DISEASE MECHANISMS

RNA mis-splicing in disease

Marina M. Scotti and Maurice S. Swanson

Abstract | The human transcriptome is composed of a vast RNA population that undergoes further diversification by splicing. Detecting specific splice sites in this large sequence pool is the responsibility of the major and minor spliceosomes in collaboration with numerous splicing factors. This complexity makes splicing susceptible to sequence polymorphisms and deleterious mutations. Indeed, RNA mis-splicing underlies a growing number of human diseases with substantial societal consequences. Here, we provide an overview of RNA splicing mechanisms followed by a discussion of disease-associated errors, with an emphasis on recently described mutations that have provided new insights into splicing regulation. We also discuss emerging strategies for splicing-modulating therapy.

Long non-coding RNAs
(lncRNAs). RNAs of >200 nucleotides in length that generally do not encode proteins.

Pseudogenes
Non-functional versions of genes that are generated either by duplication and mutation or by retrotransposition.

Splicing factors
Proteins that participate in splicing regulation but are not stable constituents of small nuclear ribonucleoprotein particles (snRNPs).

Single-nucleotide polymorphisms
(SNPs). Variations in individual nucleotides that are common in the human genome and can influence splicing regulation.

Department of Molecular Genetics and Microbiology, Center for NeuroGenetics and the Genetics Institute, University of Florida, College of Medicine, Gainesville, Florida 32610–3610 USA.
Correspondence to M.S.S.
mswanson@ufl.edu

doi:10.1038/nrg.2015.3
Published online 23 Nov 2015

Recent analysis from the Encyclopedia of DNA Elements (ENCODE) project[1] (GRCh38, Ensembl79) indicates that most of the human genome is transcribed and consists of ~60,000 genes (~20,000 protein-coding genes, ~16,000 long non-coding RNAs (lncRNAs), ~10,000 small non-coding RNA and 14,000 pseudogenes). Although this gene inventory will change with further analysis, the number of protein-coding genes is surprisingly low given the proteomic complexity that is evident in many tissues, particularly the central nervous system (CNS). High-resolution mass spectrometry studies have identified peptides encoded by most of these annotated genes[2,3], but the number of isoforms expressed from this gene set has been estimated to be at least 5–10-fold higher. For example, long-read sequence analysis of adult mouse prefrontal cortex neurexin (*Nrxn*) mRNAs indicates that only three *Nrxn* genes produce thousands of isoform variants[4]. This diversity is primarily generated by alternative splicing, with >90% of human protein-coding genes producing multiple mRNA isoforms[5–7]. Given the complexity of the precursor RNA sequence elements and *trans*-acting splicing factors that control splicing, it is not surprising that this RNA processing step is particularly susceptible to both hereditary and somatic mutations that are implicated in disease[8]. The central importance of splicing regulation is highlighted by the observation that many disease-associated single-nucleotide polymorphisms (SNPs) in protein-coding genes have been proposed to influence splicing. Although splicing efficiency may vary between individuals owing to variants in the *cis*-acting RNA sequence elements or in the genes encoding *trans*-acting factors that control splicing, most (>90%) disease-associated SNPs lie outside of protein-coding regions[9]. Thus, it is noteworthy that some non-coding RNAs (ncRNAs), including lncRNAs and circular RNAs (circRNAs), have been implicated in splicing regulation[10,11].

In this Review, we focus on RNA mis-splicing in disease, providing background information on splicing mechanisms in BOX 1. We describe why splicing can be prone to errors with potentially pathological consequences, and then summarize mutations in both *cis*-acting RNA sequence elements and *trans*-acting splicing factors that are associated with various diseases, with an emphasis on recently described mutations. The emerging issue of mutation-induced splicing factor aggregation, which is particularly notable in some neurological diseases, is also reviewed, followed by an examination of current studies focused on splicing modulatory therapies to treat human disease.

Splicing errors and disease

The division of eukaryotic genes into exons and introns has clear evolutionary advantages, including regulatory, mutation buffering and coding capacity benefits[12]. However, this split gene architecture introduces a requirement for an intricate splicing regulatory network that consists of an array of RNA regulatory sequences, RNA–protein complexes and splicing factors. Although splicing is composed of a fairly simple set of reactions, the task of the splicing machinery to find authentic 5′ splice sites (5′ss) and 3′ss is problematic for several reasons (BOX 1). First, 5′ss and 3′ss pairs must be carefully identified, particularly in coding regions where a single-nucleotide mistake often results in a frameshift and consequent nonsense-mediated decay (NMD) of the transcript. Second, mammalian gene architecture complicates the difficult task of site selection owing to extensive alternative splicing (BOX 2) and because alternative splice sites may be preferentially selected during embryonic and fetal development as a mechanism to control the levels of the final gene products. Third, human exons are often small, with ~80% of exons <200 bp in length,

and masked by a much larger intronic sequence pool. Fourth, splicing is primarily a co-transcriptional process that is modulated by the rate of transcriptional elongation by RNA polymerase II (RNA Pol II), so multiple regulatory machineries must properly interface to ensure correct splice site selection[13].

As detailed below, recurring themes in splicing regulation and disease presentation are the genetic issues of penetrance and expressivity. Incomplete penetrance and variable expressivity may result from allelic variations, modifier genes and/or environmental factors.

Common cause: pre-mRNA mutations and mis-splicing. The most common type of mutations that alter splicing patterns are *cis*-acting and are located in either core consensus sequences (5′ss, 3′ss and branch point (BP)) or the regulatory elements that modulate spliceosome recruitment, including exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE) and intronic splicing silencer (ISS) elements[8] (BOX 1). Mutations in these regulatory elements have been documented in multiple diseases that have characteristic effects on many tissues (TABLE 1, *cis*). An early splicing mutation described soon after the discovery of splicing was a point mutation that generates an alternative 3′ss in *HBB*, which encodes β-globin, resulting in β$^+$-thalassaemia, a condition that is characterized by reduced β-globin protein levels and anaemia[14–16]. More recent examples include: splice site mutations in dystrophin (*DMD*), which result in loss of dystrophin function and Duchenne muscular dystrophy[17,18] (discussed in further detail below); polymorphic UG and U tracts near the 3′ss of *CFTR* (cystic fibrosis transmembrane conductance regulator) exon 9, which modify the severity of cystic fibrosis[19,20]; and ESE, ESS and 5′ss mutations in *MAPT* (microtubule-associated protein tau) exon 10, which cause frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17)[21].

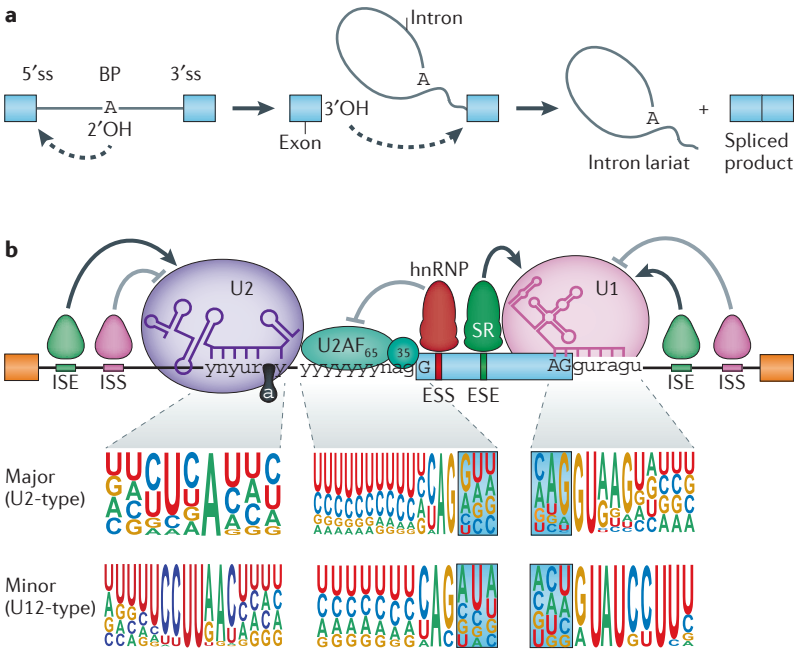
In contrast to the above gene mutations that cause a single type of disease, mutations in several types of sequence elements in *LMNA*, the gene encoding lamins A, C, Δ10 and C2 result in multiple pathological phenotypes[22]. Lamins are type V intermediate filament proteins of the nucleus that have crucial roles in differentiated cell nuclear architecture (peripheral lamins) and gene expression (nucleoplasmic lamins). Laminopathies comprise a heterogeneous group of over 14 diseases, including cardiomyopathies, hereditary peripheral neuropathies, lipodystrophies, muscular dystrophies and premature ageing (progeroid) syndromes[23].

Interestingly, 5′ss mutations in *LMNA* cause two progressive but distinct disorders (FIG. 1): limb girdle muscular dystrophy 1B (LGMD1B) primarily affects the proximal muscles of the shoulders and hips, whereas familial partial lipodystrophy type 2 (FPLD2) is characterized by a selective loss and abnormal distribution of body fat[24,25]. Both 5′ss mutations lead to intron retention (albeit for different introns), frameshifting and the generation of a premature termination codon (PTC) that should activate NMD and increase *LMNA* RNA turnover. However, the different disease presentations suggest that distinct truncated LMNA proteins may be produced by intron 8 versus intron 9 retention (FIG. 1a,b). Moreover, an unrelated premature ageing disease, Hutchinson–Gilford progeria syndrome (HGPS), is caused by the utilization of an alternative 5′ss in *LMNA* exon 11, resulting in a 150 nucleotide deletion that generates progerin, a carboxy-terminal truncated protein that is

Box 1 | **RNA splicing regulation**

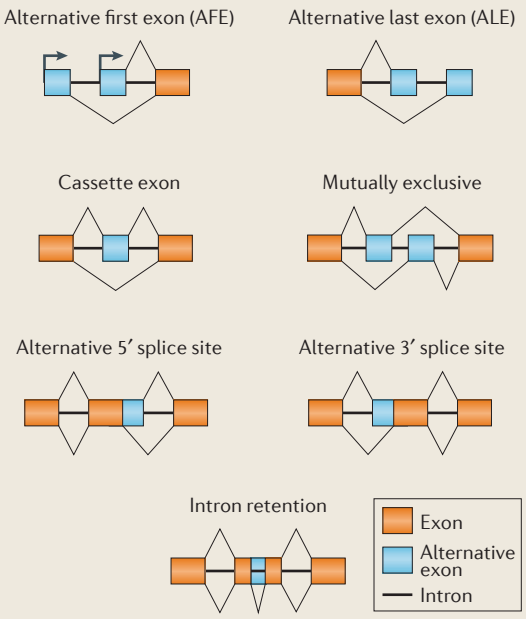
RNA splicing, which is the removal of introns followed by exon ligation, is a two-step biochemical process. Sequential transesterification reactions are initiated by a nucleophilic attack of the 5′ splice site (5′ss) by the branch adenosine (branch point; BP) in the downstream intron resulting in the formation of an intron lariat intermediate with a 2′, 5′-phosphodiester linkage. This is followed by a 5′ss-mediated attack on the 3′ss, leading to the removal of the intron lariat and the formation of the spliced RNA product (see the figure, part **a**).

