Splicing in disease: disruption of the splicing code and the decoding machinery

Guey-Shin Wang and Thomas A. Cooper*‡*

Abstract | Human genes contain a dense array of diverse *cis*-acting elements that make up a code required for the expression of correctly spliced mRNAs. Alternative splicing generates a highly dynamic human proteome through networks of coordinated splicing events. *Cis*- and *trans*-acting mutations that disrupt the splicing code or the machinery required for splicing and its regulation have roles in various diseases, and recent studies have provided new insights into the mechanisms by which these effects occur. An unexpectedly large fraction of exonic mutations exhibit a primary pathogenic effect on splicing. Furthermore, normal genetic variation significantly contributes to disease severity and susceptibility by affecting splicing efficiency.

Spliceosome
The basal splicing machinery, which is made up of 5 small nuclear ribonucleoproteins (snRNPs) and more than 150 additional proteins.

Pseudoexons
Intronic sequences that fortuitously resemble exons because of matches to 3′ and 5′ splice sites that are not normally spliced.

**Departments of Pathology, Baylor College of Medicine, Houston, Texas 77030, USA.*
‡Department of Molecular and Cellular Biology, Baylor College of Medicine.
Correspondence to T.A.C.
e-mail: tcooper@bcm.edu
doi:10.1038/nrg2164
Published online
29 August 2007

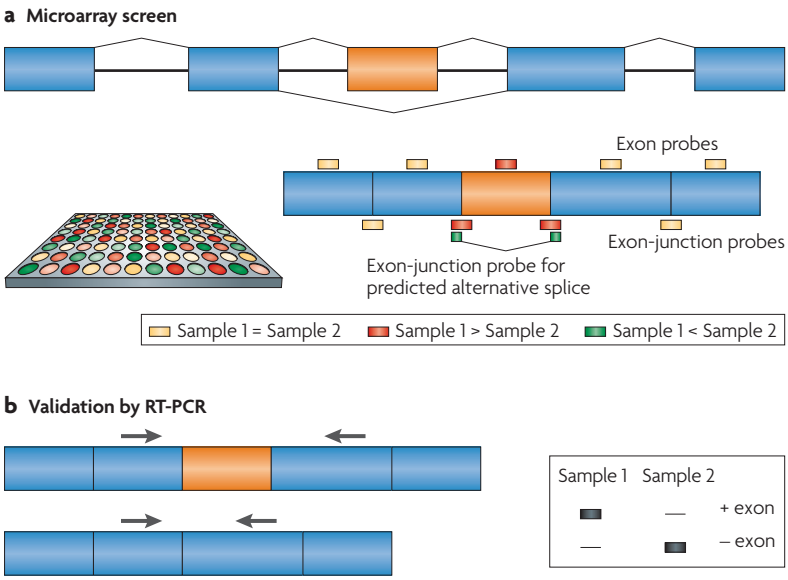
Human genes express remarkably complex pre-mRNAs, containing an average of eight exons, with introns making up 90% of the transcription unit. For each pre-mRNA that is expressed, its exons — which are separated by introns of up to hundreds of thousands of nucleotides in length — must be precisely joined together to maintain the reading frame. In addition, at least 70% of human genes express multiple mRNAs through alternative splicing of exons or exon segments[1]. The challenges to the splicing machinery (called the spliceosome) are to identify bone fide exons, ignore pre-mRNA segments that resemble exons (pseudoexons), join contiguous exons without inadvertent skipping and appropriately regulate alternative splicing to meet the physiological requirements of cells and tissues. These challenges are met through several intercommunicating layers of *cis*-acting elements that distinguish exons from introns, direct the spliceosome to the correct nucleotides for exon joining and intron removal, and serve as binding sites for auxiliary factors that regulate alternative splicing. These elements make up what is now recognized as a 'splicing code', which appears to be particularly dense within and around exons.

Alternative splicing greatly expands the information content and versatility of the transcriptome through the expression of multiple different mRNAs from individual genes[2]. The result of alternative splicing is the introduction of variable segments within otherwise identical mRNAs. About 80% of this variability falls within ORFs, greatly expanding the human proteome[3], and the 20% that falls within untranslated regions affects *cis*-acting elements that control mRNA stability, translation efficiency (including miRNA binding sites), and mRNA localization. Furthermore, one-third of alternative splicing events introduce premature termination codons (PTCs), which cause mRNA degradation by nonsense-mediated decay (NMD)[4]. Therefore, regulation of alternative splicing controls the temporal and spatial expression of functionally diverse isoforms, on–off regulation by NMD, or other post-transcriptional regulatory responses. In the past few years, rapid advances in microarray and associated computational technologies have allowed detailed analyses of individual exons on a genome-wide scale (BOX 1). Analyses of global splicing transitions in different biological systems has revealed coordinated regulation of subsets of alternative splicing events, representing a previously unappreciated layer of gene expression control[5].

In this Review, we describe recent advances towards understanding the diverse mechanisms by which disrupted splicing has a determinative role in disease, as a direct cause, a modifier of disease severity and a determinant of disease susceptibility. An unexpectedly large fraction of exonic mutations cause disease by disrupting the splicing code, rather than by the predicted disruption of the protein reading frame. Similarly, normal genetic variation in the splicing code creates differences in splicing efficiencies, which modifies disease severity. It is now clear that genetic variants that are linked with a phenotype, whether these are disease-causing mutations or common SNPs, need to be evaluated for disruption of the correct splicing pattern(s). Mutations that alter the

trans-acting splicing environment are also implicated in various diseases, potentially affecting the expression of multiple genes. Recent studies have identified examples in which *trans*-acting mechanisms contribute to disease, including a potentially important role of one *trans*-regulator of splicing in cancer. In addition to providing insights into disease mechanisms, investigation of the pathogenic roles of splicing has also significantly contributed to the discovery and understanding of the intricacies of normal splicing and its regulation.

Box 1 | **Global and computational analyses of alternative splicing**



Splicing microarrays are used for the large-scale identification of splicing differences between two RNA populations. Microarrays typically contain spotted oligonucleotide probes that are complementary to individual exons and/or exon–exon junctions for thousands of genes, and are hybridized to Cy3- and Cy5-labelled cDNA derived from the RNA samples that are being compared (corresponding to the green and red hybridization signals in panel **a**). Alternative splicing results in a loss of signal from specific probes relative to the other probes, and a gain of signal for probes that are specific for junctions in alternatively spliced variants. Erratic hybridization characteristics among individual oligonucleotide probes leads to variability independently of mRNA levels, which can be difficult to interpret. Splicing microarray reliability can be improved by higher coverage using both exon probes and exon-junction probes. For example, exon skipping results in a loss of signal from one probe when only exon probes are used, a loss of signal from two probes when only exon-junction probes are used, and a loss of signal from three probes when both exon-junction and exon probes are used. When the alternative splicing pattern is known, exon-junction probes designed for an exon-skipping variant (represented by the green probe on the mRNA in panel **a**) provide a corresponding gain of signal. A second challenge is to detect a change in exon use in a background of dramatically different or low mRNA steady-state levels. Uses for splicing microarrays range from a survey of all the exons in a large number of genes for the discovery of novel alternative splicing patterns[1], to a quantitative readout for all know splice variants from a limited number of genes[85]. The predicted splicing events are validated by RT-PCR (panel **b**).

Splicing microarrays typically contain 3–4 times more probes per mRNA than standard microarrays, and coverage is more uniform because standard microarrays are biased towards the mRNA 3′ end. Therefore, splicing microarrays also provide a robust readout of mRNA steady-state levels. The simultaneous analysis of mRNA levels and alternative splicing in different biological systems has revealed that qualitative modulation of isoform composition is typically independent of quantitative changes in the total gene output[85–87].

It is important to note that although splicing microarrays and standard RT-PCR approaches identify variability within limited regions of an mRNA, these approaches do not identify the predominating full-length mRNAs and protein isoforms that result from the combinatorial output of multiple variably spliced regions. An approach for identifying predominant mRNAs is discussed in BOX 2.

In an example of the successful use of splicing microarrays, this approach was used to identify splicing events that are regulated by the neuron-specific NOVA1 and NOVA2 proteins, which regulate splicing by binding to the UCAY motif[15]. Comparisons between wild-type and *Nova2*$^{-/-}$ or *Nova1*$^{-/-}$ knockout mice, using microarrays combined with genome-wide identification of clustered UCAY motifs and RT-PCR validation, were used to identify 49 NOVA2-dependent splicing events in the mouse neocortex. Functional annotation showed a strong enrichment for genes involved in modulating neuronal synaptic plasticity[86].

Sequences that lie within and that flank co-regulated exons can be computationally analysed for conserved sequence motifs, for enriched or depleted motifs of known splicing regulators, and for novel motifs that are enriched or depleted. For example, microarray analysis of 6,216 alternative-splicing events in 22 adult mouse tissues identified 171 cassette exons that exhibited brain-specific regulation (either preferred inclusion or skipping)[88]. Computational analysis of the flanking introns compared with a control set of constitutive exons found that NOVA and FOX binding sites were enriched downstream of brain-included but not brain-skipped exons. Furthermore, comparison of the 171 brain-regulated exons with the control set identified a novel motif in the upstream 150 nucleotides of the exons that were preferentially included in the brain.

snRNPs
Complexes of small nuclear RNAs (snRNAs) associated with ~20 proteins. The snRNA components of snRNPs form the catalytic core of the spliceosome.

SR proteins
A family of 11 highly conserved proteins that were originally identified as being required for constitutive and alternative splicing.

