 797–799 (1993).
76. Lo Ten Foe, J. R. *et al.* Exon 6 skipping in the Fanconi anaemia C gene associated with a nonsense/missense mutation (775C→T) in exon 5: the first example of a nonsense mutation in one exon causing skipping of another downstream. *Hum. Mutat.* (Suppl. 1), S25–S27 (1998).
77. Cooper, T. A. & Mattox, W. The regulation of splice-site selection, and its role in human disease. *Am. J. Hum. Genet.* **61**, 259–266 (1997).
78. Mazoyer, S. *et al.* A *BRCA1* nonsense mutation causes exon skipping. *Am. J. Hum. Genet.* **62**, 713–715 (1998).
79. Liu, H. X., Cartegni, L., Zhang, M. Q. & Krainer, A. R. A mechanism for exon skipping caused by nonsense or missense mutations in *BRCA1* and other genes. *Nature Genet.* **27**, 55–58 (2001).
- In this paper, exon skipping that is associated with a premature termination codon is shown to be due to the inactivation of an exonic splicing enhancer. A missense mutation that similarly disrupts the enhancer motif also causes exon skipping, whereas a nonsense mutation that does not disrupt the motif maintains exon inclusion.**
80. Kronenberg, M., Siu, G., Hood, L. E. & Shastri, N. The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition. *Annu. Rev. Immunol.* **4**, 529–591 (1986).
81. Fang, W. *et al.* Frequent aberrant immunoglobulin gene rearrangements in pro-B cells revealed by a *bcl-xL* transgene. *Immunity* **4**, 291–299 (1996).
82. Hernandez-Boussard, T., Rodriguez-Tome, P., Montesano, R. & Hainaut, P. IARC p53 mutation database: a relational database to compile and analyze p53 mutations in human tumors and cell lines. International Agency for Research on Cancer. *Hum. Mutat.* **14**, 1–8 (1999).
83. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).
84. Lorson, C. L., Hahnen, E., Androphy, E. J. & Wirth, B. A single nucleotide in the *SMN* gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl Acad. Sci. USA* **96**, 6307–6311 (1999).
85. Monani, U. R. *et al.* A single nucleotide difference that alters splicing patterns distinguishes the SMA gene *SMN1* from the copy gene *SMN2*. *Hum. Mol. Genet.* **8**, 1177–1183 (1999).
- References 84 and 85 show that a single, translationally silent nucleotide change in exon 7 of *SMN1* converts this constitutive exon into an alternative exon. This observation shows why the closely related *SMN2* gene acts as a disease modifier in spinal muscular atrophy.**
86. Lorson, C. L. & Androphy, E. J. An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene *SMN*. *Hum. Mol. Genet.* **9**, 259–265 (2000).
87. Jablonka, S., Rossoll, W., Schrank, B. & Sendtner, M. The role of *SMN* in spinal muscular atrophy. *J. Neural.* **247**(Suppl. 1), I37–I42 (2000).
88. Cartegni, L. & Krainer, A. R. Disruption of an SF2/ASF-dependent exonic splicing enhancer in *SMN2* causes spinal muscular atrophy in the absence of *SMN1*. *Nature Genet.* 2002 March 4 (DOI 10.1038/ng854)
- In this study, exon 7 in *SMN1* is shown to contain an exonic splicing enhancer that is recognized by a specific SR protein. This motif is abrogated by the single nucleotide change in *SMN2* that inhibits exon inclusion.**
89. Lee, V. M., Goedert, M. & Trojanowski, J. Q. Neurodegenerative tauopathies. *Annu. Rev. Neurosci.* **24**, 1121–1159 (2001).
90. Hasegawa, M., Smith, M. J., Iijima, M., Tabira, T. & Goedert, M. *FTDP-17* mutations N279K and S305N in tau produce increased splicing of exon 10. *FEBS Lett.* **443**, 93–96 (1999).
91. Varani, L. *et al.* Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc. Natl Acad. Sci. USA* **96**, 8229–8234 (1999).
92. D'Souza, I. & Schellenberg, G. D. Determinants of 4-repeat tau expression. Coordination between enhancing and inhibitory splicing sequences for exon 10 inclusion. *J. Biol. Chem.* **275**, 17700–17709 (2000).
93. D'Souza, I. *et al.* Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. *Proc. Natl Acad. Sci. USA* **96**, 5598–5603 (1999).
94. Hurst, L. D. & Pal, C. Evidence for purifying selection acting on silent sites in *BRCA1*. *Trends Genet.* **17**, 62–65 (2001).
95. Orban, T. I. & Olah, E. Purifying selection on silent sites — a constraint from splicing regulation? *Trends Genet.* **17**, 252–253 (2001).
96. Iida, K. & Akashi, H. A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. *Gene* **261**, 93–105 (2000).