The difficult tasks of splice site identification and regulated splicing is accomplished principally by two exceptionally dynamic macromolecular machines, the major (U2-dependent) and minor (U12-dependent) spliceosomes. Each spliceosome contains five small nuclear ribonucleoprotein particles (snRNPs): U1, U2, U4, U5 and U6 snRNAs for the major spliceosome (which processes ~95.5% of all introns[126]); and U11, U12, U4atac, U5 and U6atac snRNAs for the minor spliceosome (see the figure, part **b**). Spliceosome recognition of consensus sequence elements at the 5′ss, 3′ss and BP sites is a crucial step in the splicing pathway, and is modulated by an array of *cis*-acting exonic and intronic splicing enhancers (ESEs and ISEs, respectively) and exonic and intronic splicing silencers (ESSs and ISSs, respectively), which are recognized by auxiliary splicing factors, including the Ser/Arg-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs). Although early studies indicated that U12-dependent introns initiated with AT and ended with AC, previously referred to as ATAC introns (this is also why the minor spliceosome snRNAs are named U4atac and U6atac), subsequent studies demonstrated that these terminal dinucleotides were not required[126]. In part **b** of the figure, the height of the residue corresponds to relative frequency of each nucleotide in each given position. U2 and U12 consensus sequence frequencies were obtained from the Splice Rack and U12 databases, respectively, and BP site data and probabilities were calculated with Pictogram (see further information). Ultimately, this intricate network of RNA and protein interactions results in the recruitment of spliceosomal components followed by snRNP remodelling, spliceosome activation, catalysis and generation of the spliced RNA product.



Box 2 | **Alternative splicing**

High-throughput RNA sequencing (RNA-seq) studies have suggested that alternative splicing is a routine activity in human cells with 90–95% of human multi-exon genes producing transcripts that are alternatively spliced[5,7]. Alternative splicing adds another layer of complexity with multiple and developmentally regulated splicing patterns including the inclusion of alternative first and last exons (AFE and ALE, respectively), retained intron, cassette exon, mutually exclusive cassettes and alternative 5′ and 3′ splice sites[5] (see the figure). Cassette exon skipping is the most common alternative splicing event in humans but a recent study demonstrates that intron retention is also routine in mammals, occurring in nearly 75% of multi-exon genes, and is a co- or post-transcriptional mechanism designed to reduce transcript levels during development[6,127]. The interrelationship of a large array of *cis*-regulatory elements and *trans*-acting RNA-binding proteins (RBPs) suggests that alternative splicing is controlled by a 'splicing code' that could be useful as a predictive tool for cell-, and tissue-specific responses to developmental transitions and environmental changes.



Nonsense-mediated decay
(NMD). A process of enhanced RNA turnover induced by a premature termination codon (PTC) which is designed to block the synthesis of truncated proteins and modulate the appearance of full-length proteins during development.

Penetrance
The percentage of individuals carrying a disease mutation who show clinical symptoms. Incomplete, or reduced, penetrance occurs when not all individuals with a particular genetic mutation develop the associated disease.

Expressivity
The degree to which a mutant gene is phenotypically expressed. Variable expressivity refers to the symptomatic range that is displayed by different individuals with the same mutation.

Core consensus sequences
Conserved RNA sequence motifs, including the 5′ and 3′ splice sites and the branch point region, which are required for spliceosome recruitment.

Branch point
(BP). A partially conserved sequence, generally <50 nucleotides upstream of the 3′ splice site (see BOX 1), that reacts with the 5′ splice site during the first step of the splicing reaction.

Spliceosome
The large RNA–protein complex that catalyses splicing and is composed of multiple small nuclear RNAs (snRNAs) and many associated protein factors. Whereas the major and minor spliceosomes both contain U5, the other snRNA components differ (for the major spliceosome, U1, U2, U4, U6; for the minor spliceosome, U11, U12, U4atac, U6atac) (see FIG. 2).

Tri-snRNP
A preassembled complex of U4 snRNA hybridized to U6 (U4/U6 or U4atac/U6atac) that also contains U5 (U4/U6.U5) together with associated proteins (see FIG. 2).

constitutively farnesylated and remains associated with the nuclear membrane during mitosis[26,27] (FIG. 1c). Other mutations at the exon 4 3′ ss cause the addition of an extra three amino acids at the exon 3–4 border and lead to dilated cardiomyopathy (DCM), a heart condition that is caused by thinning and dilation of the ventricles and progressive atrioventricular block, a cardiac conduction defect[28] (FIG. 1d). Thus, splicing mutations result in abnormal LMNA proteins, which themselves lead to a range of pathological effects.

Core spliceosome mutations in retinitis pigmentosa. In addition to mutations that alter precursor RNA sequence elements which regulate splicing, mutations in core constituents of the spliceosome also underlie a discrete set of diseases, including retinal degenerative disorders and cancer (TABLE 1, spliceosome). Retinitis pigmentosa (RP), a progressive disease that is characterized by initial night blindness and loss of peripheral vision followed by central retinal degeneration, is the most prevalent form of hereditary blindness[29]. Multiple gene mutations cause autosomal dominant RP (adRP), but six of these genes encode components, or associated factors, of the spliceosomal U4/U6.U5 tri-small nuclear ribonucleoprotein particle (tri-snRNP; namely pre-mRNA processing factor 3 (PRPF3), PRPF4, PRPF6, PRPF8, PRPF31 and SNRNP200 (also known as BRR2)) (FIG. 2). Among surveyed tissues, the human retina expresses the highest levels of some housekeeping genes and major, as well as minor, spliceosomal snRNAs. Therefore, *PRPF* mutations might cause global splicing dysregulation that manifests in the retina because of enhanced splicing activity[30].

Studies on patient lymphoblasts with *PRPF3*, *PRPF8* and *PRPF31* mutations show that these mutant cells also display impaired constitutive and alternative splicing together with changes in snRNA stoichiometry and tri-snRNP composition. Variable expressivity and/or incomplete penetrance have also been reported for RP-associated mutations in U4/U6.U5 tri-snRNP components[31–33]. For *PRPF31*, incomplete penetrance in some families with adRP results from increased expression of the wild-type *PRPF31* allele in asymptomatic carriers owing to reduced expression of the repressive modifier gene *CNOT3* (CCR4-NOT transcription complex, subunit 3)[34]. These results support a haploinsufficiency model for spliceosome-associated adRP and highlight the importance of optimal U4/U6.U5 tri-snRNP function in the retina.

However, gain-of-function pathogenesis models have also been proposed. For example, some *PRPF3* mutations lead to large aggregate formation and apoptosis in photoreceptor, but not epithelial or fibroblast cell lines, suggesting that these aggregates may directly interfere with normal cellular activities[35]. Loss-of-function mutations may also occur in adRP. This possibility is exemplified by PRPF8, a highly conserved and large protein that is a component of the U5 snRNP and is positioned in the catalytic core of the spliceosome. PRPF8 C-terminal mutations cause an early onset and severe form of adRP, and structural analysis of this region shows that it inserts into the RNA-binding tunnel of the SNRNP200 helicase. This insertion event intermittently blocks, and thus controls, the ATP-dependent U4/U6 unwinding activity of this helicase. Thus, loss of the PRPF8 RNA-blocking activity owing to some adRP-linked mutations would be expected to cause mis-regulation of U4/U6 unwinding due to premature helicase activity[36].

Different mutations in a single gene may also result in haploinsufficiency and dominant-negative effects in adRP. Heterozygous single-nucleotide variants and deletion variants in the *PRPF4* coding and promoter

regions, respectively, have been reported recently[37]. *PRPF4* encodes a 60 kDa protein that is important for U4/U6 di-snRNP stability. Whereas the promoter deletion causes decreased *PRPF4* expression in patient fibroblasts, the coding region variant (p.Pro315Leu) results in the upregulation of *PRPF4* together with several U4/U6.U5 tri-snRNP components (*PRPF3*, *PRPF6*, *PRPF8* and *PRPF31*) and other splicing factors (*SRSF1* and *SRSF2*). In addition, overexpression of human PRPF4 in which the Pro at position 315 is mutated to Leu

Table 1 | **Disease-associated splicing alterations**

Disease	Gene (mutation)	Mechanism	Splicing effect	Inheritance
Cis				
Limb girdle muscular dystrophy type 1B (LGMD1B)	*LMNA*[24] (c.1608+5G>C)	5′ss mutation	Intron 9 retention resulting in NMD	Dominant
Familial partial lipodystrophy type 2 (FPLD2)	*LMNA*[25] (c.1488+5G>C)	5′ss mutation	Intron 8 retention resulting in NMD	Dominant
Hutchinson–Gilford progeria syndrome (HGPS)	*LMNA*[26] (c.1824C>T)	Alternative 5′ss	150 nt deletion in exon 11, resulting in progerin generation	Dominant
Dilated cardiomyopathy (DCM)	*LMNA*[28] (c.640-10A>G)	Alternative 3′ss	Extension of exon 4 adding 3 amino acids to lamin A/C	Dominant
Familial dysautonomia (FD)	*IKBKAP*[128] (c.2204+6T>C)	Decreased U1 recruitment	Exon 20 skipping	Recessive
Duchenne muscular dystrophy (DMD)	*DMD*[129] Exon 45–55 deletions are common	Exon deletions and skipping	Frameshift resulting in NMD	X-linked
Becker muscular dystrophy (BMD)	*DMD*[130] (c.4250T>A)	ESS creation	Exon 31 partial in-frame skipping	X-linked
Early-onset Parkinson disease (PD)	*PINK1* (REF. 131) (c.1488+1G>A)	U1 5′ss mutation	Cryptic splice site usage, resulting in exon 7 skipping	Recessive
Frontotemporal dementia with parkinsonism chromosome 17 (FTDP-17)	*MAPT*[132] (c.892A>G)	ESS mutation	Increased exon 10 inclusion	Dominant
X-linked parkinsonism with spasticity (XPDS)	*ATP6AP2* (REF. 133) (c.345C>T)	Novel ESS creation	Increased exon 4 exclusion	X-linked
Spliceosome				
Retinitis pigmentosa (adRP)	*PRPF6* (REF. 134) (c.2185C>T)	Abnormal nuclear localization	Decreased U4/U6 interaction affecting spliceosome assembly and recycling	Dominant
	SNRNP200 (REF. 135) (c.3260C>T), (c.3269G>T)	• Decreased helicase activity • Decreased proof-reading	Compromised splice site recognition, leading to mis-spliced mRNAs	Dominant
Myelodysplastic syndromes (MDS)	*U2AF1* (REF. 46) (c.101G>A)	Altered 3′ss preference	Increased alternative 3′ss usage	Somatic
Microcephalic osteodysplastic primordial dwarfism type 1 (MOPD I)	*RNU4ATAC*[54–56] (g.30G>A), (g.50G>A), (g.50G>C), (g.51G>A), (g.53C>G), (g.55G>A), (g.111G>A)	5′ and 3′ stem loop mutations & secondary structure disruption	Compromised minor spliceosome activity	Recessive
Trans				
Spinal muscular atrophy (SMA)	*SMN1* (REFS 136,137) (c.922+6T/G), deletion	Loss of SMN full-length protein	Altered RNP biogenesis[98]	Recessive
Amyotrophic lateral sclerosis (ALS)	*TARDP*[77] (c.991C>A), (c.1009A>G)	C-terminal mutations alter protein-protein interactions	TDP-43 target mis-splicing	Sporadic and Dominant
	FUS[138] (c. 1566C>T), (c. 1561T>G)	• Decreased U1 interaction • Increased SMN binding	FUS target mis-splicing	Dominant
Dilated cardiomyopathy (DCM)	*RBM20* (REF. 139) (c.1962T>G)	Altered R/S RNA binding domain	*TTN* mis-splicing	Dominant
Limb-girdle muscular dystrophy 1G (LGMD1G)	*HNRPDL*[140] (c. 1667G>A), (c. 1667G>C)	Altered import of HNRPDL into nucleus	HNRPDL target mis-splicing	Dominant
Autosomal dominant leukodystrophy (ADLD)	*LMNB1* (REF. 141) duplication	Increased *RAVER2* expression	PTBP1 target mis-splicing mediated by RAVER2	Dominant

ATP6AP2, ATPase, H+ transporting, lysosomal accessory protein 2; *DMD*, dystrophin; ESS, exonic splicing silencer; *HNRPDL*, heterogeneous nuclear ribonucleoprotein D-like; *IKBKAP*, inhibitor of κ-light polypeptide gene enhancer in B cells, kinase complex-associated protein; *LMNA*, lamin A; *MAPT*, microtubule-associated protein tau; NMD, nonsense-mediated decay; *PRPF6*, pre-mRNA processing factor 6; PTBP1, polypyrimidine tract binding protein 1; RNP, ribonucleoprotein; *SMN1*, survival of motor neuron 1; ss, splice site.