The splicing code and its utilization

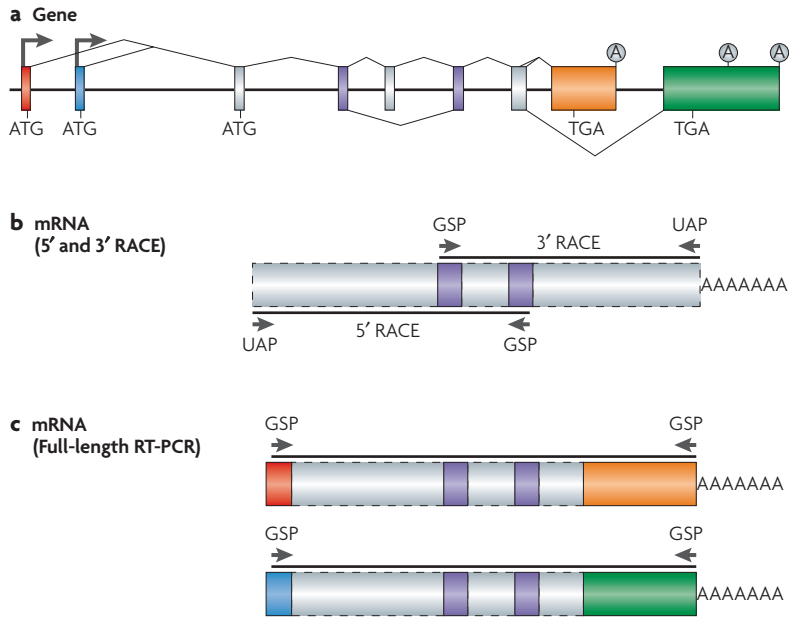
The first layer of the splicing code consists of consensus splice site sequences positioned at exon–intron boundaries that are essential for the splicing of all exons. These sequences are recognized primarily by base pairing with small nuclear RNAs (snRNAs, assembled as small nuclear ribonucleoprotein particles or snRNPs), which make up the core of the spliceosome (FIG. 1). It is this RNA–RNA base-pairing that specifies which nucleotides are involved in the precise cut-and-paste reactions that join exons[6]. A second layer of information is an extensive and complex array of diverse intronic and exonic splicing enhancer (ISE and ESE) and suppressor (ESS and ISS) elements (FIG. 1), which direct the cut-and-paste machinery to the appropriate sites and inhibit use of potential cryptic splice sites. The best characterized ESEs promote splicing by binding to the SR protein family, whereas the best characterized ESSs and ISSs repress splicing by binding to heterogeneous nuclear ribonucleoproteins (hnRNPs)[7]. Intriguingly, there are many exonic and intronic elements for which *trans*-acting mediators remain to be identified.

Whereas the consensus splice sites were relatively easy to identify from alignments of exon–intron boundary sequences, the remainder of the code remains to be fully deciphered. Most exons, and possibly all of them — whether constitutive or alternative — contain splicing elements, and combined computational and functional analyses have provided a basis for their identification[8–12]. However, these and other analyses continue to identify novel elements, particularly within exons, and have also demonstrated a bewildering interconnectedness of the splicing code. Systematic functional analyses have shown that exonic elements are composite and context-dependent[13]. In addition, the position of a splicing-factor binding site relative to the exon can determine whether it acts positively or negatively[14,15].

Alternative splicing utilizes the same consensus splice-site sequences and exonic and intronic elements as constitutive splicing. Investigations into individual splicing events and, more recently, combined microarray and computational analyses (BOX 1) have identified subsets of commonly regulated splicing events that share *cis*-acting elements that seem to be associated only with alternative splicing. These elements are bound by factors that are not generally associated with the spliceosome, such as FOX proteins (the vertebrate homologues of the worm FOX-1 protein), CUG-BP- and ETR-3-like (CELF) proteins, muscleblind-like (MBNL), the neuro-oncological ventral antigen (NOVA) proteins and TIA1 (REFS 2,16). Regulation of alternative splicing in vertebrates involves a dynamic interplay of antagonistic regulatory factors; for example, between the SR and hnRNP protein families[2,16], and between pairs of proteins including NOVA–polypyrimidine tract binding protein (PTB), CELF–PTB, CELF–MBNL, TIA1–PTB, and PTB–FOX[17–21]. As we will see below, disruption of this antagonism can have significant consequences resulting in disease.

A large fraction of the splice variants that are generated by alternative splicing remain to be identified. Furthermore, for genes that express multiple splice variants, the predominant splice variant(s) in any given tissue or cell type are typically unknown. In one study, nearly two-thirds of a set of 50 well-characterized genes were found to express novel splice variants in normal tissues and, for a striking 40% of these, the novel isoforms were predominant in normal tissues[22]. When the predominating isoform is unknown, a consequence is that functional analysis of disease-causing mutations can be carried out in the context of the splice variant(s) that are not relevant to the affected tissue, and the pathogenic effects of the mutation can be obscured by functional differences between isoforms[23]. Approaches to the identification of the predominant mRNA splice variants are discussed in BOX 2.

Box 2 | **Identifying the predominant splice variants**



In panel **a**, many vertebrate genes are highly complex, containing multiple transcription initiation sites (first exons), polyadenylation sites (A), and alternatively spliced internal and terminal exons. Splicing microarrays and standard RT-PCR do not identify the predominant full-length mRNAs, and amplification of full-length cDNA by RT-PCR requires knowledge of the 5′ and 3′ ends of the mRNAs in the tissue or cells of interest. Even the best analysis using EST, cDNA and gene-prediction approaches miss a significant fraction of alternative 5′- or 3′-terminal exons and internal alternative splicing. In the figure, constitutive exons are purple and alternative first and last exons are illustrated in different colours. In panel **b**, the predominant full-length splice variants in cells of interest can be identified using 5′ and 3′ rapid amplification of cDNA ends (RACE) to identify the predominant 5′ and 3′ termini. This is followed by RT-PCR amplification of full-length transcripts using primers that have been designed for those termini. 5′ and 3′ RACE is applied using gene-specific primers (GSPs) to constitutive exons and universal amplification primers (UAPs). Exons are defined as constitutive on the basis of available EST and cDNA data. The 5′ and 3′ RACE products are designed such that the priming site for the GSP 5′ oligonucleotide is included in the 3′ RACE product and vice versa, so that the cDNAs obtained can be used to confirm constitutive splicing of the exons used for the GSPs. In panel **c**, once the predominant 5′ and 3′ ends are identified, full-length cDNAs are amplified by RT-PCR using appropriate pairs of GSPs. The PCR products are then cloned and sequenced.

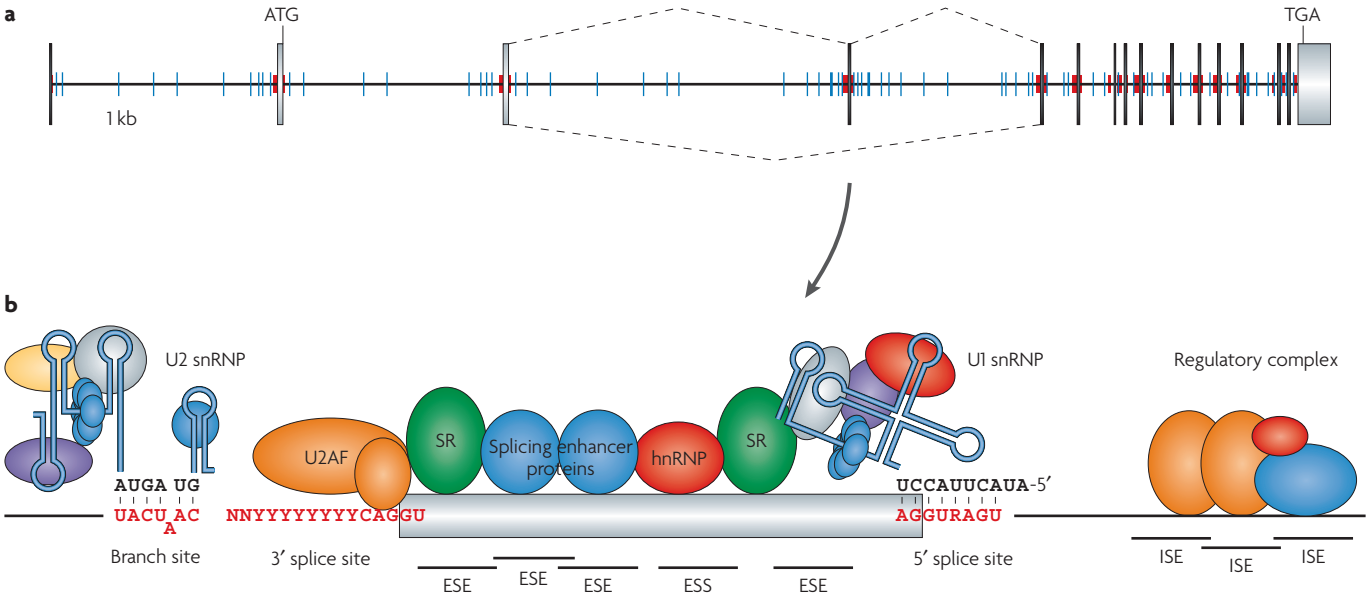


Figure 1 | **The splicing code. a** | A pre-mRNA as it might appear to the spliceosome. Red indicates consensus splice-site sequences at the intron–exon boundaries. Blue indicates additional intronic *cis*-acting elements that make up the splicing code. **b** | *cis*-elements within and around an alternative exon are required for its recognition and regulation. The 5′ splice site and branch site serve as binding sites for the RNA components of U1 and U2 small nuclear ribonucleoprotein (snRNPs), respectively. This RNA:RNA base pairing determines the precise joining of exons at the correct nucleotides. Mutations in the pre-mRNA that disrupt this base pairing decrease the efficiency of exon recognition. Exons and introns contain diverse sets of enhancer and suppressor elements that refine bone fide exon recognition. Some exon splicing enhancers (ESEs) bind SR proteins and recruit and stabilize binding of spliceosome components such as U2AF. Exon splicing suppressors (ESSs) bind protein components of heterogeneous nuclear ribonucleoproteins (hnRNP) to repress exon usage. Some intronic splicing enhancers (ISEs) bind auxiliary splicing factors that are not normally associated with the spliceosome to regulate alternative splicing.

Splicing as a primary cause of disease

Alterations in splicing can cause disease directly, modify the severity of the disease phenotype or be linked with disease susceptibility. In each case, the mechanisms causing altered splicing involve disruption of either *cis*-acting elements within the affected gene or *trans*-acting factors that are required for normal splicing or splicing regulation. The distinction between *cis*- and *trans*-acting effects has important mechanistic implications. Effects in *cis* have a direct impact on the expression of only one gene, whereas effects in *trans* have the potential to affect the expression of multiple genes. Examples and implications for *cis* and *trans* effects of each of the three mechanisms are outlined below.