97. Brookes, A. J. The essence of SNPs. *Gene* **234**, 177–186 (1999).
98. Cargill, M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* **22**, 231–238 (1999).
99. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
100. Sierakowska, H., Agrawal, S. & Kole, R. Antisense oligonucleotides as modulators of pre-mRNA splicing. *Methods Mol. Biol.* **133**, 223–233 (2000).
101. Lim, S. R. & Hertel, K. J. Modulation of survival motor neuron pre-mRNA splicing by inhibition of alternative 3' splice site pairing. *J. Biol. Chem.* **276**, 45476–45483 (2001).
102. Andreassi, C. *et al.* Aclarubicin treatment restores SMN levels to cells derived from type I spinal muscular atrophy patients. *Hum. Mol. Genet.* **10**, 2841–2849 (2001).
103. Chang, J. G. *et al.* Treatment of spinal muscular atrophy by sodium butyrate. *Proc. Natl Acad. Sci. USA* **98**, 9808–9813 (2001).
104. Zhang, M. L., Lorson, C. L., Androphy, E. J. & Zhou, J. An *in vivo* reporter system for measuring increased inclusion of exon 7 in *SMN2* mRNA: potential therapy of SMA. *Gene Ther.* **8**, 1532–1538 (2001).
105. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
106. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61–68 (1989).
107. Cooper, T. A. *In vivo* SELEX in vertebrate cells. *Methods Mol. Biol.* **118**, 405–417 (1999).
108. Maquat, L. E. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 849–868 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2000).
109. Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. *EMBO J.* **19**, 6860–6869 (2000).
- This study shows that there is an exon–junction complex that retains information about intron position after the completion of pre-mRNA splicing. This complex might be used to distinguish authentic from premature termination codons during NMD.**
110. Ishigaki, Y., Li, X., Serin, G. & Maquat, L. E. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cel* **106**, 607–617 (2001).
111. Hofmann, Y., Lorson, C. L., Stamm, S., Androphy, E. J. & Wirth, B. Htra2- $\beta$ 1 stimulates an exonic splicing enhancer and can restore full-length *SMN* expression to survival motor neuron 2 (*SMN2*). *Proc. Natl Acad. Sci. USA* **97**, 9618–9623 (2000).
112. Cavaloc, Y., Bourgeois, C. F., Kister, L. & Stévenin, J. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* **5**, 468–483 (1999).
113. Lou, H., Neugebauer, K. M., Gagel, R. F. & Bergel, S. M. Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol. Cell. Biol.* **18**, 4977–4985 (1998).
114. Tacke, R., Chen, Y. & Manley, J. L. Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc. Natl Acad. Sci. USA* **94**, 1148–1153 (1997).
115. Tacke, R., Tohyama, M., Ogawa, S. & Manley, J. L. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cel* **93**, 139–148 (1998).
116. Ozsahin, H. *et al.* Adenosine deaminase deficiency in adults. *Blood* **89**, 2849–2855 (1997).
117. Das, S. *et al.* Diverse mutations in patients with Menkes disease often lead to exon skipping. *Am. J. Hum. Genet.* **55**, 883–889 (1994).
118. Kerem, E. *et al.* A missense cystic fibrosis transmembrane conductance regulator mutation with variable phenotype. *Pediatrics* **100**, E5 (1997).
119. Tzetis, M., Efthymiadou, A., Doudounakis, S. & Kanavakis, E. Qualitative and quantitative analysis of mRNA associated with four putative splicing mutations (621 + 3A → G, 2751 + 2T → A, 296 + 1G → C, 1717-9T → C-D565G) and one nonsense mutation (E822X) in the *CFTFR* gene. *Hum. Genet.* **109**, 592–601 (2001).
120. Theophilus, B. D., Enayat, M. S., Williams, M. D. & Hill, F. G. Site and type of mutations in the factor VIII gene in patients and carriers of haemophilia A. *Haemophilia* **7**, 381–391 (2001).
121. Dreumont, N. *et al.* A missense mutation (Q279R) in the fumarylacetoacetate hydrolase gene, responsible for hereditary tyrosinemia, acts as a splicing mutation. *BMC Genet.* **2**, 9 (2001).
122. Babcock, D., Gasner, C., Francke, U. & Maslen, C. A single mutation that results in an Asp to His substitution and partial exon skipping in a family with congenital contractural arachnodactyly. *Hum. Genet.* **103**, 22–28 (1998).
123. Wang, X. Molecular characterization of a novel defect occurring *de novo* associated with erythropoietic