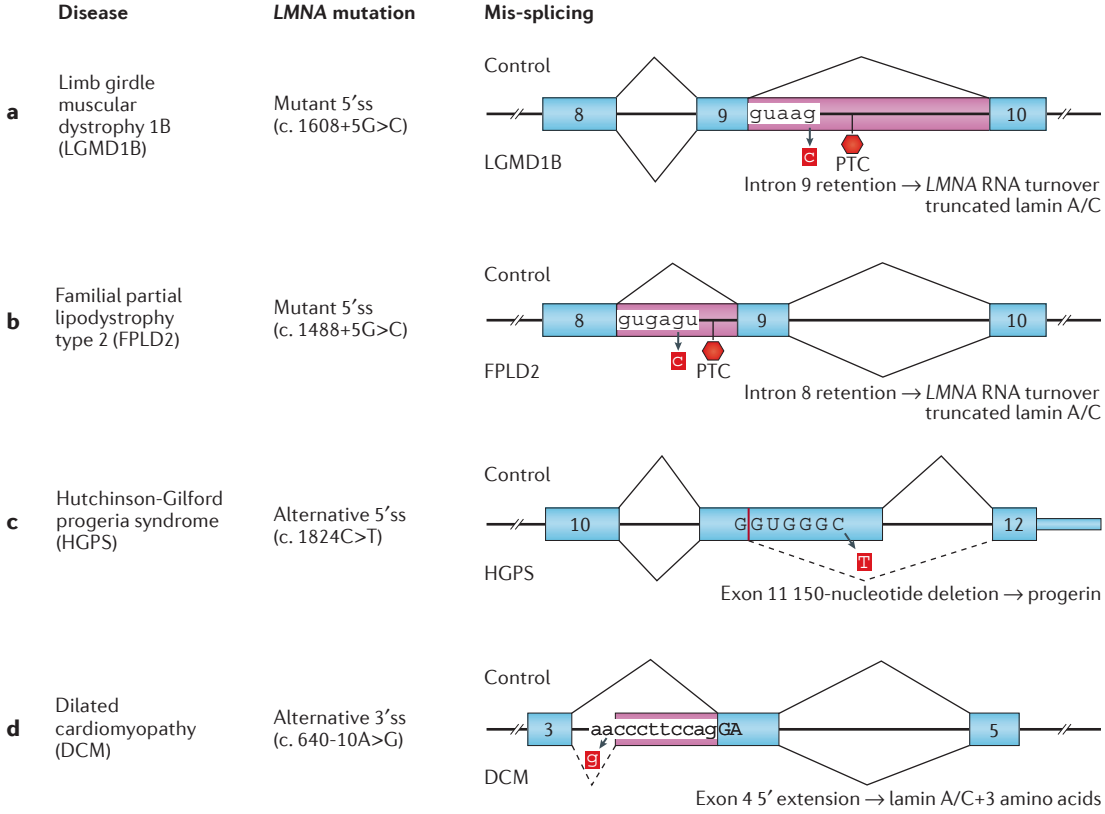


Figure 1 | **Mis-splicing of a single gene results in different diseases.** Aberrant splicing of lamin A (*LMNA*) pre-mRNA is associated with multiple hereditary disorders. Normal exons are shown in blue, introns are shown as thick black lines, normal splicing is indicated by thin black lines, and disease-associated splicing is indicated in dotted lines or purple boxes (intron retention). **a** | Limb girdle muscular dystrophy type 1B (LGMD1B) is caused by a G>C 5′ splice site (5′ss) mutation that results in intron 9 retention, a premature termination codon (PTC) and nonsense-mediated decay (NMD). c.1608 + 5 indicates that the mutations occurs 5 nucleotides into the intron that follows coding position (c) 1608. However, a lamin A/C protein truncated in intron 9 with a unique carboxy-terminal sequence may also be produced. **b** | In familial partial lipodystrophy type 2 (FPLD2), a G>C transversion mutation occurs in the exon 8 5′ss, leading to intron 8 retention, NMD and potential translation of another truncated lamin A/C with a unique C-terminal region. **c** | A common cause of Hutchinson–Gilford progeria syndrome (HGPS) is a C>T transition in exon 11, which activates a cryptic 5′ss and results in a 150 nucleotide deletion that is translated into the ageing-associated protein progerin. **d** | For LMNA-linked dilated cardiomyopathy (DCM), an alternative 3′ss is generated by an A>G mutation upstream of the normal exon 4 3′ ss so that nine additional nucleotides are inserted in-frame between exons 3 and 4, resulting in a 3-amino-acid insertion in the resultant protein.

(PRPF4^{Pro315Leu}) in zebrafish results in larval deformity and retinal phenotypes. Although the central question of why the U4/U6.U5 tri-snRNP is particularly important for normal retinal function remains unanswered, several intriguing clues are starting to emerge. The initial events in RP include loss of rod photoreceptors, the cells responsible for vision under low light conditions, and of the retinal pigment epithelium (RPE), the monolayer of cells that carry out functions such as the phagocytosis of the outer segments of photoreceptors (~10% of rod cell volume)[38]. Thus, the RPE has a high rhythmic metabolic burden, and recent results suggest that the RPE may be the primary cell type affected by *PRPF* mutations. Mouse RPE morphology is sensitive to *Prpf* mutations, and RPE phagocytic function is inhibited in *Prpf3*$^{Thr494Met/Thr494Met}$ and *Prpf8*$^{His2309Pro/His2309Pro}$ knock-in, as well as *Prpf31*$^{+/-}$ hemizygous RPE cell cultures[39]. Further studies designed to investigate how these *Prpf* mutations alter splicing patterns in these mutants should provide mechanistic insights into U4/U6.U5 tri-snRNP dynamics in normal versus diseased retinal cells.

Haploinsufficiency
A condition due to inactivating mutations in one copy of a gene when expression from the remaining copy is insufficient to produce an unaffected phenotype.

Spliceosome dysregulation in cancer. Mis-regulation of alternative splicing is an important factor in several types of cancer[40]. In addition, somatic mutations that affect the expression of core spliceosome components have an important role in cancer progression. For example, *PRPF6*, a U5 snRNP protein that mediates interactions between U5 and the U4/U6 di-snRNP to form the U4/U6.U5 tri-snRNP (FIG. 2), is overexpressed in colorectal carcinoma owing to chromosomal instability, copy number gain and possibly other factors[41], and this promotes cancer cell proliferation. The increased *PRPF6* expression in cancer cell lines correlates with an alternative

splicing event that generates an oncogenic form of the stress-activated kinase ZAK. A direct role for *PRPF8* somatic mutations and hemizygous deletion has also been proposed for myelodysplastic syndromes (MDS), which are the most prevalent forms of adult myeloid malignancies and are characterized by abnormal growth or development of blood cells[42]. In contrast to *PRPF6*, reduced expression of *PRPF8* correlates with increased cell proliferation, and *PRPF8* heterozygous mutations or hemizygous deletions result in widespread alternative splicing defects owing to enhanced activation of suboptimal splice sites[43].

Exome and whole-genome sequencing studies have also uncovered frequent somatic mutations in a key group of spliceosome-associated components, including SF3B1, U2AF1 and U2AF2 (see FIG. 2) in several types of myeloid neoplasms[44,45]. High-throughput RNA sequencing (RNA-seq) results indicate that *U2AF1* mutations alter haematopoiesis and cause changes in 3′ss recognition, resulting in the mis-splicing of hundreds of gene transcripts[46–48]. Mutations in some splicing factor genes also occur frequently in myelodysplastic syndromes and chronic myelomonocytic leukaemia. One example is *SRSF2*, which encodes a Ser/Arg-rich (SR) splicing factor (see BOX 1). Similar to *U2AF1* mutations, these *SRSF2* mutations alter the RNA-binding characteristics of SRSF2 and result in extensive changes in splicing patterns and impairment of haematopoietic cell differentiation[49–51]. Importantly, antitumour drugs that target the spliceosome have been described, including for cancers that are driven by overexpression of the *MYC* oncogene and by increased levels of nascent RNAs[52,53].

Development and stress: key roles for the minor spliceosome. New roles for minor spliceosomal snRNAs and the U4atac/U6atac.U5 tri-snRNP during fetal development have been reported (FIG. 2). Homozygous mutations in *RNU4ATAC*, which codes for the U4atac snRNA, leads to microcephalic osteodysplastic primordial dwarfism type 1 (MOPD I; also known as Taybi–Linder syndrome), an autosomal recessive developmental disorder that is characterized by intrauterine growth