***Cis*-acting splicing mutations: disruption of the splicing code.** What fraction of disease-causing mutations disrupt the *cis*-acting elements that are required for splicing? An often-cited estimate of 15% from 1992 preceded knowledge of the splicing code and reflects only mutations that are known to affect the splice sites[24]. It is now clear that a large but still unknown fraction of mutations affect splicing by disrupting other components of the splicing code, by creating cryptic splice sites or by altering secondary structure[25]. When assayed directly for individual genes, up to 50% of disease-causing mutations are found to affect splicing[7,26]. It has recently been proposed that 60% of mutations that cause disease do so by disrupting splicing[27]. Of the mutations in the Human Gene Mutation Database (HGMD)[28] that are not within splice sites, 78% are single-nucleotide substitutions within exons (56.9%), or microdeletions or microinsertions of up to six nucleotides that occur primarily within exons (21.7%). The pathogenicity of exonic mutations is generally assumed to result from the predicted effect on the reading frame and protein function. However, it is likely that the primary pathogenic effect for many of these exon mutations is at the level of splicing. An extensive and systematic mutation analysis of exons 9 and 12 of the *CFTR* gene, which is mutated in cystic fibrosis, showed that about one-quarter of even synonymous substitutions resulted in altered splicing[29]. Accurate numbers for mutations that affect splicing in *cis* remain to be determined for many genes and, because exons vary in the density of splicing elements required for efficient splicing, there is likely to be a wide range of sensitivities to nucleotide substitutions and microdeletions. Still, most *cis*-acting splicing mutations probably remain undetected.

The wide range for the predicted frequency of splicing mutations (15–60%) reflects our incomplete knowledge of the splicing code and the fact that mRNAs from mutant alleles are rarely assayed for splicing abnormalities. A long-term goal of deciphering the splicing code is to acquire the power to predict which disease-associated

hnRNPs
Heterogeneous nuclear ribonucleoproteins are an abundant class of predominantly nuclear RNA binding proteins that contain an RNA recognition motif (RRM) type of RNA binding domain.

nucleotide alterations are likely to affect splicing. In the meantime, systematic surveys of various disease-causing exonic mutations for aberrant splicing would be enlightening, both for the individual mutations that are analysed and for a broad-based analysis of the impact of splicing as a primary mechanism of disease. Screens could be carried out using high-throughput RT-PCR, or splicing microarrays could be designed to cover at least the 2,577 known disease genes in the HGMD. The main limitation will be obtaining RNA from the disease-relevant tissues.

Knowing that the primary effect of an exonic mutation is a splicing defect, rather than the predicted protein-coding mutation, is crucial for understanding the detailed pathogenic mechanism of a disease. For example, it is important to know that a mutation results in loss of expression due to aberrant splicing and NMD-mediated degradation, rather than by causing expression of wild-type levels of a protein containing a missense mutation. In addition, identification of abnormal splicing as the primary mechanism of disease raises the possibility of therapeutic approaches that target splicing (FIG. 2).

Trans-acting mutations: disruption of the splicing machinery. The structural and functional complexity of the spliceosome rivals that of the ribosome. The spliceosome is composed of five snRNPs and more than 150 proteins, including kinases, phosphatases and helicases, many of which are required for spliceosomal function, as well as associated proteins such as mRNA-export factors and transcription factors[30]. Two diseases, spinal muscular atrophy (SMA) and retinitis pigmentosa[31,32], are caused by mutations in genes involved in snRNP assembly and function, respectively. Strikingly, not only are both diseases neuron-specific, they affect different subsets of neurons. This cell specificity is surprising given the expectation that substantial loss of spliceosomal function is cell-lethal. SMA is an autosomal recessive disorder affecting motor neurons, and is caused by loss of the survivor of motor neuron-1 (*SMN1*) gene product, which is required for the assembly of core snRNPs in the cytoplasm before final maturation and nuclear import. Loss of snRNP production has been directly linked with disease[33]; however, the implication that disease results from altered splicing has yet to be demonstrated, and loss of other neuron-specific functions of *SMN1* could have a major role[31].

Retinitis pigmentosa is one of the most common forms of blindness, affecting 1 in 4,000 people worldwide. The disease results from retinal degeneration due primarily to progressive loss of photoreceptor cells. Most cases are sporadic; however, mutations in more than 30 genes cause familial forms of the disease, which can be autosomal recessive, autosomal dominant or X-linked. Surprisingly, three dominant retinitis pigmentosa genes (pre-mRNA-processing factor gene homologues *PRPF31*, *PRPF8* and *HPRP3*) encode proteins required for proper assembly and function of the U4•U5•U6 tri-snRNP, a core and essential component of the spliceosome[34–36]. As for SMA, there is no direct evidence that any of these three retinitis pigmentosa mutations affect splicing of endogenous pre-mRNAs in the affected neurons. However, disease is probably due to disruption of spliceosome function, because it seems unlikely that all three genes or the U4•U5•U6 tri-snRNP have alternative functions. Cell-type specificity is likely to result from sensitivity of one or more photoreceptor-specific pre-mRNAs to loss of U4•U5•U6 tri-snRNP function. Most of the other 12 loci that cause dominantly inherited retinitis pigmentosa are photoreceptor-cell-specific (see the RetNet web site), setting the precedent that decreased expression of individual gene products causes the disease.

Prader–Willi Syndrome (PWS) is the first known example of a genetic disease in which pathogenesis might be due to mutation of a gene encoding a splicing regulatory factor. PWS is a congenital disease defined by a constellation of symptoms including mental retardation, obesity and abnormal behavioural traits. It is caused by loss of the paternal allele of the maternally imprinted 1-Mb region of 15q11–13 (REF. 37). This region contains the paternally expressed *SNURF* (SNRPN nucleoprotein N upstream reading frame)–*SNRNP* (small nuclear ribonucleoprotein N) locus, which encodes 47 copies of the HBII-52 small nucleolar RNA (snoRNA) genes. SnoRNAs typically serve as guide RNAs that target modifications of ribosomal RNA (rRNAs), tRNAs or snRNAs; however, HBII-52 snoRNA was found to have a conserved 18-bp region that is complementary to the serotonin receptor 2C ($5\text{-}HC_{2C}R$) gene near an alternative 5′ splice site[38]. A recent study provides evidence that HBII-52 snoRNA regulates splicing of $5\text{-}HC_{2C}R$, and that loss of HBII-52 snoRNA expression in PWS results in aberrantly regulated splicing of $5\text{-}HC_{2C}R$[39]. It will be interesting to determine whether HBII-52 snoRNA regulates other splicing events, and whether alteration of splicing contributes to the PWS phenotype.

Trans-dominant effects on splicing regulation: RNA gain of function. Naturally occurring repeats of 3–10 nucleotides within transcribed genomic regions that expand beyond pathogenic thresholds cause microsatellite-expansion disorders[40,41]. Trinucleotide expansions within coding regions cause diseases such as Huntington disease, spinocerebellar ataxias and oculopharyngeal muscular dystrophy owing to a loss and/or toxic gain of function for the encoded proteins, which contain polyglutamine or polyalanine tracts[40]. A subset of microsatellite-expansion disorders are caused by large expansions within non-coding regions, and cause disease by a mechanism in which the RNA that is transcribed from the expanded alleles has a toxic gain of function[42]. The best characterized of these is myotonic dystrophy (also known as dystrophia myotonica), a dominantly inherited disorder that is the second most common cause of muscular dystrophy. Myotonic dystrophy is caused by a CTG trinucleotide expansion in the 3′ untranslated region (3′ UTR) of *DMPK*, or by a CCTG expansion in intron 1 of *ZNF9* (also known as CNBP).