- protoporphyrin. *Biochim. Biophys. Acta* **1316**, 149–152 (1996).
124. Wakamatsu, N., Kobayashi, H., Miyatake, T. & Tsuji, S. A novel exon mutation in the human  $\beta$ -hexosaminidase  $\beta$ -subunit gene affects 3' splice site selection. *J. Biol. Chem.* **267**, 2406–2413 (1992).
  125. Mustajoki, S. *et al.* Three splicing defects, an insertion, and two missense mutations responsible for acute intermittent porphyria. *Hum. Genet.* **102**, 541–548 (1998).
  126. Kanai, N. *et al.* A G to A transition at the last nucleotide of exon 6 of the  $\gamma$ c gene (868G  $\rightarrow$  A) may result in either a splice or missense mutation in patients with X-linked severe combined immunodeficiency. *Hum. Genet.* **104**, 36–42 (1999).
  127. Vockley, J. *et al.* Exon skipping in IVD RNA processing in isovaleric acidemia caused by point mutations in the coding region of the IVD gene. *Am. J. Hum. Genet.* **66**, 356–367 (2000).
  128. Nystrom-Lahti, M. *et al.* Missense and nonsense mutations in codon 659 of *MLH1* cause aberrant splicing of messenger RNA in HNPCC kindreds. *Genes Chromosomes Cancer* **26**, 372–375 (1999).
  129. Chun, K. *et al.* Mutations in the X-linked E1 $\alpha$  subunit of pyruvate dehydrogenase: exon skipping, insertion of duplicate sequence, and missense mutations leading to the deficiency of the pyruvate dehydrogenase complex. *Am. J. Hum. Genet.* **56**, 558–569 (1995).
  130. Vuillaumier-Barrot, S. *et al.* Characterization of the 415G  $\rightarrow$  A (E139K) *PMM2* mutation in carbohydrate-deficient glycoprotein syndrome type  $\alpha$  disrupting a splicing enhancer resulting in exon 5 skipping. *Hum. Mutat.* **14**, 543–544 (1999).
  131. Huang, C. H. *et al.* Molecular basis for Rh(null) syndrome: identification of three new missense mutations in the *Rh50* glycoprotein gene. *Am. J. Hematol.* **62**, 25–32 (1999).
  132. Montero, M. *et al.* A silent mutation in exon 14 of the *APC* gene is associated with exon skipping in a FAP family. *J. Med. Genet.* **38**, 863–867 (2001).
  133. Hellwinkel, O. J. *et al.* A unique exonic splicing mutation in the human androgen receptor gene indicates a physiologic relevance of regular androgen receptor transcript variants. *J. Clin. Endocrinol. Metab.* **86**, 2569–2575 (2001).
  134. Chen, W. *et al.* Silent nucleotide substitution in the sterol 27-hydroxylase gene (*CYP27*) leads to alternative pre-mRNA splicing by activating a cryptic 5' splice site at the mutant codon in cerebrotendinous xanthomatosis patients. *Biochemistry* **37**, 4420–4428 (1998).
  135. Ploos van Amstel, J. K. *et al.* Hereditary tyrosinemia type 1: novel missense, nonsense and splice consensus mutations in the human fumarylacetoacetate hydrolase gene; variability of the genotype–phenotype relationship. *Hum. Genet.* **97**, 51–59 (1996).
  136. Akli, S. *et al.* A 'G' to 'A' mutation at position –1 of a 5' splice site in a late infantile form of Tay-Sachs disease. *J. Biol. Chem.* **265**, 7324–7330 (1990).
  137. Llewellyn, D. H. *et al.* Acute intermittent porphyria caused by defective splicing of porphobilinogen deaminase RNA: a synonymous codon mutation at 22 bp from the 5' splice site causes skipping of exon 3. *J. Med. Genet.* **33**, 437–438 (1996).
  138. Jin, Y. *et al.* Glanzmann thrombasthenia. Cooperation between sequence variants in *cis* during splice site selection. *J. Clin. Invest.* **98**, 1745–1754 (1996).
  139. Ries, S. *et al.* Different missense mutations in histidine-108 of lysosomal acid lipase cause cholesteryl ester storage disease in unrelated compound heterozygous and hemizygous individuals. *Hum. Mutat.* **12**, 44–51 (1998).
  140. Kohonen-Corish, M. *et al.* RNA-based mutation screening in hereditary nonpolyposis colorectal cancer. *Am. J. Hum. Genet.* **59**, 818–824 (1996).
  141. Fahsold, R. *et al.* Minor lesion mutational spectrum of the entire *NF1* gene does not explain its high mutability but points to a functional domain upstream of the GAP-related domain. *Am. J. Hum. Genet.* **66**, 790–818 (2000).
  142. Chao, H. K., Hsiao, K. J. & Su, T. S. A silent mutation induces exon skipping in the phenylalanine hydroxylase gene in phenylketonuria. *Hum. Genet.* **108**, 14–19 (2001).
  143. De Meirleir, L. *et al.* Aberrant splicing of exon 6 in the pyruvate dehydrogenase-E1 $\alpha$  mRNA linked to a silent mutation in a large family with Leigh's encephalomyelopathy. *Pediatr. Res.* **36**, 707–712 (1994).
  144. Kanno, H. *et al.* Frame shift mutation, exon skipping, and a two-codon deletion caused by splice site mutations account for pyruvate kinase deficiency. *Blood* **89**, 4213–4218 (1997).
  145. Jacobsen, M. *et al.* A point mutation in *PTPRC* is associated with the development of multiple sclerosis. *Nature Genet.* **26**, 495–499 (2000).
  146. Imamura, T., Okano, Y., Shintaku, H., Hase, Y. & Isshiki, G. Molecular characterization of 6-pyruvoyl-tetrahydropterin synthase deficiency in Japanese patients. *J. Hum. Genet.* **44**, 163–168 (1999).
  147. Auricchio, A. *et al.* Double heterozygosity for a RET substitution interfering with splicing and an EDNRB missense mutation in Hirschsprung disease. *Am. J. Hum. Genet.* **64**, 1216–1221 (1999).
  148. Ferrari, S. *et al.* Mutations of *CD40* gene cause an autosomal recessive form of immunodeficiency with hyper IgM. *Proc. Natl Acad. Sci. USA* **98**, 12614–12619 (2001).
  149. Mendez, M. *et al.* Familial porphyria cutanea tarda: characterization of seven novel uroporphyrinogen decarboxylase mutations and frequency of common hemochromatosis alleles. *Am. J. Hum. Genet.* **63**, 1363–1375 (1998).