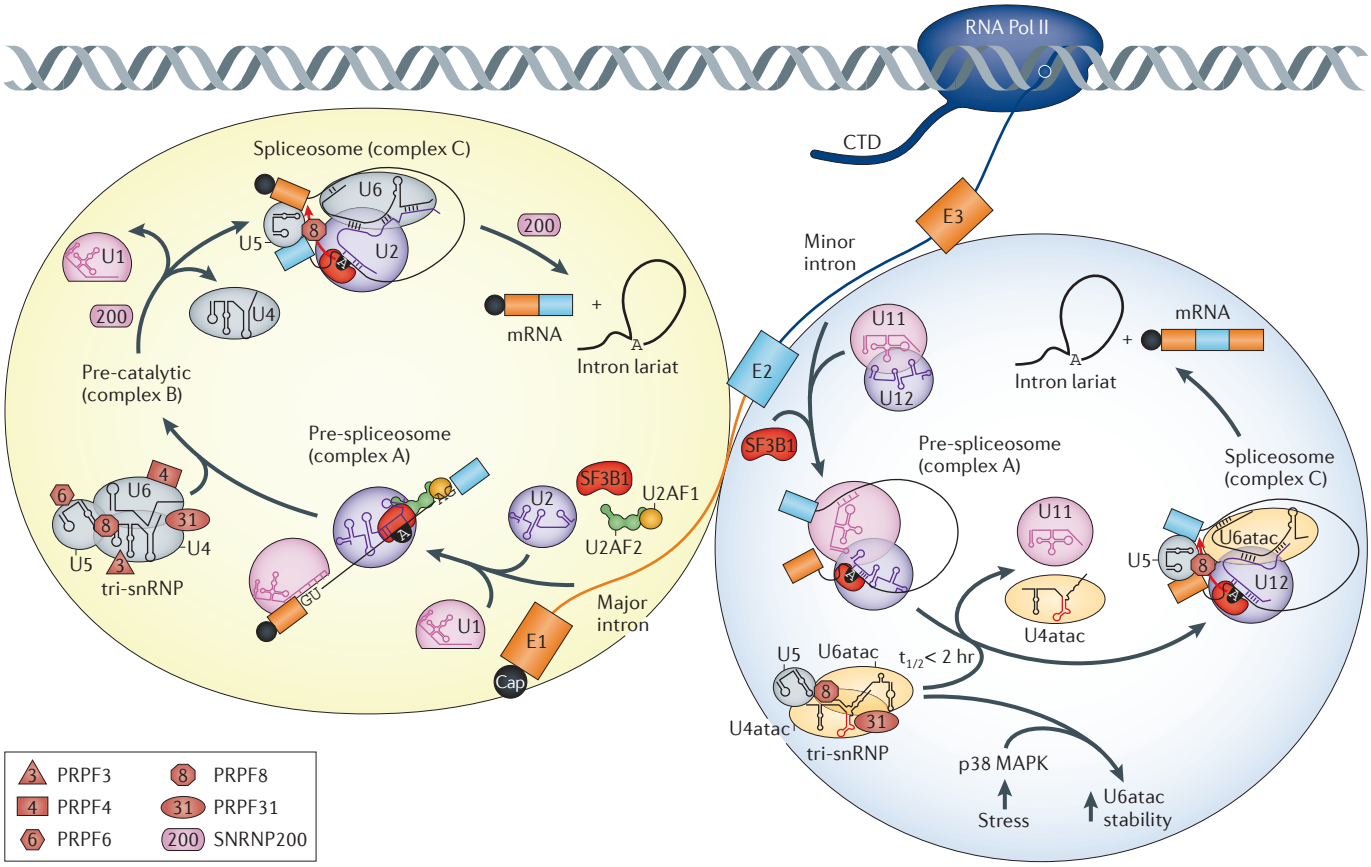


Figure 2 | **Major and minor spliceosome mutations.** The figure shows the splicing steps and core spliceosomal components of both the major (U2-dependent) and minor (U12-dependent) spliceosomes, including their interactions in the pre-spliceosomal complex (complex A) and spliceosome (complex C). Pre-mRNA processing factor 3 (PRPF3), PRPF4), PRPF6, PRPF8 and PRPF31 components of the U4/U6.U5 tri-small nuclear ribonucleoprotein (tri-snRNP) dysregulated in autosomal dominant retinitis pigmentosa (adRP) are shown. Also indicated is the SNRNP200 helicase, which is required at several dissociation steps in the spliceosomal cycle. Several PRPF components are common to both the U4/U6.U5 tri-snRNP and the U4atac/U6atac.U5 tri-snRNP complexes. Some mutations in the U4atac snRNA 5′ stem-loop found in microcephalic osteodysplastic primordial dwarfism type 1 (MOPD I) are highlighted in red. In addition, stress-induced upregulation of p38 mitogen-activated protein kinase (MAPK) leads to increased stability of U6atac ($t_{1/2}$<2 hours).

retardation and multiple tissue abnormalities that lead to early death[54–56] (TABLE 1). Most MOPD I mutations disrupt the U4atac snRNA 5′ stem-loop, inhibit binding of U4atac/U6atac di-snRNP proteins and decrease the levels of U4atac/U6atac.U5 tri-snRNP so that minor intron splicing is impaired[57]. Alternatively, a transition mutation (124G>A) near the Sm protein-binding site results in reduced levels of U4atac snRNA.

The idea that minor spliceosome levels have key regulatory roles in gene expression is also supported by the observation that minor introns act as molecular switches that modulate stress-induced expression of their host genes[58]. Under normal conditions, U6atac is unstable ($t_{1/2}$ <2 hours) relative to other snRNAs. However, the stress-induced kinase p38 mitogen-activated protein kinase (MAPK) increases U6atac snRNA stability, resulting in enhanced splicing of genes that contain minor introns, including the tumour suppressor *PTEN*[58].

Splicing factor dysregulation: perils of imbalance. Recruitment of the spliceosome to splice sites is modulated by the dynamic association of splicing factors with RNA regulatory elements on nascent transcripts. Alternative splicing is particularly important in the brain, and mutations in factors connected with splicing regulation cause a range of neurological diseases, from spinal muscular atrophy (SMA) to amyotrophic lateral sclerosis (ALS; also known as motor neuron disease) (both discussed in more detail below). RNA-binding proteins (RBPs) are generally multi-functional and many have key roles in splicing regulation. Moreover, ultraviolet light-induced RNA–protein crosslinking protocols (such as high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP; also known as CLIP-seq)[59], photoactivatable-ribonucleoside-enhanced-CLIP (PAR-CLIP)[60] and individual-nucleotide resolution-CLIP (iCLIP)[61]) have been used to identify hundreds of RBPs in HeLa and mouse embryonic stem cells[62,63], as well as other cell types[64]. Thus, it is not surprising that mutations in the genes encoding RBPs have been linked to multiple disorders.

An abundant class of nuclear RBPs are factors that associate with both the RNA Pol II carboxy-terminal domain (CTD) and nascent transcripts (FIG. 3). Early electron microscopy studies on the co-transcriptional assembly of large pre-mRNA–protein complexes indicated that mRNA biogenesis involves an orchestrated series of RBP–RNA assembly and remodelling transitions[65]. These, and other studies suggested that both stable and transient protein–protein and protein–RNA interactions play key parts in splicing regulation. Some of the most abundant RBPs in the nucleus contain structural motifs — such as RNA recognition motifs (RRMs), heterogeneous nuclear RNP (hnRNP) K homology domains, zinc fingers and zinc knuckles — that bind RNA, as well as auxiliary, or low complexity (LC), regions composed of limited amino acid diversity (for example, [Gly/Ser]Tyr[Gly/Ser] tripeptide repeats in the RBP FUS). In some RBPs, LC regions also contain prion-like domains that can function as self-templating protein conformers and which are hotspots for disease-associated mutations[66,67]. The importance of LC regions for protein–protein interactions has been highlighted by studies using the small molecule biotinylated isoxazole (b-isox)[68], treatment with which results in the precipitation of numerous RBPs from cell lysates. The LC regions of these proteins are necessary and sufficient to drive this insolubility, and these regions alone undergo a concentration-dependent phase transition to a hydrogel that consists of polymerized amyloid-like fibres. This observation suggests that dynamic interactions between the LC regions of RBPs are a fundamental principle underlying the co-transcriptional assembly of RBP–RNA protein complexes during splicing (FIG. 3a). These association–dissociation events may also be modulated by signal-induced post-translational modifications, such as phosphorylation of Ser-rich LC regions[69].

If protein–protein interaction dynamics are crucial for the assembly and disassembly of RBP–RNA complexes, then mutations in these protein interaction regions might result in coordinate perturbations in protein homeostasis (proteostasis) and RNA splicing (FIG. 3b,c). Indeed, this scenario is exemplified by studies designed to understand the molecular aetiology of ALS, a neurodegenerative disease that affects motor neurons and that leads to paralysis, respiratory failure and death generally within a few years of symptom onset[70]. Although ~90% of ALS cases are sporadic (sALS), the remaining 10% are familial (fALS), and a number of genes have been implicated in fALS by either linkage analysis or candidate gene studies. Relevant to this Review, several of these genes encode RBPs, including *FUS* (also known as *ALS6, TLS* and *HNRNPP2*), *TDP-43* (TAR DNA-binding protein; also known as *TARDBP* and *ALS10*), *ATXN2* (ataxin 2; also known as *ALS13*), *HNRNPA1*, *HNRNPA2B1* and *MATR3* (matrin 3). In addition, microsatellite GGGGCC expansions ($G4C2^{exp}$) in *C9orf72* (REF. 71) cause chromosome 9p21-linked ALS and frontotemporal dementia (C9ALS/FTD), and one disease model proposes that these expansions are transcribed into $rG4C2^{exp}$ RNAs that disrupt splicing by sequestering splicing factors[72,73]. Indeed, a recent transcriptome analysis of autopsied brains (specifically, the cerebellum) from patients with C9ALS/FTD reported thousands of alternative splicing changes in fALS, and to lesser extent in sALS, and motif analysis indicated that many of the alternative cassette exon splicing changes might be regulated by hnRNP H, which has previously been implicated in C9ALS/FTD pathogenesis[74,75]. Mutations in *TDP-43* and *FUS* account for 4% and 5%, respectively, of fALS cases, and most of these mutations occur in the prion-like domains of the encoded proteins (FIG. 3b), including the TDP-43 Gly-rich C-terminal region and the FUS Gln-, Gly-, Ser-, Tyr-rich and Gly-rich amino-terminal regions, although additional FUS mutations occur in the C-terminal Arg- and Gly-rich and nuclear localization signal regions[67,72]. Whereas earlier studies have demonstrated that TDP-43 plays a direct part in the splicing of specific exons, TDP-43 is a member of the hnRNP family of proteins that function in multiple

HITS-CLIP
(High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation; also known as CLIP–seq). A technique to map the binding sites of splicing, and other, factors on target RNAs. Related techniques include photoactivatable-ribonucleoside-enhanced-CLIP (PAR-CLIP) and individual-nucleotide resolution CLIP (iCLIP).

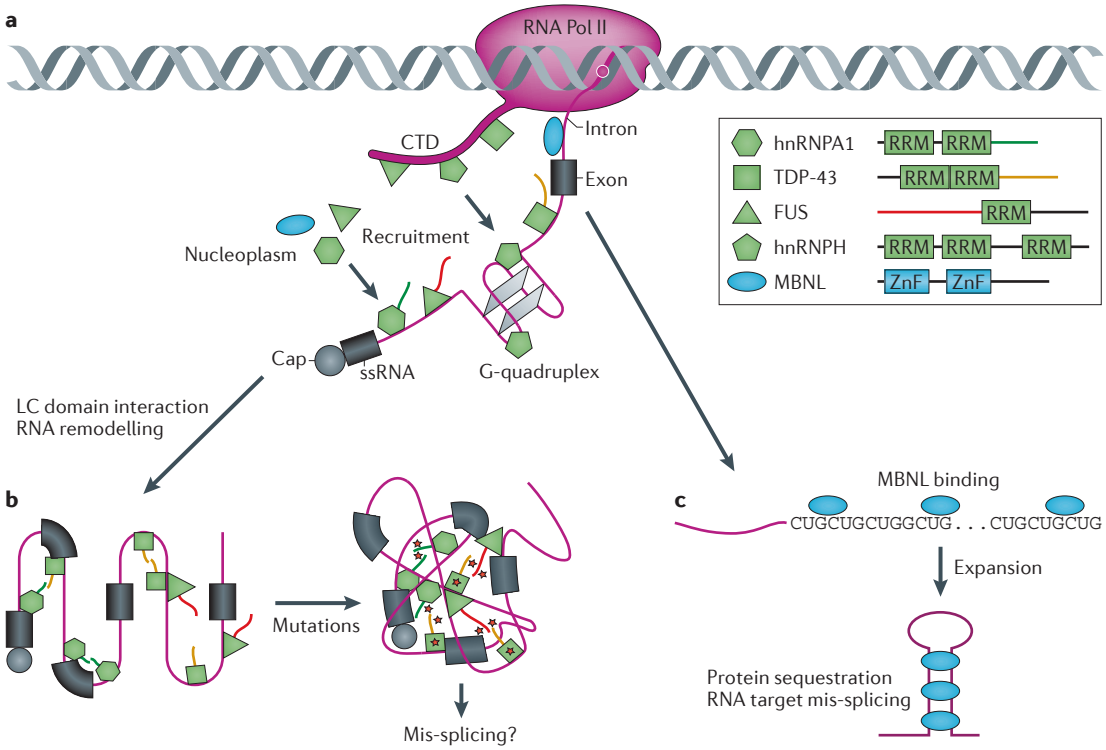


Figure 3 | **Co-transcriptional splicing factor recruitment and disease mutations.** Models for splicing factor and precursor RNA mutations and disease-associated mis-splicing. **a** | Splicing factors recognize and bind to RNA polymerase II (RNA Pol II) transcripts in the nucleoplasm or directly at the carboxy-terminal domain (CTD) of RNA Pol II. These factors may contain RNA-binding motifs (such as RNA recognition motifs (RRMs) or zinc fingers (ZnFs)), as well as auxiliary domains composed of low complexity (LC) regions with prion-like domains in heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1), TDP-43 and FUS (LC regions shown as green, yellow or red lines for hnRNPA1, TDP-43 and FUS, respectively), or other regions that either mediate protein–protein interactions (in muscleblind-like (MBNL)) or function as flexible linkers between RRMs (in hnRNPH). Splicing factors might bind to single-stranded RNA (ssRNA) motifs or pre-formed RNA structures (for example, G-quadruplexes), resulting in the formation of dynamic RNA–RNP complexes that are continuously remodelled by RNA helicases and protein–protein interactions before nuclear export. **b** | Mutations (red star) in the LC regions of hnRNPA1, TDP-43 and FUS could cause mis-folding of RNA–RNP complexes and lead to abnormal splicing. **c** | For diseases caused by microsatellite expansions, splicing factors such as MBNL, which recognize a motif within the repeated sequence, are sequestered by the repeat expansion (ssRNA, top; RNA hairpin, bottom), leading to loss-of-function and mis-splicing.