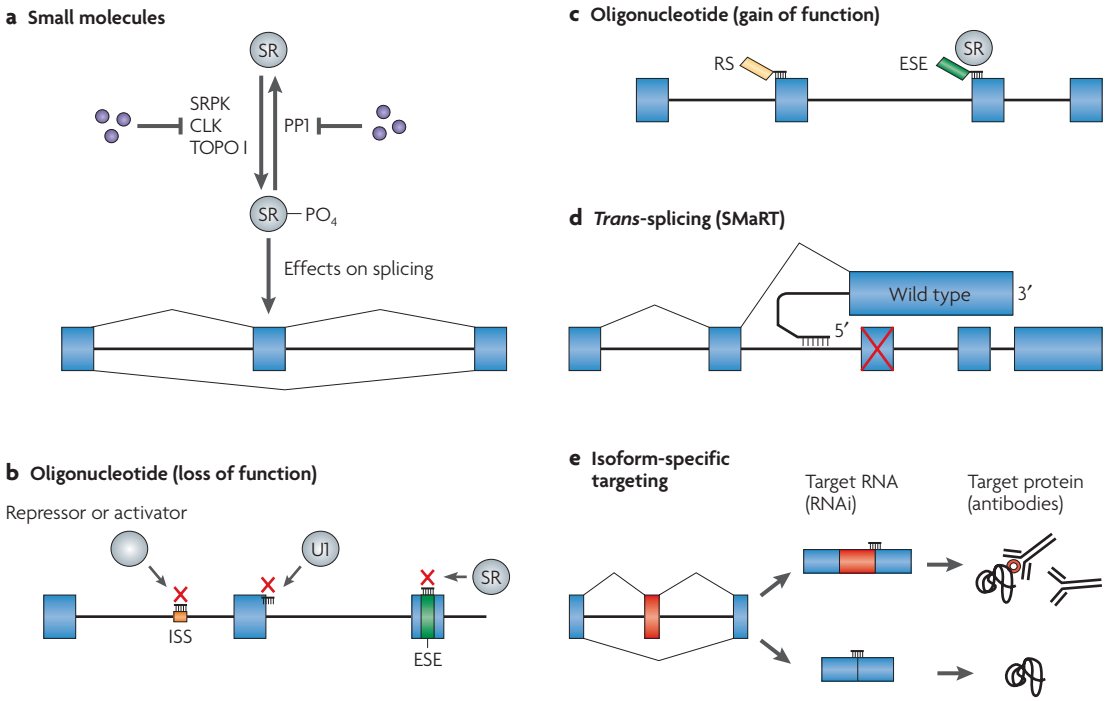


Figure 2 | **Therapeutic approaches that utilize splicing.** Therapeutic approaches can either alter the splicing patterns of target genes (panels **a**–**d**) or target specific splice variants at the RNA or protein level to achieve a therapeutic affect (panel **e**). **a** | The use of small-molecule drug therapy is a particularly attractive approach to modifying splicing patterns because of the relative ease of delivery and dosage control. Many chemical compounds have been found to affect the splicing of numerous genes[67,98]. Although the mechanisms by which splicing patterns are altered are unknown for most of these compounds, several of them affect splicing by altering the steady-state levels, phosphorylation state or intracellular distribution of SR proteins[98]. Molecules that alter the SR phosphorylation state are shown here as an example. SR proteins are phosphorylated by SR protein kinases (SRPKs) 1 and 2, DNA topoisomerase I (TOPO I), and a family of CDC2-like kinases (CLKs) and are dephosphorylated by protein phosphatase-1 (PP1)[98]. The phosphorylation status of SR proteins affects their RNA binding specificity, protein–protein interactions and intracellular distribution[99,100], so small molecules that affect the activities of these enzymes can be used to alter splicing patterns. The lack of specificity in this approach foreshadows a high potential for off-target effects; however, the recent development of splicing reporters with fluorogenic readouts allows for high-throughput screening for compounds with highly specific effects[101,102]. **b**,**c** | Antisense oligonucleotides that are modified to enhance stability and prevent RNase H-mediated degradation of the target RNA provide a gene-specific approach to altering splicing patterns using loss- or gain-of-function approaches[103]. In loss-of-function approaches (panel **b**), an oligonucleotide is used to block access of splicing factors to either the splice sites or other components of the splicing code, such as intronic and exonic splicing enhancer (ISE and ESE) and suppressor (ESS and ISS) elements. For example, a systematic analysis of overlapping antisense oligonucleotides that are complementary to the survival of motor neuron 2, centromeric (*SMN2*) gene exon identified oligonucleotides with dramatic positive or negative effects on exon inclusion, presumably by blocking ESE and ESS elements and/or secondary structures that are important for controlling the amount of exon inclusion[104]. Oligonucleotides have been used to circumvent the inclusion of mutated exons to restore production of a truncated but functional protein. In an interesting twist to this approach, oligonucleotides have been used for a gain of splicing function (panel **c**). Bifunctional oligonucleotides that contain targeting (complementary) and effector domains can be created to enhance splicing of the targeted exon. In one case, a chimeric peptide nucleic acid molecule was made that contained arginine-serine (RS) peptides to mimic the functional domain of SR proteins bound to an ESE[105]. A different approach utilized an oligonucleotide that contained an ESE to serve as a binding site for SR proteins[106]. **d** | *Trans*-splicing is used to reprogramme the mRNA that is expressed from mutated genes. The most common approach is to express an exogenous pre-mRNA that contains a replacement wild-type sequence that is flanked upstream by a strong 3′ splice site and a segment that is complementary to an intron located upstream of the mutation (indicated by a red cross) in the target pre-mRNA. The exogenous RNA is tethered to the endogenous pre-mRNA from the mutated allele, which positions the wild-type segment of the mRNA to be spliced in place of the mutant segment. This approach, dubbed SmaRT (spliceosome-mediated RNA *trans*-splicing) has been used successfully in cell culture and in mouse models of disease[103]. Non-spliceosomal *trans*-splicing approaches using group I ribozymes and a tRNA-splicing endonuclease from the archeon *Methanococcus jannaschii* have also been demonstrated in cell culture[103]. **e** | Alternative or aberrant splicing produces mRNAs that contain unique sequences and proteins with unique epitopes that can be targeted using RNAi-mediated degradation or antibodies, respectively. These features can be utilized in cases in which removing one isoform is of benefit. As with all therapies, splicing-based approaches have issues with off-target effects, toxicity, efficiency and delivery. The potential applications are diverse; however, most applications are gene-specific and are therefore limited to specific diseases so they must be optimized for individual genes.

Box 3 | **RNA gain-of-function disorders.**

An RNA gain-of-function mechanism is likely to be a component of pathogenesis in three microsatellite-expansion disorders in addition to myotonic dystrophy: fragile-X tremor ataxia syndrome (FXTAS), spinocerebellar ataxia 8 (SCA8), and Huntington disease-like 2 (HDL2). Fragile-X syndrome is caused by an expanded (>200) CGG repeat in the 5′ untranslated region (UTR) of the X-linked *FMR1* gene, resulting in transcriptional silencing. FXTAS is a late-onset neurological syndrome that includes ataxia, parkinsonism and decreased cognitive function that develops in fragile-X carriers with pre-mutation repeat sizes of 55 to <200 CGG units[89]. Several results support an RNA gain of function in FXTAS. FMR1 mRNA levels are increased in FXTAS CNS tissues as a result of increased transcription, whereas protein levels are moderately decreased or unchanged owing to reduced translation[90]. A knock-in of 98 CGG repeats into the 5′ UTR of the mouse *Fmr1* gene produced an age-dependent FXTAS-like phenotype as well as increased *Fmr1* mRNA, but did not change protein levels[91]. Expression of non-coding CGG repeats in flies resulted in neuronal degeneration[92]. Ubiquitin-positive nuclear inclusions in neurons and astrocytes of affected individuals contain expansions in the RNA of *FMR1* as well as muscleblind-like 1 (*MBNL1*), suggesting parallels with myotonic dystrophy pathogenesis[93].

Spinocerebellar ataxias (SCAs) are late-onset neurodegenerative disorders, including polyglutamine diseases, caused by coding CAG microsatellite expansions[40]. An expanded CTG repeat has been linked with SCA8, although with variable penetrance[94]. Recent analysis using a BAC transgenic mouse model has revealed that SCA8 pathogenesis involves expression of both a polyglutamine protein from one strand and an expanded CUG RNA from the other[95].

HDL2 phenotypically resembles Huntington disease and is caused by a CTG expansion within an alternative exon of the junctophilin-3 (*JPH3*) gene. Depending on the splicing pattern, the CUG tract occurs within an intron or 3′ UTR, or it encodes polyleucine or polyalanine tracts[96]. Recently, *in situ* hybridization combined with immunofluorescence staining demonstrated that *JPH3* mRNA containing expanded CUG repeats colocalizes with MBNL1 in nuclear foci in cortical neurons of HDL2 patients[97]. Although there is strong evidence for a role for RNA in the mechanism of pathogenesis of these three disorders, splicing abnormalities have not been identified.

In both forms of the disease, RNA from the expanded allele accumulates in nuclear foci. Evidence from transgenic mouse models has demonstrated that CUG-repeat-containing RNA is necessary and sufficient to cause disease[43–46].

A key molecular feature of myotonic dystrophy is the misregulation of developmentally regulated alternative splicing for a subset of genes, such that embryonic and neonatal splicing patterns are retained in adult myotonic dystrophy tissues (reviewed in (REFS 42,43,46)). Disease symptoms such as myotonia and insulin resistance result from the inappropriate expression of embryonic splicing patterns in adult tissues. The RNA disrupts normal postnatal alternative splicing transitions that are regulated by two families of proteins: the MBNL and CELF families. The best characterized members of these families are MBNL1 and CUGBP1, which were first identified as CUG-repeat RNA binding proteins. For the genes tested, these two proteins regulate the same splicing events antagonistically by binding to separate regulatory elements. The mechanism of normal postnatal transitions seems to involve a loss of nuclear CUGBP1 owing to decreased protein expression[47,48], and a gain of nuclear MBNL1 activity owing to translocation from the cytoplasm[47]. The activities of both proteins are disrupted in myotonic dystrophy by the toxic RNA. Nuclear MBNL1 is depleted as a result of its sequestration into RNA foci, whereas CUGBP1 steady-state levels increase owing to phosphorylation and increased protein half-life[49]. Mouse models support a primary role for loss of MBNL1 function as well as a gain of CUGBP1 function in the splicing abnormalities. However, disruption of a cytoplasmic function for CUGBP1 as a translation regulator has also been proposed[50]. The results suggest that expression of toxic RNA initiates a combination of mechanisms that result in myotonic dystrophy and other microsatellite-expansion disorders (BOX 3).

Splicing in disease modification and susceptibility

Given the density of the splicing code and the prevalence of alternative splicing, it is not surprising that normal genetic variation affects splicing efficiency and alters either the total output of a gene or the ratio of alternatively spliced variants. One large-scale analysis of human genetic variation found that 51 (1.3%) of the SNPs identified within and around the exons of 313 genes studied were within the consensus splice sites[51]. Analysis of SNP frequency, specifically within splice sites, found variability at all positions except the first and last two nucleotides of the intron, which are essential for splicing[52]. This variation within the non-essential nucleotides of the splice sites affects splicing efficiency; for example, haplotype-specific splicing patterns of the major histocompatibility complex, class II, DQβ 1 (*HLA-DQB1*) gene exon 4 were mapped to three SNPs (out of more than 150 polymorphisms within the locus) that were located within the branch site and polypyrimidine tract of intron 3 (REF. 53). As described in a previous section, one-quarter of synonymous substitutions introduced into an exon altered splicing efficiency[29], demonstrating that natural variation within coding sequence (cSNPs) can affect splicing efficiency.

Genetic variation also affects alternative splicing. To determine how much variability in alternative splicing exists in a population, the ratios of 6 alternative splicing events in 5 genes were compared among 147 age-matched, genetically diverse mice. Variation in alternative splicing was remarkably low within the population: all 6 splicing events were among the 15 least variable of 150 phenotypic measurements[54]. However, in a subpopulation of individuals, the ratio of splice variants differed significantly from the average, reflecting variation in either the *cis*-acting elements within the genes that was assayed or the *trans*-acting splicing environment between individuals. In another study, an analysis of 70 cassette-alternative exons in lymphoblast cell lines from 22 individuals identified 6 exons that exhibited variable skip-inclusion ratios between individuals that correlated with closely associated SNPs. Interestingly, most of the SNPs (four out of six) that were linked with altered ratios were within the alternative exons[55].