Acknowledgements

We thank M. Hastings for valuable comments on the manuscript and members of our labs for stimulating discussions. Work in the authors' labs is supported by the National Institute of General Medical Sciences, National Institute of Neurological Diseases and Stroke, National Cancer Institute and Andrew's Buddies Corporation (L.C. and A.R.K.), and by the Joint Research Board of St Bartholomew's Hospital (S.L.C.).

 Online links

**DATABASES**

The following terms in this article are linked online to:  
**LocusLink:** <http://www.ncbi.nlm.nih.gov/LocusLink/list.cgi>  
*ATM* | *BRCA1* | *DMD* | *Dscam* | eIF4E | eIF4G | *FANCC* | *FBN1* | *FN1* | *FTDP17* | hnRNP | hnRNP A1 | *HPRT1* | *IGHM* | *MAPT* | *NF1* | *PTB* | *SF2/ASF* | *SMN1* | *SMN2* | *TP53* | *TRB* | *U2AF*  
**OMIM:** <http://www.ncbi.nlm.nih.gov/Omim/>  
 Alzheimer disease | ataxia telangiectasia | Duchenne and Becker muscular dystrophy | Marfan syndrome | neurofibromatosis type 1 | spinal muscular atrophy

**FURTHER INFORMATION**

**Adrian Krainer's lab:**  
<http://www.cshl.org/public/SCIENCE/krainer.html>  
**Encyclopedia of Life Sciences:** <http://www.els.net>  
 Spliceosomal machinery  
**Human Gene Mutation Database:**  
<http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html>  
**IARC TP53 Mutation Database:** <http://www.iarc.fr/p53/>  
**Shern Chew's lab:**  
<http://www.mds.qmw.ac.uk/endocrinology/chew.html>  
**Access to this interactive links box is free online.**