cellular pathways, so it is not clear how many of these ALS-associated splicing changes are directly regulated by TDP-43 versus secondary effects of TDP-43 nuclear depletion and/or cytoplasmic accumulation[76]. For example, transgenic mice expressing human TDP-43^{Gln331Lys} and TDP-43^{Met337Val} mutants at levels similar to endogenous TDP-43 develop progressive motor neuron degeneration with target-specific splicing alterations in the absence of TDP-43 nuclear depletion or aggregation[77]. FUS also recognizes a GU-rich motif (GUGGU) and binds to thousands of gene transcripts to regulate splicing in the CNS[78,79]. Interestingly, FUS targets conserved introns within genes encoding RBPs that are important for splicing regulation, such as intron 7 in *SNRNP70*, which encodes the U1 snRNP-associated 70K protein (U1-70K).

In addition to splicing factors, snRNP components may also be prone to aggregation in some diseases. A recent surprise is the potential connection between U1 snRNP activity, splicing regulation and Alzheimer disease[80]. Mass spectrometry analysis of the sarkosyl-insoluble proteome from the brains of patients with Alzheimer disease, which includes Aβ peptide and tau protein, revealed that several U1 snRNP proteins, including SNRNP70/U1-70K and SNRPA/U1-A, form tangle-like cytoplasmic inclusions that associate with tau neurofibrillary tangles. This aggregate formation correlates with the accumulation of unspliced precursor RNAs. Both *SNRNP70* knockdown and blocking U1 snRNA with an antisense oligonucleotide (ASO) leads to increased levels of the amyloid precursor protein, suggesting that loss of U1 snRNP splicing activity may be an important feature of Alzheimer disease.

Large introns and microexons in neurological disorders. Although introns and exons are highly variable in length, current studies have shown that long (>100 kb) introns and small (≤51 nucleotides[81]) exons present

particular challenges for the splicing machinery. For example, recursive splicing (RS) — the processing of long introns by sequential events that regenerate a 5′ss — was described originally in the *Drosophila melanogaster* Ultrabithorax (*Ubx*) gene. Recently, RS has also been documented in nine human genes by identifying RNA-seq sawtooth read patterns within introns[82,83]. In contrast to *D. melanogaster*, human RS requires the definition of an RS exon downstream of the RS site, and it has been suggested that mutations near these sites might contribute to disease. In this regard, it is noteworthy that the RS read pattern is similar to the FUS-binding distribution determined by iCLIP. However, the relationship between large intron splicing and disease has been brought to the forefront by studies on TDP-43 in ALS. High-throughput strategies (HITS-CLIP and iCLIP) have been used to map binding sites for TDP-43 on RNAs in mouse and human brains[84,85]. The preferred binding motif for this protein is UG/GU-rich clusters, and the mouse brain contains thousands (>6,000) of genes with TDP-43-binding sites, often in distal regions of introns. Transcriptome analysis — using RNA-seq and splicing-sensitive microarrays — of control versus ASO-induced TDP-43 knockdowns in the striatum of the brain revealed that TDP-43 regulates the alternative splicing of target RNAs and is also important for maintaining wild-type levels of some transcripts with large introns[84]. A mechanistic connection between these observations has recently emerged from the demonstration that a normal function of TDP-43 is to repress the splicing at non-conserved cryptic splice sites, which are often located in distal introns, as inclusion of the resulting cryptic exons often makes the RNA susceptible to NMD[86]. Because TDP-43 is a member of the hnRNP A/B protein family, it will be interesting to determine whether depletion of other hnRNPs results in splicing of additional non-conserved cryptic exons.

At the other end of the scale, mis-splicing of microexons is involved in autism spectrum disorders (ASDs). ASDs are a clinically heterogeneous group of neurodevelopmental disorders distinguished by impaired social interactions and communication combined with repetitive behaviours, possibly due to cortical circuit hyperexcitability[87]. Microexons are characterized by a high level of evolutionary conservation and have a prominent regulatory function during neurogenesis of the mammalian brain[81,88]. These exons encode peptides that modulate interactions between neurogenesis factors during brain development, and alternative splicing of microexons is regulated by the SR-related protein SRRM4 (Ser/Arg repetitive matrix protein 4; also known as nSR100), RBFOX (RNA-binding protein fox-1 homologue 3) and PTBP1 (polypyrimidine tract-binding protein 1)[88]. Importantly, RBFOX1 regulates the alternative splicing of genes that are important for neuronal function, and point, translocation and copy number mutations in *RBFOX1* occur in several neurological disorders, including ASDs[89–91]. Moreover, HITS-CLIP analysis has determined that RBFOX targets splicing events for multiple autism-susceptibility genes in mice, including *Shank3* (SH3 and multiple ankyrin repeat domains 3) and *Tsc2* (tuberous sclerosis 2)[92].

Therapies to modulate RNA mis-splicing

The prevalence of *cis*-, and *trans*-acting splicing mutations and dysregulation as the underlying cause of an array of diseases has led to the development of several therapeutic approaches that are currently in clinical trials[93]. Here, we review the two main strategies that have been pursued — ASOs and small molecule compounds — for three diseases. ASOs are designed either to recognize specific RNA splicing regulatory elements and modulate splicing or to bind nascent transcripts and promote RNase H-mediated degradation in the nucleus. Small molecules have been developed that target splicing factors to modulate their activities or RNA sequences and/or structures (such as hairpins or G-quadruplexes) in an effort to block the abnormal recruitment of splicing factors to mutant sequences.

Antisense oligonucleotides. Duchenne muscular dystrophy (DMD) is a progressive muscle disease that affects ~1 in 3,500 newborn males. It is caused by mutations, often deletions, in the largest annotated human gene (2.4 Mb, 79 exons), *DMD*, which encodes dystrophin[94]. This protein is a key factor in muscle maintenance because it provides an essential link between the dystroglycan complex within the muscle cytoplasmic membrane, or sarcolemma, and the intracellular actin network. Thus, loss of dystrophin results in continuous cycles of myofibre necrosis, satellite cell activation and muscle regeneration, ultimately leading to premature muscle wasting and death. *DMD* mutations are often multiexon deletions that cause frameshifts, and a common deletion results in a frameshift at exon 51. However, the reading frame can be restored by skipping of exon 51, mediated by ASOs that target an exon 51 ESE. This leads to the production of internally deleted DMD proteins that retain partial function (FIG. 4a). To induce exon 51 skipping, two ASO drug candidates, drisapersen (a 2′*O*-methylphosphorothioate ASO (2′OMePS)) and eteplirsen (a phosphorodiamidate morpholino oligomer, (PMO)), have progressed through clinical trials, although inflammatory responses to drisapersen have been noted[95]. A similar strategy has been used to reduce abnormal progerin expression and increase lifespan in a mouse model of HGPS by dual targeting of *LMNA* exon 10 and cryptic exon 11 5′ splice sites with vivo-morpholino ASOs[96].

Recent studies have also supported the efficacy of ASOs for treating SMA (also known as proximal/5q SMA), an autosomal recessive neuromuscular disorder that is characterized by progressive degeneration of spinal cord anterior horn α-motor neurons[97]. SMA is the leading genetic cause of infant mortality (1 in ~10,000 live births) and is clinically subdivided by age-of-onset and severity. It is caused by loss-of-function mutations and/or deletions in the survival of motor neuron 1 (*SMN1*) gene, which encodes the SMN protein required for the assembly of Sm proteins onto snRNAs to form functional snRNPs[98]. A paralogous gene, *SMN2*, also encodes SMN. However, it varies in sequence from

Cryptic splice sites
Splice sites that are not normally recognized by the spliceosome but can be activated either by mutations in *cis*-acting elements or *trans*-acting factors.

Splicing regulatory elements
RNA sequence motifs in either exons or introns that modulate splicing primarily by binding *trans*-acting splicing regulatory factors.

Morpholino
An antisense oligomer with standard nucleic acid bases but instead of deoxyribose contains a six-member morpholine ring linked with phosphorodiamidate (PMO). PMOs function by steric blocking and *vivo*-morpholinos, composed of a morpholino oligomer covalently attached to an octa-guanidine dendrimer, are optimized for *in vivo* delivery.

SMN1 by a C>T transition in exon 7, which abrogates ESE splicing mediated by the splicing factor SRSF1 (REF. 99), thus promoting skipping of this exon and resulting in an unstable SMN isoform, denoted SMNΔ7 protein that is expressed at low levels (FIG. 4b). To activate SMN2 exon 7 splicing, a 2′-*O*-methoxyethyl (MOE) ASO (ASO-10-27), which blocks an ISS in *SMN2* intron 7 (REF. 100), has been used to increase SMN levels in type 1 (severe) SMA infants and children following intrathecal injection and is currently being tested in Phase III clinical trials[97]. Interestingly, systemic administration of ASO-10-27 in neonates in a mouse model of SMA is effective at rescuing the mutant phenotype, suggesting that SMA is also a peripheral tissue splicing disease[101].