There is a growing realization that splicing efficiency is a significant contributor to phenotypic variability[56], and the contribution of splicing to phenotype

Table 1 | **Disease modification and susceptibility mediated through splicing effects**

Gene	Effect	Disease	Variant location	Effect on splicing / protein	Refs
Modifies disease phenotype					
CFTR	*cis*	Cystic fibrosis	(TG)n and Tn polymorphisms in *CFTR* intron 8	Affects the amount of exon 9 skipping	107,108
MCAD	*cis*	Medium-chain acyl-CoA dehydrogenase deficiency	ESS within exon 5	Prevents effect of disease-causing ESE mutation	60
SCN1A	*cis*	Susceptibility to anti-epileptics	5′ splice site of neonatal alternative exon	Increased use of neonatal alternative exon	68
CFTR	*cis* and *trans*	Cystic fibrosis	Point mutation in intron 19 creates a variably spliced 84-nucleotide exon	Variable level of cryptic exon inclusion influences severity	109
IKBKAP	*trans*	Familial dysautonomia	n/a	Tissue-specific differences in recognition of mutant 5′ splice site	62
Scn8a	*cis* and *trans*	Neurological disorder (mouse)	4-bp deletion within the 5′ splice site of exon 3	5′ splice-site mutation modified by *Scnm1*	64
Linked with disease susceptibility					
IRF5	*cis*	Systemic lupus erythematosus (SLE)	One SNP between alternative promoters creates 5′ splice site	SNP creates 5′ splice site and new first exon	110
CTLA4	*cis*	Autoimmune diseases	Two SNPs in 3′ UTR (exon 4)	Increased exon 3 skipping; reduced soluble isoform	111
NCAM1	*cis*	Bipolar disorder	Two SNPs, one within cluster of alternative exons	Decreased expression of secreted splice variants	112
ERBB4	*cis*	Schizophrenia	One SNP in intron 12 and SNPs near exon 3 linked with splicing of exons 16 and 26, respectively	Increased use of exons 16 and 26	113
OLR1	*cis*	Myocardial infarction	Six SNPs; three in intron 4, two in intron 5, one in the 3′ UTR in exon 6	Exon 5 skipping results in an isoform with reduced apoptotic effects	114
OAS1	*cis*	Type I diabetes	Intron 6 AG→AA variant shifts 3′ splice site by 1 nucleotide, changing the reading frame	SNP moves splice site by 1 nucleotide resulting in a longer protein	115
TNNT2	*cis*	Cardiac hypertrophy	5-bp deletion affects intron 3 splice site	Results in E4 skipping (minigene analysis)	116
GPRA	*cis*	Asthma	Three SNPs distal to alternative site	Increased use of the more distal of two terminal exons	117
MAPT	*cis*	Tauopathies	238-bp insertion into intron 9	Enhanced exon 10 inclusion	118
PTPRC (CD45)	*cis*	Altered immune function	A138G polymorphism exon 6	Enhanced exon 6 skipping	119
PTPRC (CD45)	*cis*	Multiple sclerosis	C77G polymorphism exon 4	Enhanced exon 4 inclusion	120,121
LDLR	*cis*	Elevated cholesterol	C688T polymorphism exon 12	Enhanced exon 12 skipping	122
SFRS8	*trans*	Asthma	n/a	None reported	66

Variation either within the relevant gene (*cis*) and in the splicing environment (*trans*) can affect splicing of the modifying or susceptibility gene to affect disease severity or the likelihood of developing disease. In the case of *SFRS8*, disease susceptibility is linked to a splicing regulator. *CFTR*, cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7); *CTLA4*, cytotoxic T-lymphocyte-associated protein 4; *ERBB4*, v-erb-a erythroblastic leukaemia viral oncogene homologue 4; ESS, exonic splicing suppressor; *GPRA*, G-protein-coupled receptor for asthma susceptibility; *IKBKAP*, inhibitor of κ-light polypeptide gene enhancer in B cells, kinase complex-associated protein; *IRF5*, interferon regulatory factor 5; ISS, intronic splicing suppressor; *LDLR*, low-density lipoprotein receptor (familial hypercholesterolaemia); *MAPT*, microtubule-associated protein tau; *NCAM1*, neural cell-adhesion molecule 1; *OLR1*, oxidized low-density lipoprotein (lectin-like) receptor 1; *OAS1*, 2′,5′-oligoadenylate synthetase 1, 40–46kDa; *PTPRC*, protein tyrosine phosphatase, receptor type, C; *SCN1A*, sodium channel, voltage-gated, type I, α-subunit; *SCN8A*, sodium channel, voltage-gated, type VIII, α-subunit; *TNNT2*, troponin T type 2 (cardiac); *SFRS8*, splicing factor, arginine/serine-rich 8 (suppressor-of-white-apricot homologue, *Drosophila*).

has become particularly apparent through its impact on modifying the severity of human disease[57] and its contribution to disease susceptibility (TABLE 1).

Cis-acting genetic modifiers. Naturally occurring variants that affect the splicing efficiency and subsequent level of expression from a mutant allele can modify disease phenotype. A well-characterized example is the variable phenotype that is found among individuals with atypical cystic fibrosis. Approximately 50% of individuals with cystic fibrosis in the United States are homozygous for the ΔF508 mutation of the *CFTR* gene, which results in a severe loss of function of the encoded protein and mortality in young adulthood. Many diverse, less severe *CFTR* loss-of-function mutations produce atypical cystic fibrosis, a non-lethal disease

affecting the lungs, testes and pancreas. The penetrance and severity of these diseases are highly variable, even within families[58]. One determinant of this variability is a double polymorphism that affects UG and U repeats located in the 3′ splice site of *CFTR* exon 9, which can increase the (normally low) level of exon 9 skipping that is observed in most individuals. The increased exon skipping associated with these polymorphisms has no phenotypic consequences in unaffected individuals with wild-type *CFTR* alleles. However, the level of exon 9 skipping becomes determinative for the phenotype when *CFTR* function is already compromised by an atypical cystic fibrosis mutation located elsewhere in the gene. Additional variants within the coding regions of exons 9 and 12 that affect splicing can also produce variability in the atypical cystic fibrosis phenotype[59].

In another example, a missense mutation that inactivates an ESE in exon 5 of the medium-chain acyl-CoA dehydrogenase gene (*MCAD*) causes exon skipping and MCAD deficiency, resulting in a disease that affects mitochondrial function. Interestingly, a silent A→C polymorphism in the same exon located 11 nucleotides upstream from the ESE mutation modulates the severity of exon skipping. Analysis using minigene reporters indicated that the C variant prevents exon skipping by inactivating an adjacent ESS[60]. This is a particularly revealing example of a *cis*-acting modifier that illustrates the complexity and commonality of the splicing code and the intricate relationships between disease-causing splicing mutations and disease-modifying natural variants that affect splicing.

Trans-acting genetic modifiers. The same splicing mutation in different *trans*-acting splicing environments can exhibit different splicing efficiencies and modify the severity of the phenotype. In familial dysautonomia, a disease affecting sensory and autonomic neurons, the mutation that is responsible for >99% of cases is a T→C mutation in the sixth position of intron 20 (the 5′ splice site of exon 20) of *IKBKAP*, the gene encoding the inhibitor of κ-light polypeptide gene enhancer in B-cells, kinase complex-associated protein. This mutation results in exon skipping and loss of IKBKAP mRNA, presumably owing to NMD[61]. Exon inclusion is reduced, but not eliminated, and neuron-specific differences in splice-site recognition results in reduced levels of wild-type mRNA in neuronal tissues correlating with the neuron-specific defects[62]. Splicing variability between affected individuals could explain the variability of disease severity[63] and would support therapeutic approaches to promote exon inclusion[61].

One of the best characterized pairs of mutant-modifier genes in mammals is the effect of the sodium channel modifier 1 gene (*Scnm1*) on the phenotype of the *med*^{J} mutation of the sodium channel gene *Scn8a* in mice[64]. The *med*^{J} mutation is a 4-nucleotide deletion within the 5′ splice site of *Scn8a* exon 3, resulting in production of only 10% of the normal mRNA levels and causing a neurological phenotype. The *med*^{J} phenotype is significantly milder in C3H compared with C57BL/6J strains, in that the former exhibit progressive ataxia and live for more than 18 months, whereas the latter become paralysed and die within 1 month. *Scnm1* was identified as a modifier by positional cloning. In C57BL/6J mice, *Scnm1* contains a nucleotide substitution that inactivates an ESE (and introduces a termination codon), inducing exon skipping and expression of multiple aberrant transcripts. The *med*^{J} allele in these mice produces only 5% of the normal wild-type SCN8A mRNA levels owing to reduced splicing of the mutated 5′ splice site. Transgenic expression of the *Scnm1* cDNA in C57 mice rescued the neurological phenotype, as well as *Scn8a* exon 3 splicing. Direct evidence for SCNM1 as a splicing regulator remains to be obtained; however, the protein localizes to the nucleus and contains a C2H2 zinc-finger RNA binding domain, which is consistent with a splicing function[64]. Of particular interest was the finding that the slight reduction in correctly spliced *Scn8a* mRNA from 10% to 5% has such a significant impact on phenotype.

Splicing and disease susceptibility. With the expanding resources that are available to analyse associations between genotypic variability and disease susceptibility[65], a growing number of reports are identifying disease-susceptibility alleles that are associated with specific splicing patterns (TABLE 1). This third classification differs from the *cis*- and *trans*-acting disease modifiers described above in that there is no existing disease phenotype that is known to be modified by splicing differences. Rather, susceptibility to a disease has been mapped to a locus and, in some cases, to particular SNPs that are associated with a specific splicing pattern. In one case, the susceptibility gene itself encodes a splicing factor (SFRS8)[66] (TABLE 1). Although it remains to be demonstrated directly in these cases that splicing of the linked allele is determinative, the functional consequences of the altered splicing patterns are often consistent with increased disease susceptibility.