For DMD and SMA, mutations result in reduced levels of the encoded proteins. By contrast, another class of diseases, the microsatellite expansion disorders, are associated with unusual RNA structures that alter splicing indirectly (FIGS 3c, 4c). Microsatellites are tandem repeats of 2–10 base pairs in length that comprise ~3% of the human genome and are prone to instability due to DNA replication, recombination and repair errors[102]. Although microsatellites undergo both expansions and contractions, expansions cause >30 hereditary diseases. When these repeats occur in non-coding regions, such as untranslated regions (UTRs) and introns, and expand beyond a particular threshold length, they gain a dominant-negative function at the RNA level by sequestering

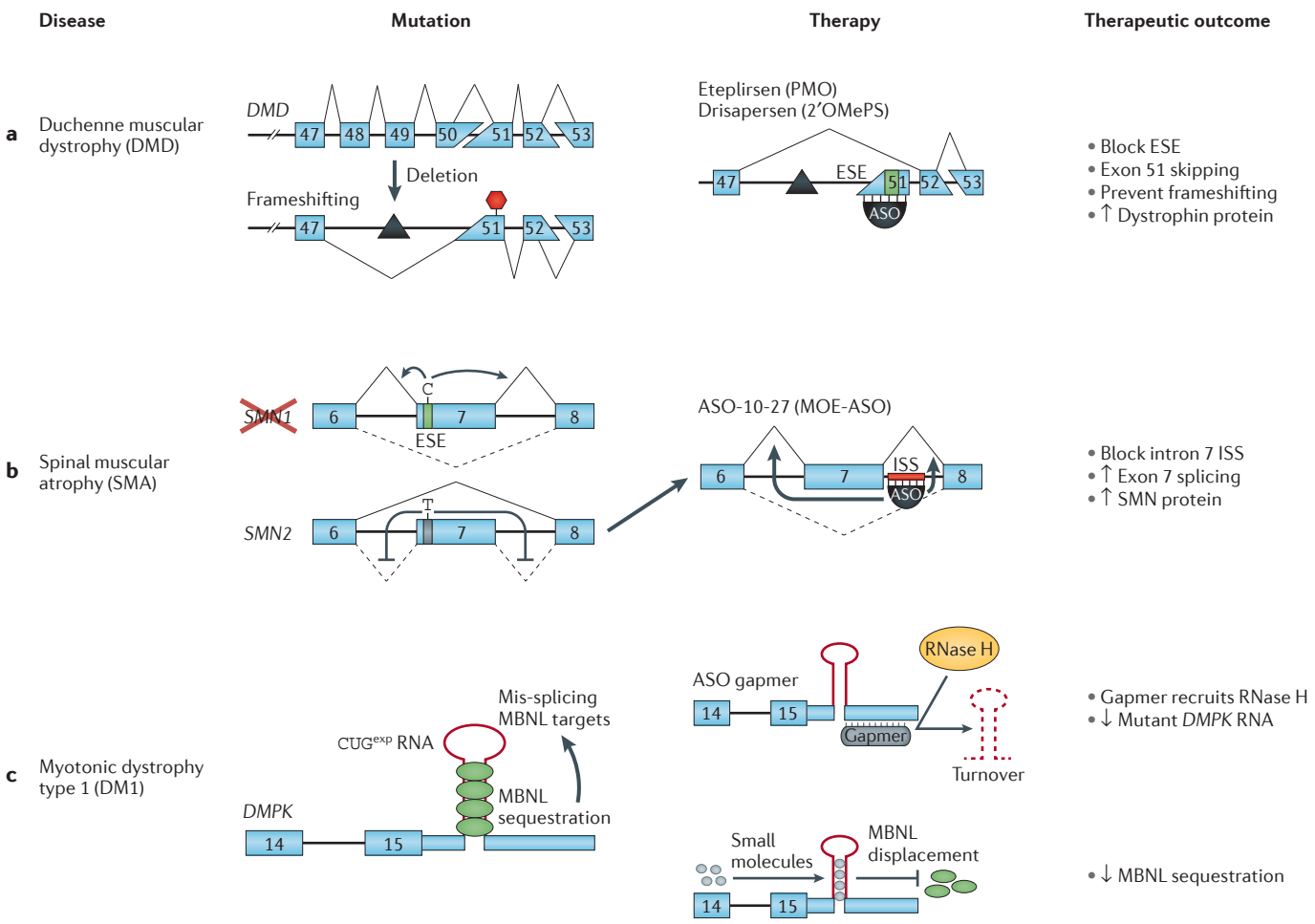


Figure 4 | **Therapeutic strategies.** Examples of therapies based on antisense oligonucleotide (ASO) and small molecule approaches. **a** | Duchenne muscular dystrophy is often caused by chromosomal deletions (black triangle) that remove exons 48–50, resulting in a frameshift (blue rectangles, exons with intact codons; trapezoids, exons with incomplete codons) and loss of dystrophin protein. The red hexagon indicates the premature stop codon resulting from frameshifted exon 51. To prevent frameshifting, both phosphorodiamidate morpholino oligomer (PMO) and 2′OMePS (2′*O*-methyl-phosphorothioate) ASOs (black semicircle) block an exon 51 exonic splicing enhancer (ESE; green rectangle) and shift splicing to the in-frame exon 52. **b** | In spinal muscular atrophy, survival of motor neuron 1 (*SMN1*), which produces the majority of SMN protein, is either deleted or inactivated by mutations, and the paralogous *SMN2* expresses low levels of SMN due to a C>T transition (grey box) that suppresses exon 7 splicing. ASO-10-27 targets an intronic splicing silencer (ISS; red bar) and enhances exon 7 splicing to produce stable SMN protein. **c** | In myotonic dystrophy type 1, CUG expansion (CUG^{exp}) RNA (red hairpin) binds muscleblind-like (MBNL) proteins (green ovals) and causes mis-splicing of MBNL RNA targets. Mutant MBNL–RNA complexes accumulate in the nucleus, and so ASO gapmers preferentially target mutant RNAs for degradation (dotted red line). Alternatively, small molecule compounds bind to mutant CUG^{exp} RNA, displace MBNL and rescue abnormal splicing. *DMPK*, DM protein kinase.

splicing factors, which results in the mis-splicing of hundreds to thousands of RNAs[103,104]. A prominent case is myotonic dystrophy (dystrophia myotonica; DM), which is the most common adult-onset muscular dystrophy and is associated with either a CTG expansion (CTG^{exp}) in the DM protein kinase (*DMPK*) 3′ UTR (for DM1 disease) or an intronic $CCTG^{exp}$ in CCHC-type zinc finger, nucleic acid binding protein (*CNBP*; for DM2 disease)[73]. Following transcription, these non-coding repeats fold into stable hairpin structures that sequester the muscleblind-like (MBNL) proteins while also triggering CELF1 hyperphosphorylation. MBNL and CELF proteins are alternative splicing factors that act antagonistically during development[105], and MBNL loss-of-function due to sequestration by expansion RNAs activates fetal splicing patterns in adult tissues, leading to the characteristic pathophysiology of DM. Although CAG25 morpholinos have been used to displace MBNL proteins from CUG^{exp} RNAs, disperse RNA foci and reverse mis-regulated RNA splicing in a mouse transgenic model of DM[106], systemic uptake requires coupling to cell-penetrating peptides, such as peptide-linked morpholinos (PPMOs)[107]. Another approach, currently being tested in Phase II clinical trials, is the use of ASO gapmers, which are composed of arms with MOE modifications for stability together with a central gap of ten unmodified nucleotides that are susceptible to RNase H-mediated cleavage. These gapmers are designed to target sequences outside the CUG expansion region in the *DMPK* transcript[108]. This is an effective strategy, as mutant allele transcripts are preferentially targeted because RNase H is localized in the nucleus (cytoplasmic activity is confined to mitochondria), and mutant *DMPK* mRNAs accumulate in nuclear RNA foci whereas normal mRNAs are efficiently transported into the cytoplasm.

Small molecules. A variety of small molecule strategies have been reported that target disease-relevant mis-splicing. For DMD, the ASOs that promote DMD exon 51 skipping result in low levels of dystrophin production in humans (~2–16% of normal levels) so small molecule screens have been used to identify drug candidates that increase ASO-induced skipping[109]. One example is dantrolene, which modulates ryanodine receptor activity and is currently used to treat malignant hyperthermia and muscle spasticity. In one study, dantrolene administered together with suboptimal ASO dose was found to increase exon 51 skipping ~10-fold compared with the vehicle (dimethyl sulfoxide) dose in DMD myotubes (patient fibroblasts that have been re-reprogrammed following MYOD1 expression)[109].

For SMA, a number of compounds that increase SMN protein levels have been identified by high-throughput screens[97,110]. Analogous to ASO splice-switching strategies, a recent study uncovered potential drug splicing modifiers that enhance *SMN2* exon 7 inclusion[111]. Using an SMN2 minigene reporter cell-based assay, the study found that treatment of SMA type 1 fibroblasts and induced pluripotent stem cell-derived motor neurons with several compounds (SMNC1, SMNC2 and SMNC3) resulted in increased full-length SMN protein levels. Moreover, RNA-seq analysis demonstrated that these compounds are fairly selective and do not cause widespread transcriptome changes[111]. Importantly, these compounds increased SMN protein levels and improved motor function in a mouse model of severe SMA[111]. A similar high-throughput screen of NSC34 motor neurons expressing a SMN2 splicing reporter was used to identify other compounds (NVS-SM1, NVS-SM2, NVS-SM3 and NVS-SM4) that can increase SMN protein levels. One of these, NVS-SM1, achieved a dose-dependent elevation of SMN protein levels and increased the lifespan of SMNΔ7 mice — which are deleted for their single endogenous *Smn* gene but express SMNΔ7 from human *SMN2* transgenes — from ~15 to >35 days[112]. The structurally similar NVS-SM2 acts by sequence-dependent stabilization of U1 snRNP bound to the *SMN2* exon 7 5′ss.

A wide array of small molecule compounds also block the toxic effects of non-coding microsatellite expansion RNAs. For DM, high-throughput screens as well as more targeted screens have identified a number of compounds that block MBNL1 sequestration and rescue mis-splicing (FIG. 4c). These include a substituted naphthyridine that interacts with UU loops in CUG^{exp} RNA (DM1)[113], a kanamycin A derivative (multivalent K-alkyne) that binds $CCUG^{exp}$ RNA (DM2)[114], the antifungal pentamidine[115], the natural antimicrobial lomofungin[116] and ligand 1 (REF. 117).

Conclusions and perspectives

Human gene structure necessitates an intricate regulatory system to generate the proper set of processed RNA products that are required by the vast assortment of developmental and adult cell types. Hereditary and somatic mutations, which underlie a wide range of diseases from retinal and developmental disorders to cancer, have been documented in both the conserved protein and RNA components of the core spliceosome. Recent studies have highlighted key roles for the U4/U6.U5 tri-snRNP in spliceosomal dynamics, particularly in some specialized cells. Although the limited number of snRNA mutations linked to disease is striking, this probably reflects the essential roles of these RNAs during embryonic development. Another emerging theme in splicing dysregulation is the importance of mutations in LC regions of some splicing factors, including TDP-43 and FUS. Although these mutations often result in aberrant aggregation of these proteins and possible loss- or gain-of-function effects on multiple pathways, mutant LC regions may also interfere with the co-transcriptional dynamics of RNA–RNP complexes that are required to modulate normal splicing patterns.

Owing to the increasing number of annotated ncRNAs in the human genome and the fact that most (>90%) disease-associated SNPs lie outside of protein-coding regions, ncRNAs have also been proposed as regulatory factors that affect both splicing regulation and disease[118]. LncRNAs, which are generally inefficiently spliced and expressed at lower levels than coding RNAs, could influence splicing through interactions with splicing factors (to act as molecular scaffolds and/or sponges) or other RNAs (to repress or activate RNA-based activities)[119,120]. Indeed, lncRNAs have been implicated in schizophrenia, a chronic and disabling brain disorder for which

Human transcriptome
All of the RNAs transcribed from the human genome.

genome-wide association studies have identified >100 independent disease-associated loci[121]. For example, the nuclear-retained myocardial infarction associated transcript (MIAT; also known as Gomafu) lncRNA binds several splicing factors *in vitro*, including quaking (QKI), which has been implicated in schizophrenia; MIAT is downregulated in the brains of those with schizophrenia, and the *MIAT* gene is located in a locus linked to schizophrenia (22q12.1)[11]. CircRNAs, which are widely expressed mammalian ncRNAs and are often generated by head-to-tail splicing of exons, have also been proposed as important regulators of gene expression, possibly by competing with linear splicing[122]. Interestingly, several splicing factors that have been implicated in disease, including QKI and MBNL, regulate circRNA biogenesis[10,123]. Additional studies should be designed to determine whether these ncRNAs have direct functional roles in splicing and disease.

The future will probably reveal many surprises in splicing regulation and mis-splicing in disease. New examples of RNA splicing errors should emerge as a result of our enhanced understanding of the human transcriptome, owing to improvements in single-cell RNA-seq and single-molecule RNA sequencing technologies. Complementary machine-learning approaches that focus on analysing sequence variants in disease should accelerate our understanding of the 'splicing code' (REF. 124). Furthermore, functional network analysis of the spliceosomal machinery using knockdown and other approaches will lead to new insights into how variations in core spliceosomal components influence differentiation, development and disease[125]. Finally, clinical trials using splicing modulatory strategies have produced some encouraging results for several diseases, and these approaches should be applicable to additional disorders caused by mis-splicing.

1. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
2. Kim, M. S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
3. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
4. Treutlein, B., Gokce, O., Quake, S. R. & Sudhof, T. C. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proc. Natl Acad. Sci. USA* **111**, E1291–E1299 (2014). **Long-read sequencing of full-length neurexin mRNAs from pre-frontal cortex is performed to determine the extent of alternative splicing and provide evidence for thousands of neurexin isoforms.**
5. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
6. Gerstein, M. B. *et al.* Comparative analysis of the transcriptome across distant species. *Nature* **512**, 445–448 (2014).
7. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
8. Singh, R. K. & Cooper, T. A. Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* **18**, 472–482 (2012).
9. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
10. Ashwal-Fluss, R. *et al.* circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* **56**, 55–66 (2014). **Demonstrates that circRNAs are generated co-transcriptionally, that the circularization process competes with linear splicing and then MBNL functions as an auto-regulatory factor in circRNA biogenesis.**
11. Barry, G. *et al.* The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol. Psychiatry* **19**, 486–494 (2014).
12. Sharp, P. A. Split genes and RNA splicing. *Cell* **77**, 805–815 (1994).
13. Dujardin, G. *et al.* How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell* **54**, 683–690 (2014).
14. Maquat, L. E. *et al.* Processing of human β-globin mRNA precursor to mRNA is defective in three patients with β$^+$-thalassemia. *Proc. Natl Acad. Sci. USA* **77**, 4287–4291 (1980).
15. Spritz, R. A. *et al.* Base substitution in an intervening sequence of a β$^+$-thalassemic human globin gene. *Proc. Natl Acad. Sci. USA* **78**, 2455–2459 (1981).
16. Busslinger, M., Moschonas, N. & Flavell, R. A. β$^+$ thalassemia: aberrant splicing results from a single point mutation in an intron. *Cell* **27**, 289–298 (1981).
17. Takeshima, Y. *et al.* Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular dystrophy cases from one Japanese referral center. *J. Hum. Genet.* **55**, 379–388 (2010).
18. Fletcher, S. *et al.* Antisense suppression of donor splice site mutations in the dystrophin gene transcript. *Mol. Genet. Genom. Med.* **1**, 162–173 (2013).
19. Chu, C. S., Trapnell, B. C., Curristin, S., Cutting, G. R. & Crystal, R. G. Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nat. Genet.* **3**, 151–156 (1993).
20. Tsui, L. C. & Dorfman, R. The cystic fibrosis gene: a molecular genetic perspective. *Cold Spring Harb. Perspect. Med.* **3**, a009472 (2013).
21. Niblock, M. & Gallo, J. M. Tau alternative splicing in familial and sporadic tauopathies. *Biochem. Soc. Trans.* **40**, 677–680 (2012).
22. Gruenbaum, Y. & Medalia, O. Lamins: the structure and protein complexes. *Curr. Opin. Cell Biol.* **32**, 7–12 (2015).
23. Luo, Y. B., Mastaglia, F. L. & Wilton, S. D. Normal and aberrant splicing of *LMNA*. *J. Med. Genet.* **51**, 215–223 (2014).
24. Muchir, A. *et al.* Identification of mutations in the gene encoding lamins A/C in autosomal dominant limb girdle muscular dystrophy with atrioventricular conduction disturbances (LGMD1B). *Hum. Mol. Genet.* **9**, 1453–1459 (2000).
25. Morel, C. F. *et al.* A *LMNA* splicing mutation in two sisters with severe Dunnigan-type familial partial lipodystrophy type 2. *J. Clin. Endocrinol. Metab.* **91**, 2689–2695 (2006).
26. Eriksson, M. *et al.* Recurrent *de novo* point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. *Nature* **423**, 293–298 (2003).
27. De Sandre-Giovannoli, A. *et al.* Lamin a truncation in Hutchinson–Gilford progeria. *Science* **300**, 2055 (2003).
28. Otomo, J. *et al.* Electrophysiological and histopathological characteristics of progressive atrioventricular block accompanied by familial dilated cardiomyopathy caused by a novel mutation of lamin A/C gene. *J. Cardiovasc. Electrophysiol* **16**, 137–145 (2005).
29. Liu, M. M. & Zack, D. J. Alternative splicing and retinal degeneration. *Clin. Genet.* **84**, 142–149 (2013).
30. Tanackovic, G. *et al. PRPF* mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Hum. Mol. Genet.* **20**, 2116–2130 (2011).
31. Linder, B. *et al.* Identification of a PRPF4 loss-of-function variant that abrogates U4/U6.U5 tri-snRNP integration and is associated with retinitis pigmentosa. *PLoS ONE* **9**, e111754 (2014).
32. Vaclavik, V., Gaillard, M. C., Tiab, L., Schorderet, D. F. & Munier, F. L. Variable phenotypic expressivity in a Swiss family with autosomal dominant retinitis pigmentosa due to a T494M mutation in the *PRPF3* gene. *Mol. Vis.* **16**, 467–475 (2010).
33. Maubaret, C. G. *et al.* Autosomal dominant retinitis pigmentosa with intrafamilial variability and incomplete penetrance in two families carrying mutations in *PRPF8*. *Invest. Ophthalmol. Vis. Sci.* **52**, 9304–9309 (2011).
34. Venturini, G., Rose, A. M., Shah, A. Z., Bhattacharya, S. S. & Rivolta, C. *CNOT3* is a modifier of *PRPF31* mutations in retinitis pigmentosa with incomplete penetrance. *PLoS Genet.* **8**, e1003040 (2012).
35. Comitato, A. *et al.* Mutations in splicing factor *PRPF3*, causing retinal degeneration, form detrimental aggregates in photoreceptor cells. *Hum. Mol. Genet.* **16**, 1699–1707 (2007).
36. Mozaffari-Jovin, S. *et al.* Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* **341**, 80–84 (2013).
37. Chen, X. *et al. PRPF4* mutations cause autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* **23**, 2926–2939 (2014).
38. Kevany, B. M. & Palczewski, K. Phagocytosis of retinal rod and cone photoreceptors. *Physiology (Bethesda)* **25**, 8–15 (2010).
39. Farkas, M. H. *et al.* Mutations in pre-mRNA processing factors 3, 8, and 31 cause dysfunction of the retinal pigment epithelium. *Am. J. Pathol.* **184**, 2641–2652 (2014).
40. David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* **24**, 2343–2364 (2010).
41. Adler, A. S. *et al.* An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth. *Genes Dev.* **28**, 1068–1084 (2014). **This study identifies the tri-snRNP complex protein PRPF6 as an oncogenic driver of colon cancer proliferation by promoting selective splicing of growth regulatory gene transcripts.**
42. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
43. Kurtovic-Kozaric, A. *et al. PRPF8* defects cause missplicing in myeloid malignancies. *Leukemia* **29**, 126–136 (2014).
44. Yoshida, K. & Ogawa, S. Splicing factor mutations and cancer. *Wiley Interdiscip. Rev. RNA* **5**, 445–459 (2014).
45. Malcovati, L. *et al. SF3B1* mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood* **126**, 233–241 (2015).
46. Shirai, C. L. *et al.* Mutant U2AF1 expression alters hematopoiesis and pre-mRNA splicing *in vivo*. *Cancer Cell* **27**, 631–643 (2015).
47. Okeyo-Owuor, T. *et al. U2AF1* mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia* **29**, 909–917 (2015).
48. Ilagan, J. O. *et al. U2AF1* mutations alter splice site recognition in hematological malignancies. *Genome Res.* **25**, 14–26 (2015).
49. Zhang, J. *et al.* Disease-associated mutation in *SRSF2* misregulates splicing by altering RNA-binding affinities. *Proc. Natl Acad. Sci. USA* **112**, E4726–E4734 (2015).