Genetic variation also affects splicing-dependent quantitative and qualitative expression of genes that control drug efficacy and toxicity. There are several examples in which differences in splicing have an impact on drug metabolism, toxic threshold or therapeutic levels (reviewed in REF. 67). In one recent example, a requirement for higher dosing of antiepileptic drugs was found to be associated with a common G→A polymorphism in the neuronal sodium channel α-subunit gene (*SCN1A*). The polymorphism is within the 5′ splice site of an alternative exon that is used predominantly in the neonatal brain, and the A variant, which is associated with higher dosing requirements, correlates with increased usage of the neonatal splice variants in adult brains as compared with the G variant[68].

A full understanding of the splicing code will allow computational predictions of the effects of specific SNPs, and particularly cSNPs linked with disease, disease severity or susceptibility. Application of splicing microarrays for analysis of association between splicing variation (alternative or constitutive) and genetic variation will provide a burst of information that is relevant to normal phenotypic variation as well as disease.

Splicing and cancer

Splicing abnormalities are a common characteristic of cancer. The potential roles for splicing in cancer are well documented and include changes in genes associated with cell migration, regulation of cell growth, hormone responsiveness, apoptosis and response to chemotherapy[69–71]. There are three broad issues with respect to the role of splicing in cancer. First, there is the question of the cause–effect relationship between splicing changes and cancer initiation or progression: does a change in the splicing of one or more gene(s) contribute to any of the multiple steps of transformation, neoplasia or metastasis? Second is the potential utility of alternative splicing signatures for diagnostic and prognostic purposes. Third is the potential to use cancer-specific splicing as an Achilles heel to target therapy specifically to cancer cells with little effect on normal tissues. The first two issues are discussed here; the third is addressed in the context of splicing-based therapies in FIG. 2.

How do splicing changes contribute to cancer development and progression? There is clear evidence that *cis*-acting mutations that affect the splicing of oncogenes, tumour suppressors and other cancer-relevant genes can have causal roles in cancer initiation and progression (reviewed in REF. 71). However, most cancer-associated splicing changes are not associated with nucleotide changes in the affected genes, implying that there is an alteration in the *trans*-acting splicing environment.

Evidence from several studies has established that changes in SR protein expression can have causative roles in neoplasia. Upregulation or changes in SR protein phosphorylation have been demonstrated in various cancers (see REFS 72,73 and the references within). Moreover, the SR protein gene *SFRS6* (also known as *SRP55*) was one of 189 genes found to be mutated in breast and colon cancers in a sequence-based screen of >13,000 genes[74]. A correlation between SR protein expression and cancer progression was first established using a mouse model of mammary gland tumour progression[75], and it has recently been demonstrated that the SR protein SF2/ASF is a proto-oncogene[72]. Expression of SF2/ASF was found to be upregulated in 20–50% of cancers screened, depending on cancer type. Moderate overexpression of exogenous SF2/ASF, comparable to that in cancer samples, transformed immortalized rodent cells and resulted in tumour formation in nude mice. SR proteins control several steps of gene expression in addition to splicing[76], and SF2/ASF is also required for genomic stability[77]. However, oncogenic activity of SF2/ASF was directly linked with its splicing activity; in particular, an oncogenic isoform of ribosomal protein S6 kinase-β1 (RPS6KB1, also known as S6K1), which is induced by SF2/ASF, strongly correlated with the oncogenic activity of this SR protein. Further evidence was provided by the reversal of the transformed phenotype, in both NIH3T3 fibroblast cells that overexpressed SF2/AS and in a human tumour lung cancer cell line, by knocking down SF2/ASF or the oncogenic S6K1 isoform[72].

In a separate study, SF2/ASF was shown to regulate alternative splicing of the tyrosine kinase receptor proto-oncogene, *recepteur d'origine nantais* (*RON*), to produce a constitutively active form (ΔRON)[73]. ΔRON expression is elevated in two-thirds of breast cancers and the activated receptor induces increased migration and invasiveness, properties that are characteristic of metastatic progression. The use of both overexpression and RNAi-mediated knockdown established a strong connection between SF2/ASF expression, production of the ΔRON isoform and cell behaviour that is consistent with metastatic potential. Both of these studies show that SF2/ASF is likely to be a causative factor in different cancers and an important potential therapeutic target.

Splicing signatures as diagnostic and prognostic indicators. Standard microarrays provided a robust signature for tumour identification[78]. The combination of mRNA steady-state levels and ratios of specific splice variants obtained from splicing microarrays has the potential for extremely high-resolution signatures to better resolve tumour subclasses and identify optimal therapies. This utility does not require the cause–effect relationships between splicing and cancer to be identified, or even that they exist, only that associations between splicing patterns and subclassifications are sufficiently consistent to be predictive. Although large-scale studies have yet to be carried out, initial results are promising. Surveys of small numbers of genes found that signature splicing events are linked with particular cancer types[79,80]. A larger study comparing standard microarrays with splicing microarrays using 1,532 isoforms from 364 prostate-cancer-relevant genes found that, for 30% of the genes included on the array, differential expression of splice variants was detected in the absence of detectable changes in overall expression. Although the inclusion of splicing data gave a modest increase in the proportion of correctly predicted cases, from 87% to 95% (REF. 81), the high success rate is a testament to the robustness of microarray assays. These results bode well for the future use of arrays with larger gene sets and in combination with additional information on regulatory modules for application to less well-characterized tumours.

Conclusions and outlook

Investigations into the roles of splicing in disease have taught us a great deal about the normal mechanisms of splicing, and its regulation. First, the early finding that most disease-causing 5′ splice-site mutations resulted in exon skipping rather than intron retention supported the concept of exon definition, in which exons are the primary units of recognition[82]. Second, the effects of disease-causing exon mutations on splicing helped reveal the existence of the splicing code. Third, the role of splicing as a disease modifier and susceptibility factor has highlighted its broader contributions to phenotypic variability. Fourth, myotonic dystrophy provides an example of how a network of developmentally regulated alternative splicing events

and the relevant regulators were identified through purification of proteins that bound to the pathogenic CUG-repeat RNA. Similarly, in the case of PWS, finding HBII-52 snoRNA genes within the deleted region sparked investigation into a regulatory role for these snoRNAs in alternative splicing.

The extent to which splicing has a role in disease, as either a direct cause, a modifier or a susceptibility factor, continues to be defined. Advances in several areas will clarify the roles of splicing in disease and reveal the mechanisms involved, and will allow routine application of the knowledge gained toward diagnosis and treatment. One major advance will be to develop the ability to predict splicing outcomes associated with genetic variants and disease-causing mutations. Success will be enhanced by computational analyses that combine a defined splicing code with additional features. For example, disease-causing amino-acid substitutions commonly have dramatic effects on protein structure[83]. Mutations could be weighted as to whether effects on splicing or protein structure are more likely. Other features such as those defined by the ENCODE project[84] could also be included to develop a probability score to rank the most likely mechanisms by which mutations cause disease and normal variants affect phenotype. Knowledge of the splicing code can also be applied towards the identification of which of multiple SNPs that are linked with a phenotype could be causative of that phenotype as a result of a splicing effect.

Another important advance would be for assays for the effects of the mutation on splicing to be included in the routine analysis of disease-causing mutations. The best approach would be RT-PCR analysis of the mRNA region that includes the affected exon or intron position and flanking exons, as this does not require *a priori* knowledge of the aberrant splicing pathway(s) that are induced by the mutation.

Important insights will be gained from the full characterization of the human transcriptome, which would provide a catalogue of all the splice variants expressed from each gene and identification of the isoforms that predominate in specific cell types and tissues. This is a significant challenge that will have a huge benefit, not the least of which is the ability to design microarrays that can be applied for the quantitative assessment of all splice variants.

Finally, another key direction for the future will be the application of genome-wide microarray assays to assess splicing differences associated with normal variation and with disease. Alternative-splicing signatures are likely to provide a useful diagnostic and prognostic tool for many diseases. As for the other challenges outlined above, the tools that are required to meet this challenge are developing rapidly.

1. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).
One of the early splicing microarray papers, in which exon–exon junctions from 10,000 human genes were probed using RNA from 52 different human tissues.
2. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **27**, 27–48 (2003).
3. Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nature Genet.* **30**, 13–19 (2002).
4. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100**, 189–192 (2003).
A quantitative computational analysis demonstrating that one-third of alternatively spliced transcripts contain a premature termination codon and are potentially subject to NMD.
5. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37–47 (2006).
An outstanding review of the application of splicing microarrays to obtain a global perspective on the regulation of alternative splicing.
6. Burge, C. B., Tuschl, T. & Sharp, P. A. Splicing of precursors to mRNAs by the spliceosomes. in *The RNA World* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 525–560 (Cold Spring Harbor Laboratory Press, New York, 1999).
7. Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Rev. Genet.* **3**, 285–298 (2002).
8. Fairbrother, W. G., Yeh, R. F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
An early use of computational analysis (RESCUE-ESE) to identify sequence motifs within exons. Functional analysis using minigene reporters demonstrated that most of the computationally identified sequences enhanced exon inclusion.
9. Zhang, X. H. & Chasin, L. A. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* **18**, 1241–1250 (2004).
10. Fairbrother, W. G. *et al.* RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res.* **32**, W187–W190 (2004).
11. Goren, A. *et al.* Comparative analysis identifies exonic splicing regulatory sequences — the complex definition of enhancers and silencers. *Mol. Cell* **22**, 769–781 (2006).
12. Smith, P. J. *et al.* An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum. Mol. Genet.* **15**, 2490–2508 (2006).
ESEfinder is a commonly used program to identify binding sites for specific SR proteins within exons.
13. Pagani, F. *et al.* New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in *CFTR* exon 12. *Hum. Mol. Genet.* **12**, 1111–1120 (2003).
14. Kanopka, A., Muhlemann, O. & Akusjarvi, G. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**, 535–538 (1996).
15. Ule, J. *et al.* An RNA map predicting NOVA-dependent splicing regulation. *Nature* **444**, 580–586 (2006).
Combined computational and biochemical analyses demonstrated that the position of the NOVA binding site relative to the exon is predictive of whether NOVA enhances or represses the alternative splicing event.
16. Matlin, A. J., Clark, F. & Smith, C. W. Understanding alternative splicing: towards a cellular code. *Nature Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
17. Polydorides, A. D., Okano, H. J., Yang, Y. Y., Stefani, G. & Darnell, R. B. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of NOVA to regulate neuron-specific alternative splicing. *Proc. Natl Acad. Sci. USA* **97**, 6350–6355 (2000).
18. Ho, T. H. *et al.* Muscleblind proteins regulate alternative splicing. *EMBO J.* **23**, 3103–3112 (2004).
19. Charlet-B., N., Logan, P., Singh, G. & Cooper, T. A. Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol. Cell* **9**, 649–658 (2002).
20. Jin, Y. *et al.* A vertebrate RNA-binding protein FOX-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.* **22**, 905–912 (2003).
21. Izquierdo, J. M. *et al.* Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol. Cell* **19**, 475–484 (2005).
22. Roy, M., Xu, Q. & Lee, C. Evidence that public database records for many cancer-associated genes reflect a splice form found in tumors and lack normal splice forms. *Nucleic Acids Res.* **33**, 5026–5033 (2005).
This work demonstrated that the normal isoforms of a panel of 50 cancer-related genes were underrepresented in GenBank because 70% of GenBank entries for these genes are derived from tumour samples. The authors found that the predominating isoforms in normal tissues for many of these genes were novel.
23. Zhong, X., Liu, J. R., Kyle, J. W., Hanck, D. A. & Agnew, W. S. A profile of alternative RNA splicing and transcript variation of *CACNA1H*, a human T-channel gene candidate for idiopathic generalized epilepsies. *Hum. Mol. Genet.* **15**, 1497–1512 (2006).
***Tour de force* identification and functional characterization of the large number of protein isoforms that are expressed from the human T-type calcium channel gene. An extensive range of biophysical properties among the isoforms is contrasted with the fact that only a few have been used for functional analysis of mutants associated with epilepsy.**
24. Krawczak, M., Reiss, J. & Cooper, D. N. The mutational spectrum of single base-pair substitutions in messenger RNA splice junctions of human genes — causes and consequences. *Hum. Genet.* **90**, 41–54 (1992).
25. Pagani, F. & Baralle, F. E. Genomic variants in exons and introns: identifying the splicing spoilers. *Nature Rev. Genet.* **5**, 389–396 (2004).
26. Pagenstecher, C. *et al.* Aberrant splicing in *MLH1* and *MSH2* due to exonic and intronic variants. *Hum. Genet.* **119**, 9–22 (2006).
27. Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G. & Guigo, R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* **579**, 1900–1903 (2005).
28. Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* **21**, 577–581 (2003).
29. Pagani, F., Raponi, M. & Baralle, F. E. Synonymous mutations in *CFTR* exon 12 affect splicing and are not neutral in evolution. *Proc. Natl Acad. Sci. USA* **102**, 6368–6372 (2005).
30. Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**, 5–14 (2003).
31. Briese, M., Esmaeili, B. & Sattelle, D. B. Is spinal muscular atrophy the result of defects in motor neuron processes? *Bioessays* **27**, 946–957 (2005).

32. Mordes, D. *et al.* Pre-mRNA splicing and retinitis pigmentosa. *Mol. Vis.* **12**, 1259–1271 (2006).
33. Winkler, C. *et al.* Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. *Genes Dev.* **19**, 2320–2330 (2005).
34. Chakarova, C. F. *et al.* Mutations in *HPRP3*, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* **11**, 87–92 (2002).
35. Vithana, E. N. *et al.* A human homolog of yeast pre-mRNA splicing gene, *PRP31*, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell* **8**, 375–381 (2001).
36. McKie, A. B. *et al.* Mutations in the pre-mRNA splicing factor gene *PRPC8* in autosomal dominant retinitis pigmentosa (RP13). *Hum. Mol. Genet.***10**, 1555–1562 (2001).
37. Goldstone, A. P. Prader–Willi syndrome: advances in genetics, pathophysiology and treatment. *Trends Endocrinol. Metab.* **15**, 12–20 (2004).
38. Cavaille, J. *et al.* Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA* **97**, 14311–14316 (2000).
39. Kishore, S. & Stamm, S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* **311**, 230–232 (2006).
This work demonstrated that a snoRNA regulates splicing of the serotonin receptor by binding to the pre-mRNA. Loss of this RNA correlates with altered splicing in Prader–Willi syndrome.
40. Gatchel, J. R. & Zoghbi, H. Y. Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Rev. Genet.* **6**, 743–755 (2005).
41. Lalioti, M. D. *et al.* Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* **386**, 847–851 (1997).
42. Ranum, L. P. & Cooper, T. A. RNA-mediated neuromuscular disorders. *Annu. Rev. Neurosci.* **29**, 259–277 (2006).
43. Cho, D. H. & Tapscott, S. J. Myotonic dystrophy: emerging mechanisms for *DM1* and *DM2*. *Biochim. Biophys. Acta* **1772**, 195–204 (2007).
44. Wang, G. S., Kearney, D. L., De Biasi, M., Taffet, G. & Cooper, T. A. Elevated CUG-BP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. *J. Clin. Invest.* (in the press).
45. Mahadevan, M. S. *et al.* Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. *Nature Genet.* **38**, 1066–1070 (2006).
46. Osborne, R. J. & Thornton, C. A. RNA-dominant diseases. *Hum. Mol. Genet.* **15**, R162–R169 (2006).
47. Lin, X. *et al.* Failure of MBNL1-dependent postnatal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.* **15**, 2087–2097 (2006).
48. Ladd, A. N., Stenberg, M. G., Swanson, M. S. & Cooper, T. A. Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. *Dev. Dyn.* **233**, 783–793 (2005).
49. Kuyumcu-Martinez, M. N., Wang, G. S. & Cooper, T. A. Increased steady state levels of CUG-BP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol. Cell* (in the press).
50. Timchenko, N. A. *et al.* RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J. Biol. Chem.* **276**, 7820–7826 (2001).
51. Stephens, J. C. *et al.* Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**, 489–493 (2001).
52. Krawczak, M. *et al.* Single base-pair substitutions in exon–intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.* **28**, 150–158 (2007).
53. Kralovicova, J., Houngninou-Molango, S., Kramer, A. & Vorechovsky, I. Branch site haplotypes that control alternative splicing. *Hum. Mol. Genet.* **13**, 3189–3202 (2004).
54. Chisa, J. L. & Burke, D. T. Mammalian mRNA splice-isoform selection is tightly controlled. *Genetics* **175**, 1079–1087 (2007).
55. Hull, J. *et al.* Identification of common genetic variation that modulates alternative splicing. *PLoS Genet.* 3, e99 (2007).
56. Marden, J. H. Quantitative and evolutionary biology of alternative splicing: how changing the mix of alternative transcripts affects phenotypic plasticity and reaction norms. *Heredity* 27 September 2006 (doi: 10.1038/sj.hdy.6800904).
57. Nissim-Rafinia, M. & Kerem, B. Splicing regulation as a potential genetic modifier. *Trends Genet.* **18**, 123–127 (2002).
An early analysis of the role of splicing as a genetic modifier of disease.
58. Noone, P. G. & Knowles, M. R. 'CFTR-opathies': disease phenotypes associated with cystic fibrosis transmembrane regulator gene mutations. *Respir. Res.* **2**, 328–332 (2001).
59. Steiner, B., Truninger, K., Sanz, J., Schaller, A. & Gallati, S. The role of common single-nucleotide polymorphisms on exon 9 and exon 12 skipping in nonmutated *CFTR* alleles. *Hum. Mutat.* **24**, 120–129 (2004).
60. Nielsen, K. B. *et al.* Seemingly neutral polymorphic variants may confer immunity to splicing-inactivating mutations: a synonymous SNP in exon 5 of *MCAD* protects from deleterious mutations in a flanking exonic splicing enhancer. *Am. J. Hum. Genet.* **80**, 416–432 (2007).
61. Hims, M. M. *et al.* Therapeutic potential and mechanism of kinetin as a treatment for the human splicing disease familial dysautonomia. *J. Mol. Med.* **85**, 149–161 (2007).
62. Cuajungco, M. P. *et al.* Tissue-specific reduction in splicing efficiency of IKBKAP due to the major mutation associated with familial dysautonomia. *Am. J. Hum. Genet.* **72**, 749–758 (2003).
63. Axelrod, F. B. Familial dysautonomia: a review of the current pharmacological treatments. *Expert Opin. Pharmacother.* **6**, 561–567 (2005).
64. Buchner, D. A., Trudeau, M. & Meisler, M. H. SCNM1, a putative RNA splicing factor that modifies disease severity in mice. *Science* **301**, 967–969 (2003).
65. Farrall, M. & Morris, A. P. Gearing up for genome-wide gene-association studies. *Hum. Mol. Genet.* **14**, R157–R162 (2005).
66. Brasch-Andersen, C. *et al.* Significant linkage to chromosome 12q24.32–q24.33 and identification of *SFRS8* as a possible asthma susceptibility gene. *Thorax* **61**, 874–879 (2006).
67. Bracco, L. & Kearsey, J. The relevance of alternative RNA splicing to pharmacogenomics. *Trends Biotechnol.* **21**, 346–353 (2003).
68. Heinzen, E. L. *et al. Nova2* Interacts with a *cis*-acting polymorphism to influence the proportions of drug-responsive splice variants of *SCN1A*. *Am. J. Hum. Genet.* **80**, 876–883 (2007).
69. Skotheim, R. I. & Nees, M. Alternative splicing in cancer: noise, functional, or systematic? *Int. J. Biochem. Cell Biol.* **39**, 1432–1449 (2007).
70. Venables, J. P. Unbalanced alternative splicing and its significance in cancer. *Bioessays* **28**, 378–386 (2006).
71. Srebrow, A. & Kornblihtt, A. R. The connection between splicing and cancer. *J. Cell Sci.* **119**, 2635–2641 (2006).
72. Karni, R. *et al.* The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature Struct. Mol. Biol.* **14**, 185–193 (2007).
This work demonstrated that the splicing factor SF2/ASF acts as a proto-oncogene.
73. Ghigna, C. *et al.* Cell motility is controlled by SF2/ASF through alternative splicing of the *RON* protooncogene. *Mol. Cell* **20**, 881–890 (2005).
This paper identified the RON tyrosine kinase receptor as a target of SF2/ASF and presents a clear demonstration that increased SF2/ASF expression promotes an epithelial-to-mesenchymal transition through altered RON splicing.
74. Sjoblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
75. Stickeler, E., Kittrell, F., Medina, D. & Berget, S. M. Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. *Oncogene* **18**, 3574–3582 (1999).
76. Huang, Y. & Steitz, J. A. SRprises along a messenger's journey. *Mol. Cell* **17**, 613–615 (2005).
77. Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378 (2005).
78. Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. From signatures to models: understanding cancer using microarrays. *Nature Genet.* **37**, S38–S45 (2005).
79. Gardina, P. J. *et al.* Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **7**, 325 (2006).
80. Relogio, A. *et al.* Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *J. Biol. Chem.* **280**, 4779–4784 (2005).
81. Zhang, C. *et al.* Profiling alternatively spliced mRNA isoforms for prostate cancer classification. *BMC Bioinformatics* **7**, 202 (2006).
82. Berget, S. M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414 (1995).
83. Steward, R. E., MacArthur, M. W., Laskowski, R. A. & Thornton, J. M. Molecular basis of inherited diseases: a structural perspective. *Trends Genet.* **19**, 505–513 (2003).
84. Consortium, E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
85. Pan, Q. *et al.* Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* **16**, 929–941 (2004).
86. Ule, J. *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nature Genet.* **37**, 844–852 (2005).
Splicing microarray analyses of alternative splicing differences in NOVA knockout mice identified enrichment for genes involved in synaptic plasticity.
87. Ip, J. Y. *et al.* Global analysis of alternative splicing during T-cell activation. *RNA* **13**, 563–572 (2007).
88. Sugnet, C. W. *et al.* Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* **2**, e4 (2006).
89. Hagerman, P. J. & Hagerman, R. J. The fragile-X premutation: a maturing perspective. *Am. J. Hum. Genet.* **74**, 805–816 (2004).
90. Tassone, F. *et al.* Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA* **13**, 555–562 (2007).
91. Willemsen, R. *et al.* The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome. *Hum. Mol. Genet.* **12**, 949–959 (2003).
92. Jin, P. *et al.* RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in *Drosophila*. *Neuron* **39**, 739–747 (2003).
93. Iwahashi, C. K. *et al.* Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**, 256–271 (2006).
94. Ikeda, Y. *et al.* Spinocerebellar ataxia type 8: molecular genetic comparisons and haplotype analysis of 37 families with ataxia. *Am. J. Hum. Genet.* **75**, 3–16 (2004).
95. Moseley, M. L. *et al.* Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nature Genet.* **38**, 758–769 (2006).
96. Margolis, R. L., Rudnicki, D. D. & Holmes, S. E. Huntington's disease like-2: review and update. *Acta Neurol. Taiwan* **14**, 1–8 (2005).
97. Rudnicki, D. D. *et al.* Huntington's disease-like 2 is associated with CUG repeat-containing RNA foci. *Ann. Neurol.* **61**, 272–282 (2007).
98. Soret, J., Gabut, M. & Tazi, J. SR proteins as potential targets for therapy. *Prog. Mol. Subcell. Biol.* **44**, 65–87 (2006).
99. Xiao, S. H. & Manley, J. L. Phosphorylation of the ASF/SF2 RS domain affects both protein–protein and protein–RNA interactions and is necessary for splicing. *Genes Dev.* **11**, 334–344 (1997).
100. Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA* **6**, 1197–1211 (2000).
101. Orengo, J. P., Bundman, D. & Cooper, T. A. A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Res.* **34**, e148 (2006).
102. Bonano, V. I., Oltean, S., Brazas, R. M. & Garcia-Blanco, M. A. Imaging the alternative silencing of *FGFR2* exon IIIb *in vivo*. *RNA* **12**, 2073–2079 (2006).
103. Garcia-Blanco, M. A. Alternative splicing: therapeutic target and tool. *Prog. Mol. Subcell. Biol.* **44**, 47–64 (2006).
104. Hua, Y., Vickers, T. A., Baker, B. F., Bennett, C. F. & Krainer, A. R. Enhancement of *SMN2* exon 7 inclusion by antisense oligonucleotides targeting the exon. *PLoS Biol.* **5**, e73 (2007).
105. Cartegni, L. & Krainer, A. R. Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nature Struct. Biol.* **10**, 120–125 (2003).
106. Skordis, L. A., Dunckley, M. G., Yue, B., Eperon, I. C. & Muntoni, F. Bifunctional antisense oligonucleotides provide a *trans*-acting splicing enhancer that stimulates *SMN2* gene expression in patient fibroblasts. *Proc. Natl Acad. Sci. USA* **100**, 4114–4119 (2003).
107. Larriba, S. *et al.* Testicular *CFTR* splice variants in patients with congenital absence of the vas deferens. *Hum. Mol. Genet.* **7**, 1739–1743 (1998).