50. Kim, E. *et al. SRSF2* mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* **27**, 617–630 (2015).
51. Komeno, Y. *et al.* SRSF2 is essential for hematopoiesis, and its myelodysplastic syndrome-related mutations dysregulate alternative pre-mRNA splicing. *Mol. Cell. Biol.* **35**, 3071–3082 (2015).
52. Hsu, T. Y. *et al.* The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* **525**, 384–388 (2015).
53. Bonnal, S., Vigevani, L. & Valcarcel, J. The spliceosome as a target of novel antitumour drugs. *Nat. Rev. Drug Discov.* **11**, 847–859 (2012).
54. Abdel-Salam, G. M. *et al.* A homozygous mutation in *RNU4ATAC* as a cause of microcephalic osteodysplastic primordial dwarfism type I (MOPD I) with associated pigmentary disorder. *Am. J. Med. Genet. A* **155**, 2885–2896 (2011).
55. Edery, P. *et al.* Association of TALS developmental disorder with defect in minor splicing component *U4atac* snRNA. *Science* **332**, 240–243 (2011).
56. He, H. *et al.* Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* **332**, 238–240 (2011).
57. Jafarifar, F., Dietrich, R. C., Hiznay, J. M. & Padgett, R. A. Biochemical defects in minor spliceosome function in the developmental disorder MOPD I. *RNA* **20**, 1078–1089 (2014).
58. Younis, I. *et al.* Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *eLife* **2**, e00780 (2013).
The authors propose that minor introns function as molecular switches which are regulated by U6atac stability, which is itself controlled by the stress-activated kinase p38 MAPK.
59. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
60. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
61. Konig, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
62. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
63. Kwon, S. C. *et al.* The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1122–1130 (2013).
64. Moore, M. J. *et al.* Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat. Protoc.* **9**, 263–293 (2014).
65. Daneholt, B. Assembly and transport of a premessenger RNP particle. *Proc. Natl Acad. Sci. USA* **98**, 7012–7017 (2001).
66. King, O. D., Gitler, A. D. & Shorter, J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.* **1462**, 61–80 (2012).
67. Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
Reports that hereditary mutations in the prion-like domains of HNRNPA1 and HNRNPA2B1 cause multisystem proteinopathy and ALS and drive the formation of cytoplasmic inclusions.
68. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
69. Han, T. W. *et al.* Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).
70. Marangi, G. & Traynor, B. J. Genetic causes of amyotrophic lateral sclerosis: new genetic analysis methodologies entailing new opportunities and challenges. *Brain Res.* **1607**, 75–93 (2014).
71. Renton, A. E., Chio, A. & Traynor, B. J. State of play in amyotrophic lateral sclerosis genetics. *Nat. Neurosci.* **17**, 17–23 (2014).
72. Ling, S. C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416–438 (2013).
73. Goodwin, M. & Swanson, M. S. RNA-binding protein misregulation in microsatellite expansion disorders. *Adv. Exp. Med. Biol.* **825**, 353–388 (2014).
74. Prudencio, M. *et al.* Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat. Neurosci.* **18**, 1175–1182 (2015).
75. Lee, Y. B. *et al.* Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep.* **5**, 1178–1186 (2013).
76. Buratti, E. & Baralle, F. E. TDP-43: gumming up neurons through protein–protein and protein–RNA interactions. *Trends Biochem. Sci.* **37**, 237–247 (2012).
77. Arnold, E. S. *et al.* ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. *Proc. Natl Acad. Sci. USA* **110**, E736–E745 (2013).
Demonstration that controlled expression of ALS-associated TDP-43 mutations in a mouse model causes splicing dysregulation of specific targets and progressive motor neuron loss in the absence of TDP-43 aggregation or nuclear depletion.
78. Lagier-Tourenne, C. *et al.* Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* **15**, 1488–1497 (2012).
79. Nakaya, T., Alexiou, P., Maragkakis, M., Chang, A. & Mourelatos, Z. FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *RNA* **19**, 498–509 (2013).
80. Bai, B. *et al.* U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **110**, 16562–16567 (2013).
81. Li, Y. I., Sanchez-Pulido, L., Haerty, W. & Ponting, C. P. RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.* **25**, 1–13 (2015).
82. Sibley, C. R. *et al.* Recursive splicing in long vertebrate genes. *Nature* **521**, 371–375 (2015).
83. Duff, M. O. *et al.* Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature* **521**, 376–379 (2015).
84. Polymenidou, M. *et al.* Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.* **14**, 459–468 (2011).
85. Tollervey, J. R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* **14**, 452–458 (2011).
86. Ling, J. P., Pletnikova, O., Troncoso, J. C. & Wong, P. C. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* **349**, 650–655 (2015).
Shows that TDP-43 depletion results in splicing of non-conserved cryptic exons that are often located in distal regions of long introns and that loss of TDP-43-mediated cryptic exon splicing repression is a potential pathogenic factor in ALS and FTD.
87. Nelson, S. B. & Valakh, V. Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron* **87**, 684–698 (2015).
88. Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014).
89. Gehman, L. T. *et al.* The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* **43**, 706–711 (2011).
90. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
91. Martin, C. L. *et al.* Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **144**, 869–876 (2007).
92. Weyn-Vanhentenryck, S. M. *et al.* HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* **6**, 1139–1152 (2014).
93. Kole, R., Krainer, A. R. & Altman, S. RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat. Rev. Drug Discov.* **11**, 125–140 (2012).
94. Ruegg, U. T. Pharmacological prospects in the treatment of Duchenne muscular dystrophy. *Curr. Opin. Neurol.* **26**, 577–584 (2013).
95. Kole, R. & Krieg, A. M. Exon skipping therapy for Duchenne muscular dystrophy. *Adv. Drug Deliv. Rev.* **87**, 104–107 (2015).
96. Osorio, F. G. *et al.* Splicing-directed therapy in a new mouse model of human accelerated aging. *Sci. Transl. Med.* **3**, 106ra107 (2011).
97. Faravelli, I., Nizzardo, M., Comi, G. P. & Corti, S. Spinal muscular atrophy — recent therapeutic advances for an old challenge. *Nat. Rev. Neurol.* **11**, 351–359 (2015).
98. Li, D. K., Tisdale, S., Lotti, F. & Pellizzoni, L. SMN control of RNP assembly: from post-transcriptional gene regulation to motor neuron disease. *Semin. Cell Dev. Biol.* **32**, 22–29 (2014).
99. Cartegni, L., Hastings, M. L., Calarco, J. A., de Stanchina, E. & Krainer, A. R. Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am. J. Hum. Genet.* **78**, 63–77 (2006).
100. Hua, Y., Vickers, T. A., Okunola, H. L., Bennett, C. F. & Krainer, A. R. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects *SMN2* splicing in transgenic mice. *Am. J. Hum. Genet.* **82**, 834–848 (2008).
101. Hua, Y. *et al.* Motor neuron cell-nonautonomous rescue of spinal muscular atrophy phenotypes in mild and severe transgenic mouse models. *Genes Dev.* **29**, 288–297 (2015).
Although SMA is a motor neuron disease, this study uses a combination of splice-switching and decoy oligonucleotides to provide evidence that disease is not specific to motor neurons in a mouse model of SMA.
102. Lopez Castel, A., Cleary, J. D. & Pearson, C. E. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* **11**, 165–170 (2010).
103. Wang, E. T. *et al.* Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150**, 710–724 (2012).
104. Charizanis, K. *et al.* Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* **75**, 437–450 (2012).
105. Wang, E. T. *et al.* Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Res.* **25**, 858–871 (2015).
106. Wheeler, T. M. *et al.* Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science* **325**, 336–339 (2009).
107. Leger, A. J. *et al.* Systemic delivery of a peptide-linked morpholino oligonucleotide neutralizes mutant RNA toxicity in a mouse model of myotonic dystrophy. *Nucleic Acid. Ther.* **23**, 109–117 (2013).
108. Wheeler, T. M. *et al.* Targeting nuclear RNA for *in vivo* correction of myotonic dystrophy. *Nature* **488**, 111–115 (2012).
109. Kendall, G. C. *et al.* Dantrolene enhances antisense-mediated exon skipping in human and mouse models of Duchenne muscular dystrophy. *Sci. Transl. Med.* **4**, 164ra160 (2012).
110. Cherry, J. J. *et al.* Enhancement of SMN protein levels in a mouse model of spinal muscular atrophy using novel drug-like compounds. *EMBO Mol. Med.* **5**, 1035–1050 (2013).
111. Naryshkin, N. A. *et al. SMN2* splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science* **345**, 688–693 (2014).
A minigene splicing reporter combined with chemical screening and optimization were used to identify small molecule compounds that activate SMN2 exon 7 splicing and increase SMN levels.
112. Palacino, J. et al. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.* **11**, 511–517 (2015); erratum **11**, 741 (2015).
113. Childs-Disney, J. L. *et al.* Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. *Nat. Commun.* **4**, 2044 (2013).
114. Childs-Disney, J. L. *et al.* Structure of the myotonic dystrophy type 2 RNA and designed small molecules that reduce toxicity. *ACS Chem. Biol.* **9**, 538–550 (2014).
115. Warf, M. B., Nakamori, M., Matthys, C. M., Thornton, C. A. & Berglund, J. A. Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl Acad. Sci. USA* **106**, 18551–18556 (2009).
116. Hoskins, J. W. *et al.* Lomofungin and dilomofungin: inhibitors of MBNL1–CUG RNA binding with distinct cellular effects. *Nucleic Acids Res.* **42**, 6591–6602 (2014).

117. Jahromi, A. H. *et al.* A novel CUG$^{exp.}$ MBNL1 inhibitor with therapeutic potential for myotonic dystrophy type 1. *ACS Chem. Biol.* **8**, 1037–1043 (2013).
118. Niland, C. N., Merry, C. R. & Khalil, A. M. Emerging roles for long non-coding RNAs in cancer and neurological disorders. *Front. Genet.* **3**, 25 (2012).
119. Yang, L., Froberg, J. E. & Lee, J. T. Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem. Sci.* **39**, 35–43 (2014).
120. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).
121. Gratten, J., Wray, N. R., Keller, M. C. & Visscher, P. M. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat. Neurosci.* **17**, 782–790 (2014).
122. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
123. Conn, S. J. *et al.* The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125–1134 (2015).
124. Xiong, H. Y. *et al.* The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
Comprehensive study showing the utility of a machine learning technique that scores sequence variants for splicing impact and reveals thousands of disease-associated mutations.
125. Papasaikas, P., Tejedor, J. R., Vigevani, L. & Valcarcel, J. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol. Cell* **57**, 7–22 (2015).
126. Turunen, J. J., Niemela, E. H., Verma, B. & Frilander, M. J. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* **4**, 61–76 (2013).
127. Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–1786 (2014).
Although intron retention is the most common form of splicing regulation in plants, this study uncovers widespread intron retention in mammals and proposes intron retention as a regulatory mechanism to control transcript levels by NMD and nuclear sequestration and turnover.
128. Ibrahim, E. C. *et al.* Weak definition of *IKBKAP* exon 20 leads to aberrant splicing in familial dysautonomia. *Hum. Mutat.* **28**, 41–53 (2007).
129. Muntoni, F., Torelli, S. & Ferlini, A. Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol.* **2**, 731–740 (2003).
130. Disset, A. *et al.* An exon skipping-associated nonsense mutation in the dystrophin gene uncovers a complex interplay between multiple antagonistic splicing elements. *Hum. Mol. Genet.* **15**, 999–1013 (2006).
131. Samaranch, L. *et al.* PINK1-linked parkinsonism is associated with Lewy body pathology. *Brain* **133**, 1128–1142 (2010).
132. Iovino, M. *et al.* The novel *MAPT* mutation K298E: mechanisms of mutant tau toxicity, brain pathology and tau expression in induced fibroblast-derived neurons. *Acta Neuropathol.* **127**, 283–295 (2014).
133. Korvatska, O. *et al.* Altered splicing of *ATP6AP2* causes X-linked parkinsonism with spasticity (XPDS). *Hum. Mol. Genet.* **22**, 3259–3268 (2013).
134. Tanackovic, G. *et al.* A missense mutation in *PRPF6* causes impairment of pre-mRNA splicing and autosomal-dominant retinitis pigmentosa. *Am. J. Hum. Genet.* **88**, 643–649 (2011).
135. Cvackova, Z., Mateju, D. & Stanek, D. Retinitis pigmentosa mutations of *SNRNP200* enhance cryptic splice-site recognition. *Hum. Mutat.* **35**, 308–317 (2014).
136. 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).
137. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).
138. Sun, S. *et al.* ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. *Nat. Commun.* **6**, 6171 (2015).
139. Guo, W. *et al. RBM20*, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat. Med.* **18**, 766–773 (2012).
140. Vieira, N. M. *et al.* A defect in the RNA-processing protein HNRPDL causes limb-girdle muscular dystrophy 1G (LGMD1G). *Hum. Mol. Genet.* **23**, 4103–4110 (2014).
141. Bartoletti-Stella, A. *et al.* Messenger RNA processing is altered in autosomal dominant leukodystrophy. *Hum. Mol. Genet.* **24**, 2746–2756 (2015).

Acknowledgements
The authors regret that many important studies were not cited owing to space limitations. Work in the authors' laboratories is funded by grants to M.S.S. from the US National Institutes of Health (NIH AR046799, NS058901), the Muscular Dystrophy Association (MDA276063), the W.M. Keck Foundation (F013635) and the Marigold Foundation. M.M.S. is the recipient of an NIH pre-doctoral traineeship (NIAMS T32 AR7605-15).

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
OMIM: http://www.omim.org/
ENCODE: http://www.gencodegenes.org/
UniProtKB: http://www.uniprot.org/
RetNet: https://sph.uth.edu/retnet/
MISO Database: https://miso.readthedocs.org/en/fastmiso/annotation.html
Splice Rack Database: http://katahdin.mssm.edu/splice/splice_matrix_poster.cgi?database=spliceNew2
U12 Database: http://genome.crg.es/cgi-bin/u12db/u12db.cgi
Pictogram: http://genes.mit.edu/pictogram.html
ALL LINKS ARE ACTIVE IN THE ONLINE PDF