108. Niksic, M., Romano, M., Buratti, E., Pagani, F. & Baralle, F. E. Functional analysis of *cis*-acting elements regulating the alternative splicing of human *CFTR* exon 9. *Hum. Mol. Genet.* **8**, 2339–2349 (1999).
109. Chiba-Falek, O. *et al.* The molecular basis of disease variability among cystic fibrosis patients carrying the 3849 + 10 kb C→T mutation. *Genomics* **53**, 276–283 (1998).
110. Graham, R. R. *et al.* A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nature Genet.* **38**, 550–555 (2006).
111. Ueda, H. *et al.* Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease. *Nature* **423**, 506–511 (2003).
112. Atz, M. E., Rollins, B. & Vawter, M. P. *NCAM1* association study of bipolar disorder and schizophrenia: polymorphisms and alternatively spliced isoforms lead to similarities and differences. *Psychiatr. Genet.* **17**, 55–67 (2007).
113. Law, A. J., Kleinman, J. E., Weinberger, D. R. & Weickert, C. S. Disease-associated intronic variants in the *ERBB4* gene are related to altered *ERBB4* splice-variant expression in the brain in schizophrenia. *Hum. Mol. Genet.* **16**, 129–141 (2007).
114. Mango, R. *et al. In vivo* and *in vitro* studies support that a new splicing isoform of *OLR1* gene is protective against acute myocardial infarction. *Circ. Res.* **97**, 152–158 (2005).
115. Bonnevie-Nielsen, V. *et al.* Variation in antiviral 2′,5′-oligoadenylate synthetase (2′5′AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the *OAS1* gene. *Am. J. Hum. Genet.* **76**, 623–633 (2005).
116. Komamura, K. *et al.* The role of a common *TNNT2* polymorphism in cardiac hypertrophy. *J. Hum. Genet.* **49**, 129–133 (2004).
117. Laitinen, T. *et al.* Characterization of a common susceptibility locus for asthma-related traits. *Science* **304**, 300–304 (2004).
118. Caffrey, T. M., Joachim, C., Paracchini, S., Esiri, M. M. & Wade-Martins, R. Haplotype-specific expression of exon 10 at the human *MAPT* locus. *Hum. Mol. Genet.* **15**, 3529–3537 (2006).
119. Stanton, T. *et al.* A high-frequency polymorphism in exon 6 of the CD45 tyrosine phosphatase gene (*PTPRC*) resulting in altered isoform expression. *Proc. Natl Acad. Sci. USA* **100**, 5997–6002 (2003).
120. Jacobsen, M. *et al.* A point mutation in *PTPRC* is associated with the development of multiple sclerosis. *Nature Genet.* **26**, 495–499 (2000).
121. Thude, H., Hundrieser, J., Wonigeit, K. & Schwinzer, R. A point mutation in the human *CD45* gene associated with defective splicing of exon A. *Eur. J. Immunol.* **25**, 2101–2106 (1995).
122. Zhu, H. *et al.* A common polymorphism decreases low-density lipoprotein receptor exon 12 splicing efficiency and associates with increased cholesterol. *Hum. Mol. Genet.* **16**, 1765–1772 (2007).

Acknowledgements

A great deal of outstanding work has been published on the topics covered in this Review, and space limitations prevent mention of all important contributions. The work in our laboratory is supported by the US National Institutes of Health (NIH) and the Muscular Dystrophy Association (USA).

Competing interests statement

The authors declare no competing financial interests.

DATABASES

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
CFTR | *DMPK* | *FMR1* | *HLA-DQB1* | *HPRP3* | *IKBKAP* | *IPH3* | *MCAD* | *PRPF8* | *PRPF31* | *RON* | *Scnm1* | *SCN1A* | *Scn8a* | *SMN1* | *SNRPN*

OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
cystic fibrosis | familial dysautonomia | FXTAS | HDL2 | Huntington disease | myotonic dystrophy | oculopharyngeal muscular dystrophy | Prader–Willi Syndrome | retinitis pigmentosa | SCA8 | spinal muscular atrophy | spinocerebellar ataxias

UniProtKB: http://ca.expasy.org/sprot
CNBP | CUGBP1 | MBNL1 | RPS6KB1 | SFRS6 | SFRS8 | SNURF | TIA1

FURTHER INFORMATION

Alternative Splicing Database Project: http://www.ebi.ac.uk/asd
Alternative Splicing Annotation Project database: http://bioinfo.mbi.ucla.edu/ASAP
Cooper Laboratory homepage: http://www.bcm.edu/pathology/labs/cooper/index.htm
ESE finder: http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home
Human Gene Mutation Database: http://www.hgmd.cf.ac.uk/ac/index.php
RESCUE-ESE Web Server: http://genes.mit.edu/burgelab/rescue-ese
RetNet: Summaries of Genes and Loci Causing Retinal Diseases: http://www.sph.uth.tmc.edu/Retnet/sum-dis.htm
Spidey mRNA-to-genomic alignment: http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/spideydoc.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF.