The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches

Elisabeth Daguenet¹²†, Gwendal Dujardin¹²† & Juan Valcárcel¹²³,*

Abstract

Removal of introns from pre-mRNA precursors (pre-mRNA splicing) is a necessary step for the expression of most genes in multicellular organisms, and alternative patterns of intron removal diversify and regulate the output of genomic information. Mutation or natural variation in pre-mRNA sequences, as well as in spliceosomal components and regulatory factors, has been implicated in the etiology and progression of numerous pathologies. These range from monogenic to multifactorial genetic diseases, including metabolic syndromes, muscular dystrophies, neurodegenerative and cardiovascular diseases, and cancer. Understanding the molecular mechanisms associated with splicing-related pathologies can provide key insights into the normal function and physiological context of the complex splicing machinery and establish sound basis for novel therapeutic approaches.

Keywords alternative splicing; disease; mutation; RNA; spliceosome

DOI 10.15252/embr.201541116 | Received 30 July 2015 | Revised 9 October 2015 | Accepted 12 October 2015 | Published online 13 November 2015

See the Glossary for abbreviations used in this article.

Introduction

The central dogma of Molecular Biology emerged originally as a collinear view of gene expression, in which the information flows from DNA to protein through messenger RNA (mRNA) molecules. The last decades of research have considerably expanded this paradigm by showing the multiplicity of transcripts that can be generated from a single DNA locus through the use of alternative promoters, termination sites, and through alternative splicing of introns, intervening sequences present in primary transcripts that need to be removed to generate translatable mRNAs. Furthermore, intertwining links between transcriptional and posttranscriptional steps in the gene expression pathway, both in the nucleus and in the cytoplasm, not only facilitate coupling between these processes but also expand their regulatory possibilities, particularly in higher eukaryotes [1].

Pre-mRNA splicing requires precise recognition of cis-acting sequences on the pre-mRNA by spliceosomal components and additional RNA-binding factors, and involves a vast network of RNA–RNA, RNA–protein, and protein–protein interactions. The realization that at least 95% of human genes produce multiple spliced RNA species via alternative exon usage has revealed the prevalence of this additional layer of gene expression regulation [2]. Indeed, alternative splicing (AS) enables individual genes to increase their coding capability and to generate a set of structurally and functionally distinct protein isoforms. The main types of AS are “cassette” exon skipping, alternative 5’ and 3’ splice site selection, alternatively retained introns, and mutually exclusive exons. Interestingly, the frequency of AS varies with species complexity and cell type, during development or upon cellular differentiation, thereby participating in the fine tuning of a gene signature both temporally and spatially [3,4].

Mis-regulation of splicing has been long known to be related to an increasing number of human pathologies, including genetic diseases, neurodegenerative disorders, and cancer. Alterations in pre-mRNA splicing can either act as drivers of disease etiology or act as modifiers that sensitize individuals to disease susceptibility and severity. Recent excellent reviews have covered multiple aspects of this topic [5–11]. In this review, we provide a general overview of the function of the spliceosome and the combinatorial rules governing the splicing code. Our focus will be on splicing aberrations in various pathological contexts and how understanding the underlying mechanistic principles can set the stage for the development of novel therapeutic approaches and at the same time shed light on the function and physiology of splicing itself.

Basics of the pre-mRNA-splicing process

Successful completion of the splicing reaction and deployment of its physiological function require both fidelity and flexibility. First, the discrimination between correct and incorrect splice sites is achieved...
through systematic, multistage proofreading of the sequences by different factors. Second, splicing commitment is subject to an elaborated and dynamic crosstalk between splicing regulatory factors in order to enforce or to repress splice site selection (for recent reviews: [12–16]).

**Exon definition & the spliceosome assembly pathway**

Intron removal is orchestrated by the multi-megadalton macromolecular ribonucleoprotein complex known as the spliceosome, which is composed of five small nuclear ribonucleoproteins (U1, U2, U4/U6, U5 snRNP) and more than 200 snRNP- and non-snRNP-associated proteins [14]. The definition of an intron relies on four consensus elements: the exon/intron junctions at the 5′ and 3′ end of the intron—the 5′ and 3′ splice sites (SS)—, the branch point sequence (BPS) located upstream of the 3′ SS, and the polypyrimidine tract located between the BPS and the 3′ SS (Fig 1A). The BPS adenosine plays a crucial role in splicing catalysis, by forming a 2′–5′ phosphodiester bond with the 5′ end of the intron after the first step of the reaction. The 5′ SS and the region surrounding the BPS are recognized through base-pairing interactions with U1 and U2 snRNAs, respectively. Additional regulatory sequences within introns and exons contribute to splice site recognition by the core splicing machinery (Fig 1B and see below). In vertebrates, given the longer length of introns compared with exons, splice sites flanking an internal exon communicate with each other to help in initial exon definition [17], and subsequently engage in interactions across the intron to allow intron removal and exon inclusion (Fig 1C). Modulation of splice site pairing...
during exon and intron definition can be the target of regulators [18–20] (Fig 1C).

Assembly of spliceosomal complexes onto pre-mRNA follows a stepwise choreography and is supported by at least eight DExD/H-type RNA-dependent ATPases/helicases whose function is either to remodel snRNP composition or to proofread specific transitions along the assembly cycle. The initial step begins with the recognition of the 5' SS by U1 snRNP and the cooperative binding of the splicing factor 1 (SF1) and of the heterodimer U2AF65/U2AF35 to the BPS region, polypyrimidine tract and 3' SS, respectively, generating complex E (Fig 2). These molecular interactions then trigger the ATP-dependent recruitment of the U2 snRNP to the BPS region through base-pairing interactions between U2 snRNA and nucleotides flanking the branch point adenosine. U2 snRNP assembly is also assisted by U2 snRNP-associated proteins and during exon and intron definition can be the target of regulators [18–20] (Fig 1C).

Assembly of spliceosomal complexes onto pre-mRNA follows a stepwise choreography and is supported by at least eight DExD/H-type RNA-dependent ATPases/helicases whose function is either to remodel snRNP composition or to proofread specific transitions along the assembly cycle. The initial step begins with the recognition of the 5' SS by U1 snRNP and the cooperative binding of the splicing factor 1 (SF1) and of the heterodimer U2AF65/U2AF35 to the BPS region, polypyrimidine tract and 3' SS, respectively, generating complex E (Fig 2). These molecular interactions then trigger the ATP-dependent recruitment of the U2 snRNP to the BPS region through base-pairing interactions that bulge out the BPS adenosine. U2 snRNP assembly is also assisted by U2 snRNP-associated proteins and during exon and intron definition can be the target of regulators [18–20] (Fig 1C).

Assembly of spliceosomal complexes onto pre-mRNA follows a stepwise choreography and is supported by at least eight DExD/H-type RNA-dependent ATPases/helicases whose function is either to remodel snRNP composition or to proofread specific transitions along the assembly cycle. The initial step begins with the recognition of the 5' SS by U1 snRNP and the cooperative binding of the splicing factor 1 (SF1) and of the heterodimer U2AF65/U2AF35 to the BPS region, polypyrimidine tract and 3' SS, respectively, generating complex E (Fig 2). These molecular interactions then trigger the ATP-dependent recruitment of the U2 snRNP to the BPS region through base-pairing interactions that bulge out the BPS adenosine. U2 snRNP assembly is also assisted by U2 snRNP-associated proteins and during exon and intron definition can be the target of regulators [18–20] (Fig 1C).

Assembly of spliceosomal complexes onto pre-mRNA follows a stepwise choreography and is supported by at least eight DExD/H-type RNA-dependent ATPases/helicases whose function is either to remodel snRNP composition or to proofread specific transitions along the assembly cycle. The initial step begins with the recognition of the 5' SS by U1 snRNP and the cooperative binding of the splicing factor 1 (SF1) and of the heterodimer U2AF65/U2AF35 to the BPS region, polypyrimidine tract and 3' SS, respectively, generating complex E (Fig 2). These molecular interactions then trigger the ATP-dependent recruitment of the U2 snRNP to the BPS region through base-pairing interactions that bulge out the BPS adenosine. U2 snRNP assembly is also assisted by U2 snRNP-associated proteins and during exon and intron definition can be the target of regulators [18–20] (Fig 1C).
engaging in RNA–protein interactions with sequences around the BPS region and in protein–protein interactions (e.g., between the SF3B1 protein and U2AF65). Subsequent to complex A formation, the pre-assembled U4-U6-U5 tri-snRNP joins the pre-spliceosome complex to establish complex B. The enzymatic activation of the machinery takes place at this stage through a series of conformational and massive compositional rearrangements (including displacement of U1 and U4 snRNP) to successively form the catalytically active complex (Bact), which upon further conformational rearrangements and changes in protein composition catalyzes the first step of the splicing reaction, leading to the formation of a lariat intermediate containing a 2'–5' phosphodiester bond (complex C). An additional conformational switch leads to the second catalytic step, rendering the spliced product and the intron lariat. Upon release of the products, the lariat intron is linearized and degraded and the snRNPs recycled for another round of assembly and catalysis in other introns. Proteins of the ATP-dependent DEAH/X helicase family (in magenta) are key to promote the multiple conformational transitions, characterized by extensive rearrangements of RNA:RNA interactions involving snRNA:snRNA and snRNA:pre-mRNA contacts.

Figure 2. The spliceosome assembly pathway.
Initial recognition of the 5' SS by U1 snRNP and of the 3' SS region by SF1 (branch point binding protein) and U2AF (complex E) is followed by the ATP-dependent recruitment of U2 snRNP to the branch point region (complex A), concomitant with displacement of SF1. Binding of the U4/5/6 tri-snRNP and the protein-only NineTeen Complex (NTC) leads to formation of complex B, involving also the displacement of proteins present in complex A. Extensive remodeling of complex B, including the destabilization/displacement of U1 and U4 snRNPs, leads to a catalytically active complex (Bact), which upon further conformational rearrangements and changes in protein composition catalyzes the first step of the splicing reaction, leading to the formation of a lariat intermediate containing a 2'–5' phosphodiester bond (complex C). An additional conformational switch leads to the second catalytic step, rendering the spliced product and the intron lariat. Upon release of the products, the lariat intron is linearized and degraded and the snRNPs recycled for another round of assembly and catalysis in other introns. Proteins of the ATP-dependent DEAH/X helicase family (in magenta) are key to promote the multiple conformational transitions, characterized by extensive rearrangements of RNA:RNA interactions involving snRNA:snRNA and snRNA:pre-mRNA contacts.
catalytically active complex B (Bact, B*) and complex C, which host, respectively, the 1st and 2nd trans-esterification reactions of the splicing reaction [14,21] (Fig 2).

**From constitutive to alternative splicing**

The core splicing sequences in higher eukaryotes are often variable and contain too little information to unambiguously define SS. Additional sequences in the pre-mRNA modulate SS recognition and are referred to as exonic or intronic splicing enhancers (ESE or ISE) or silencers (ISS or ESS). These sequence elements are recognized by trans-acting splicing factors that balance splice site selection and alternative splicing decisions (Fig 1B). Trans-acting factors include the serine/arginine-rich domain-containing (SR) protein and heterogeneous nuclear ribonucleoprotein (hnRNP) families, which display cooperative or antagonistic effects on the recruitment of the core splicing machinery, typically at early stages of spliceosome formation [22]. In addition, several tissue-restricted regulators have been identified, including the neuronal-specific determinants neuro-oncological ventral antigen (NOVA), RNA-binding Fox (RBFOX) or muscleblind (MBNL). Two recurrent themes are that the same factors can act as activators or repressors depending on the position of their binding sites relative to the regulated SS and that their precise activity depends on the context of other cognate sites for other regulatory factors with which they can establish cooperative or antagonistic interactions [23]. This leads to a complex interplay of regulatory sequences, positional effects, and trans-acting factor interactions that establish the functional framework of a splicing code [23–25]. In addition, variations in the levels or activity of core splicing factors, even those acting late in the spliceosome assembly pathway, can also modulate SS choice [18,26–30]. Furthermore, an increasing number of studies revealed that splicing regulation is also subjected to complex interaction with the transcription and chromatin machineries. Indeed, changes in the kinetics of RNA polymerase II elongation can markedly affect SS selection by influencing the ability of splicing regulators to bind to nascent mRNAs [31,32]. In addition, histone marks and nucleosome positioning are also key features that participate in splicing reactions by helping the recruitment of splicing regulators and collaborating in exon definition, respectively [13,25,33,34]. Finally, signal transduction cascades represent another regulatory level through the modulation of post-translational modifications of splicing regulators, which may modify their interactions, activities, and localization [23,35].

**Alterations of splicing in pathological conditions**

There is growing evidence from both human genetics and genomewide studies that splicing control can impact a variety of pathologies at three levels (Fig 3): (i) Mutations or genetic variants that affect cis-acting sequences by decreasing the specificity or fidelity of SS selection or activating cryptic SS that are normally not used. These alterations impinge on single genes. (ii) Functional alterations in trans-acting splicing factors, including core spliceosomal components and regulatory factors. Such perturbations potentially modify the expression of multiple RNA targets. (iii) Stoichiometric imbalance of splicing factors following their sequestration in repetitive elements. Such squelching mechanism brings about widespread gene expression changes. While abundant examples of such pathogenic mechanisms exist, it is expected that further combined experimental and computational approaches will greatly expand the repertoire—and possibly the categories of mechanisms—of mutations that determine predisposition, onset and/or progression of pathologies, opening novel opportunities for diagnosis, and translational research.

**Cis-acting mutations: breaking the splicing code**

Genetic variation within splice site and regulatory sequences frequently causes aberrant splicing in human hereditary diseases and cancer. Single nucleotide substitutions affecting the 5’ or the 3’ SS are the most common splicing mutations, resulting either in exon skipping, activation of a cryptic SS, or to a lesser extent in intron retention. Similarly, intronic mutations and exonic variations (e.g., missense, nonsense, or even otherwise silent mutations) can often trigger splicing perturbations through a loss and/or a gain of enhancers/silencers. This is illustrated by the analysis of the mutational landscape of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene related to cystic fibrosis [36], or of the neurofibromatosis gene, where 50% of the disease-causing mutations lead to splicing defects [37]. These alterations can occur in both constitutive and alternative exons and consequently generate aberrant transcripts that miss a constitutive exon or result in changes in the ratio between spliced isoforms (Fig 3A).

According to the Human Gene Mutation Database (HGMD release 2014.4), mutations that disrupt normal splicing have been estimated to account for up to a third of all disease-causing mutations [38–41]. In silico tools have been developed for predicting the penetrance associated with de novo mutations and the molecular consequences for disease formation [42]. These efforts should be complemented with predictions of the impact of frequent intronic regulatory sequences as well as exonic silent mutations (or even missense mutations) that affect splicing outcomes. In fact, synonymous variants have already been demonstrated to contribute to human diseases and be particularly prevalent modulators of exon usage [36,43–45]. Consistent with the observation that cancer-related genes may harbor a greater susceptibility toward aberrant splicing [46], recurrent synonymous mutations in ESE and ESS have been identified for a subset of important oncoproteins, representing an extra mechanism for oncogene activation. More globally, half of synonymous drivers is estimated to alter splicing [44], further emphasizing the importance of splicing in cancer biology.

The MutPred Splice algorithm aims to determine functional relationships between exonic variants and mis-splicing in inherited diseases and cancer [47]. The results indicate that in inherited disease, loss of natural splice sites represents the principal category of splice-altering variants (SAV), whereas ESE loss and/or ESS gain leading to exon skipping is more frequent in cancer [47]. More recently, Xiong, Frey, and colleagues assessed the effects of 650,000 single nucleotide intronic and exonic variants using a machine-learning computational pipeline that uncovered an extensive impact of mutation-associated splicing alterations. Their results estimate that intronic mutations alter splicing nine times more frequently than other common variants and that disease-associated missense exonic mutations are five times more likely to interfere with splicing than non-disease-associated variants [45], further illustrating the potential impact of splicing alterations on human pathologies. This approach led to the identification of splicing alterations with potential roles in autism and provided an explanation for the penetrance of synonymous mutations in colorectal cancer [45].
The following examples illustrate how disease-causing mutations can be tightly linked to multiple aspects of splice site recognition and, in fact, help to illustrate the delicate balance of sequence signals and interactions that tune splice site choice and can potentially inform therapeutic approaches (see also Figs 3 and 4 and section on therapies).

The vast majority of patients with familial dysautonomia (FD) contain a point mutation at the sixth position of intron 20 of the \( IKBKAP \) gene, which encodes the transcription regulator protein IKAP [48]. This position is the last of the intronic nucleotides that establish base-pairing interactions with the 5' end of U1 snRNA as part of the initial step in 5' SS recognition (Fig 1A). Just the lack of this single base pair leads to defects in 5' SS identification and in exon definition such that the whole exon 20 is skipped, generating a mRNA with a premature stop codon and defective expression of IKAP [49]. The critical role of this single base pair became evident...
Pathogenicity of splicing defects

Elisabeth Daguenet et al

Published online: November 13, 2015

EMBO reports

in experimental systems where restoring base-pairing between the mutated 5′ SS and U1 snRNA also restored exon 20 inclusion [50]. Remarkably, the extent of the splicing defect is tissue dependent, being very limited in lymphoblasts but extensive in the brain, explaining the severe brain abnormalities and demyelination-associated symptoms of the disease [51]. The basis for the tissue-specific effect of the mutation remains obscure.

More than 2,500 sequences may function as 5′ SS [52]. It is thus not surprising that small sequence variations can lead to the activation of cryptic splice sites, as dramatically illustrated by Hutchinson–Gilford Progeria, a premature aging syndrome where most affected individuals harbor a single C>T silent mutation in exon 11. The mutation activates a cryptic 5′ SS, leading to an mRNA encoding a dominant negative form of lamin A (progerin) that causes nuclear and genomic instability. The balance between the activities of the SR proteins SRSF1 and SRSF6 determines the level of cryptic site activation [53], which can be modulated using anti-sense oligonucleotides or morpholinos, with therapeutic effects in cell and mouse models of the disease [54,55]. Strikingly, it has been proposed that even the wild-type sequence can be used as a 5′ SS

Figure 4. Summary of therapeutic approaches based upon splicing modulation.

Two main strategies are outlined. Splice-switching antisense oligonucleotides target splice sites or splicing regulatory sequences to prevent the binding of cognate factors to modulate splice site selection. Examples include blocking of an ISS element that promotes exon 7 inclusion in the SMN2 gene, leading to restoring SMN protein expression and motoneuron function in SMA; induction of skipping of a mutation-containing exon in the DMD gene, leading to in-frame deletion of a nonessential part of the dystrophin protein and restoration of muscle function; prevention of sequestration of the MBNL splicing regulator in CUG repeat expansions in DMPK transcripts, leading to restoration of abnormal splicing patterns in DM. Small molecules, some of them targeting core splicing components like SF3B, modulate alternative splicing of cell cycle control genes and display anti-tumoral properties. Other drugs can induce SMN2 exon 7 inclusion and therefore raise hope as oral treatments for SMA.
and that its use increases with age, possibly contributing to physiological aging [54].

Another example of how single nucleotide changes near 5′ SS regions can dramatically affect splicing outcomes and disease is tauopathies associated with mutations in exon 10 of the MicrotubuleAssociated Protein Tau (MAPT) gene. These mutations, found in thirteen families with the autosomal dominant condition frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17), alter the ratio between spliced isoforms, promoting exon inclusion and leading to Tau protein aggregation, which has been linked with personality disturbances, dementia, and motor dysfunction [56]. The mutations are not located at the splice sites, but rather induce the opening of an RNA stem-loop that normally partially sequesters the 5′ SS, preventing full inclusion of exon 10 [56,57]. This in fact provides one of the best-documented examples of how secondary structures in the pre-mRNA can influence splice site recognition.

A now classic example of how sequence variation in exonic sequences can influence exon recognition with profound consequences for human disease is spinal muscular atrophy (SMA). SMA is one of the most frequent genetic diseases, an autosomal recessive neuromuscular disorder characterized by the selective loss of spinal motor neurons, leading to severe skeletal muscle weakness and atrophy. SMA etiology relates to insufficient amount of SMN protein whose function is to chaperone the biogenesis and assembly of snRNPs [12]. SMN insufficiency results from loss-of-function mutations or deletion of the Survival Motor Neuron 1 gene (SMN1) [58]. Despite its high homology to SMN1, the SMN2 gene fails to prevent SMA development due to a synonymous nucleotide difference, C6T, in exon 7, which causes exon 7 skipping and generation of a truncated, unstable, and rapidly degraded version of the SMN protein [59]. Understanding the mechanisms behind the differential effects of C vs. T on exon 7 splicing may indeed be instrumental to offer novel therapeutic approaches because the penetrance of the disease is inversely correlated with the levels of exon 7 inclusion in SMN2, which differ in different patient populations [60]. Extensive analyses of the mechanistic impact of the C→T transition initially revealed the loss of an ESE recognized by the SR protein SR5F1 as well as the gain of an ESS recognized by hnRNPA1 [61,62]. Later work revealed additional contributions of intronic silencers that collectively repress exon 7 [63–65]. SMA can be considered as a complex multifactorial pathology as the absence of SMN implies perturbations of the snRNP repertoire with widespread splicing changes [66] and also correlates with splicing alterations of U12 minor spliceosome-dependent events in mouse and Drosophila SMA models [67–69]. A key and still largely unresolved question is why reduced/unbalanced snRNP production leads to a specific motor neuron defect rather than ubiquitously compromising multiple aspects of RNA processing and gene expression and therefore the function of most cell types.

Trans-acting factor mutations and alterations: breaking the splicing machinery

Perhaps the first pathology identified with a link to components of the splicing machinery was systemic lupus erythematosus (SLE), a condition in which antibodies against nuclear antigens, including almost invariably Sm proteins (present in U1, U2, U4, and U5 snRNPs), induce complex autoimmune responses [70]. Strikingly, mutations in genes encoding protein components of the core splicing machinery and snRNP biogenesis are often associated with distinct pathological conditions and most likely with splicing alterations affecting specific subsets of splicing events [5,40], opening interesting and challenging questions about the requirement of these factors for pre-mRNA splicing in general. Answers to these specificity issues may bring key insights into the physiopathology of these diseases and also inform possible therapeutic routes (Figs 3B and 4).

Mutations affecting snRNP biogenesis

As described above, mutation of the SMN1 gene leads to defects in snRNP biogenesis and altered snRNP characteristics in the prevalent motor neuron disease SMA. Interesting possible molecular links involving snRNP function in SMA and other motor/neurodegenerative disorders have emerged. Amyotrophic lateral sclerosis (ALS) is a common motor/neurodegenerative disease caused by mutations in over 20 genes with different functions [71], including the RNA-binding proteins FUS and TDP-43. FUS is an hnRNP-like protein that interacts with U1 and U2 snRNAs [72–74]. FUS mutant proteins also bind to these snRNAs but are retained in the cytoplasm, causing a reduction in the pool of U1 and U2 snRNPs in the nucleus. Furthermore, FUS interacts with SMN and mutated FUS proteins also appear to alter the cellular localization of SMN, potentially contributing also to disease-related splicing alterations [75,76]. TDP-43, another RNA-binding protein frequently mutated in ALS, interacts with FUS [76] and its dysregulation also affects SMN localization and snRNA abundance [77]. Recent results indicate that TDP-43 represses splicing of cryptic exons and that activation of these exons in TDP-43-deficient embryonic stem cells induces cell death [78]. However, the multifunctional nature of TDP-43, FUS, and other related factors makes it difficult to restrict their pathogenic effects to splicing alterations.

TDP-43 forms insoluble aggregates [79], which is a hallmark of other neurodegenerative diseases as well, such as Alzheimer’s disease (AD) or Parkinson’s disease. The recent sequencing of the protein content of AD aggregates uncovered the accumulation of several U1 snRNP components [80], and immunohistochemical analyses revealed that U1-70K and U1A form tangle-like aggregates in the cytoplasm of AD brain samples, but not in other neurodegenerative diseases. Furthermore, RNA-Seq analyses revealed global splicing defects, with accumulation of unspliced RNAs, in AD samples [80], suggesting a possible role for RNA processing in the etiology of the disease. Interestingly, silencing of U1-70K in HEK293 cells led to an increase in amyloid precursor protein (APP), suggesting a possible role for pre-mRNA splicing in APP metabolism [80]. Taken together, these observations imply that alterations in snRNP accumulation and consequently pre-mRNA splicing can contribute to the etiology of multiple neurodegenerative disorders, perhaps through common mechanisms and targets.

Alterations in U6 snRNA biogenesis have been detected in Clericuzio-type poikiloderma with neutropenia (PN), a rare autosomal recessive skin disease frequently associated with chronic neutropenia and bone marrow abnormalities potentially leading to myelodysplasia and increased risk of leukemic transformation [81]. The disease is associated with mutations in the C16orf57 gene encoding hMps1/Usb1 [82–87]. This 3′→5′ exoribonuclease is involved in correct processing of U6 snRNA, removing a tail of uridine residues, and generating a characteristic 2′-3′-cyclic phosphate that stabilizes the snRNA [81]. Reduced levels of this enzyme in lymphoblasts from PN patients result in higher U6 snRNA degradation. However,
no global perturbation of splicing was observed in these cells [81], which is remarkable considering the critical role that U6 snRNA plays in the two catalytic steps of the splicing process. These results suggest that the disease is either associated with specific discrete effects in splicing of certain target RNAs, or that the enzyme plays roles in the metabolism of other RNA classes.

A common concept underlying these examples is that limiting amounts of particular snRNPs exert discrete molecular effects that result in tissue-specific phenotypic disturbances, suggesting that the levels of snRNPs may be tuned in a cell type-specific manner for achieving physiological gene regulation.

**Mutations in core spliceosomal proteins** Recent whole-genome and whole-exome sequencing data of abnormal blood cells from patients with hematopoietic disorders of both lymphoid and myeloid lineages, including myelodysplastic syndrome (MDS), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myelomonocytic leukemia (CML), revealed recurrent somatic mutations in genes encoding splicing factors. These include proteins involved in 3' SS recognition, like the two U2AF subunits, SF1/BBP or ZRSR2, and components of U2 snRNP subcomplexes, like SF3A1 and SF3B1 [88–90]. Mutations were also found in PRP40, a protein involved in bridging 5' and 3' SS-recognizing complexes, and in SRSF2, a splicing regulator of the SR protein family [89] (see below). The most frequently mutated factor was SF3B1, which plays a key role in splice site recognition. Its interactions with U2AF and the pre-mRNA contribute to U2 snRNP recruitment, its phosphorylation is coupled with splicing catalysis, and even its association with chromatin influences splicing outcomes [91–93]. SF3B1 mutations have been identified in up to 85% of MDS patients with refractory anemia with ring sideroblasts (RARS), 15% of CLL and 5% of AML and CMMML patients. Strikingly, while SF3B1 mutations correlate with favorable prognosis in MDS, they correlate with unfavorable prognosis and resistance to fludarabine in CLL [88,94], suggesting that the molecular effects of these mutations are strongly influenced by the cellular context. This concept is further highlighted by the recent identification of mutations in another component of the U2 snRNP-associated SF3B subcomplex (SF3B4) in patients with Nager syndrome, an acrofacial dysostosis characterized by craniofacial and limb malformations [95]. Of interest, ultra-deep sequencing analyses of blood DNA from over 4,000 individuals revealed that expansion of hematopoietic clones harboring mutations in SF3B1 and SRSF2 greatly increases with age, consistent with the higher incidence of MDS in advanced age [96].

The common concept emerging from follow-up transcriptome studies is that mutations in these core spliceosomal components display specific, rather than global effects on splice site recognition, leading to alterations in the profiles of mRNAs and their isoforms that can explain at least part of the associated phenotypes [97,98]. For example, expression of a mutated U2AF35 protein in transgenic mice led to altered hematopoiesis correlating with splicing alterations in RNA processing and ribosomal genes (possibly triggering a cascade of posttranscriptional changes) and in genes frequently mutated in MDS and AML [98].

Classical views of splicing regulation involve the initial steps of spliceosome assembly as the targets of splicing regulators [22]. It would therefore be expected that splicing factors involved in later steps of the process would be more generally required and their mutation more detrimental for intron removal. However, mutations in components of the U4/U6/U5 tri-snRNP also lead to very specific syndromes or disease conditions. For example, mutations in the EFTUD2 gene, which encodes the U5-116 kDa component of the U5 snRNP, are responsible for mandibulofacial dysostosis with microcephaly (MFD), a multiple malformation syndrome [99]. Moreover, mutations in PRPF31, PRPF8, PRPF6, PRPF3, PAP-1, and SNRNP200/BRR2, some of which are involved in the latest spliceosomal rearrangements leading to splicing catalysis, have been causally linked with retinitis pigmentosa (RP), a large group of inherited degenerative disorders of the retina characterized by progressive misfunction of the photoreceptors and the pigment epithelium of the eye [100]. Why should mutations in splicing factors key for the catalytic process lead to a tissue-specific pathology? One possibility is that reduced splicing activity may be more detrimental for rapidly dividing cell types in tissues requiring fast regeneration, like the retina. As a matter of fact, a genome-wide screen for factors important for correct cell division revealed a substantial enrichment in spliceosomal components [101]. Recent reports argue that splicing of genes encoding proteins involved in sister chromatid cohesion and removal of cohesion at the metaphase–anaphase transition, including sororin and APC2, are particularly sensitive to decreased function of splicing factors, including U2AF, components of the SF3A, SF3B, and NTC complex, SNW1, PRPF8, and UBL5, a spliceosome-associated ubiquitin-like protein [102–105]. Detailed structure–function studies revealed that positive regulation and negative regulation of the U4/6 unwinding activity of BRR2 by PRPF8 play key roles in proper activation of splicing catalysis [106,107]. Strikingly, the effects of mutants associated with the most severe forms of RP (RP13) can be correlated with this intermittent block of BRR2 activity [108]. This may reflect the impact of these mutations in splicing in general, in splicing of specific introns (e.g., sororin, APC2), or lead to differential splice site selection and alternative splicing changes, as observed upon knockdown of tri-snRNP components including BRR2 and PRPF8 [29].

While less abundant and of more restricted effects in RNA metabolism, the minor spliceosome can also be the target of disease-causing mutations. Thus, mutations in the gene encoding U4atac snRNA are responsible for the microcephalic osteodysplastic primordial dwarfism type 1 (MOPD1, also known as Taybi–Linder syndrome), characterized by neurological and skeletal abnormalities [109,110], and aberrant splicing of U12-type introns is also the hallmark of mutations in ZRSR2 found in MDS [111].

**Mutations and misregulation of splicing regulators** Considering the strong impact of alternative splicing in cancer [9], it is not surprising that regulatory factors also appear frequently mutated, or their expression changed, in a variety of tumors [8,112,113] (Fig 3B). These include SR (e.g., SRSF1, SRSF2, SRSF6) and hnRNP (A1, I, H) as well as other families of regulatory proteins like TIA-1/TIAR, Sam68, HuR, NOVA, SON, RBM5, and RBM10 [40]. Some selected examples illustrate this point. SRSF1 is upregulated in multiple tumors, partly by gene amplification, and this is sufficient to transform cells by modulating alternative splicing of tumor suppressor and kinase genes, including S6K1—resulting in cell transformation [114], and Ron—leading to increased cell motility and metastasis [115]. Mutations in the SR protein SRSF2 contribute to myelodysplasia by modifying the protein’s RNA-binding specificity, thus...
differentially regulating the activity of exonic enhancers and causing misregulation of key hematopoietic regulators [116–118]. hnRNP I/PTB and hnRNP A1/B2 are upregulated in glioblastomas by Myc oncogene activation and this leads to a switch in pyruvate kinase alternative splicing resulting in efficient energy production in cancer cells by aerobic glycolysis [119]. RBM10 is one of the most frequently mutated genes in lung adenocarcinomas [120], and a mutant version of the protein fails to regulate alternative splicing of the Notch pathway regulator NUMB, thus favoring cell proliferation [121]. Interestingly, mutations in RBM10 also cause TARP (Talipes equinovarus, Atrial septal defects, Robin sequence, and Persistent left superior vena cava) syndrome, a congenital disorder characterized by palate and jaw abnormalities, clubfoot, and cardiac defects, correlating with dysregulation of multiple alternative splicing events [122]. This variety of phenotypes of RBM10 mutations illustrates the multiple—but discrete—functions of splicing regulators during cell differentiation and development.

Mutations in splicing regulators have also been described in several neurodegenerative diseases, such as ALS and frontotemporal lobar degeneration (mutations in TDP-43 and FUS), autism (mutations in RBFOX1) [40], and Huntington’s disease (mutations in SRSF6, see below). Recently, a class of microexons (3–15 nucleotides in length) was discovered that are specifically included in neurons and which are frequently misregulated in patients with autism [123]. It will be of particular interest to investigate whether mutations or changes in activity of key regulators, including the SR protein SR100/SRRM4, can explain the disruption in the program of microexon inclusion associated with autistic disorders.

**Nucleotide repeat expansions: titrating splicing factors (and more)**

The best-understood disease associated with microsatellite expansions is myotonic dystrophy (DM), where expanded CTG repeats in the 3’ UTR of the DMPK gene and expanded CCTG repeats in intron 1 of the ZNF9 gene act as gain of function mutations responsible for type 1 (DM1) and type 2 DM (DM2), respectively [5,6,124,125]. The expanded CUG and CCUG repeats in the pre-mRNAs of these genes can fold on themselves to form relatively long stretches of double-stranded RNA, with two distinct molecular consequences impacting on the function of splicing regulators (Fig 3C). On the one hand, the repeats, which accumulate in nuclear foci, are bound by and sequester MBNL1, a regulator of AS important for cell differentiation [126]. On the other hand, CUG repeats stabilize CUG-binding protein 1 (CUG-BP1) via increased phosphorylation mediated by the activation of PKC [127]. The combined effects of decreasing MBNL1 and increasing CUG-BP1 activity lead to changes in developmentally regulated AS that can explain key features of the disease. For example, delayed muscle relaxation (myotonia) is related to aberrant inclusion of an ORF-disrupting alternative exon in the muscle-specific chloride channel CLCN1 [128,129], while insulin resistance is related to an exon-skipping event in the insulin receptor pre-mRNA that leads to the production of a less sensitive receptor [130]. The multiple effects of changes in the activities of splicing regulatory factors in DM illustrate the combinatorial nature and developmental logic of AS regulation. It also illustrates how, despite these multi-systemic effects, particular phenotypes can be attributed to specific AS changes, and be reversed by these key targets [131].

DM can serve as a paradigm to understand other disorders caused by repeat expansions (Fig 3C) [5], including some forms of ALS, fragile X-associated tremor/ataxia, spinocerebellar ataxia, and, possibly, also Huntington’s disease (HD). The neurodegenerative HD disorder, characterized by involuntary movements, psychiatric symptoms, and dementia, is caused by CAG repeat expansions in Huntingtin (Htt) gene exon 1 [132]. If the number of CAG repeats is higher than 40, symptoms appear, the number of repeats correlating with the severity and the early appearance of the disease. Classically, the pathological effects of the expansion have been attributed to a gain of function due to the increase in polyglutamine peptides, encoded by the CAG repeats, in the Htt protein. This correlates with the accumulation of Htt or of proteolytic degradation products of its glutamine-rich N-terminal region, in inclusion bodies. A recent study reported a cryptic polyadenylation site within Htt intron 1, which becomes activated upon CAG expansion [133]. This RNA species is actively translated into shorter polypeptides containing polyglutamine tracts, therefore offering an alternative explanation for the generation of presumably toxic N-terminal peptides. Interestingly, the SR protein SRSF6 seems to bind to the expanded CAG repeats in these transcripts (which harbor consensus sites for this protein) and may promote inhibition of intron 1 splicing [133], facilitating use of the alternative polyadenylation site. One interesting possible mechanism is that SRSF6 prevents the association of U1 snRNP with the intron 1 5’ SS, explaining both splicing inhibition and cryptic polyadenylation site activation, a common additional function of U1 snRNP binding [134]. An interesting link has been recently made between HD and tauopathies [135]. In this study, an imbalance between Tau isoforms similar to that found in FTDP-17 (see section 1) was observed in brains of HD patients, along with Tau protein deposits in neuronal nuclei that contribute to the motor phenotype of Htt transgenic mice. Remarkably, SRSF6 is a known regulator of Tau exon 10 splicing [136], and the association of SRSF6 with CAG-expanded transcripts (which leads to increased phosphorylation of the protein—somewhat similar to the PKC-mediated phosphorylation of CUG-BP1 by CUG repeat expansions in DM1) can potentially explain not only the generation of shorter Htt transcripts, but also the alterations in Tau splicing [135]. Therefore, changes in activity of SRSF6 induced by expanded CAG repeats may underlie aberrant RNA processing and protein deposits of both Htt and Tau.

**Considering therapeutic approaches**

**Antisense/splice site switching oligonucleotides (ASO/SSO)**

ASO/SSO strategies aim to influence the ratio between mRNA isoforms to restore normal splicing or enforce expression of particular variants with potential therapeutic effects [137]. They do so by base pairing with splice sites or splicing regulatory sequences preventing their recognition by the splicing machinery or cognate regulatory factors. Remarkable success of these approaches has been achieved in mouse models of SMA, where 2’-O-methyl phosphorothioate-modified oligonucleotides complementary to an ISS recognized by hnRNP A1 promote inclusion of SMN2 exon 7, restoring functional levels of SMN protein in vivo and reversing SMA-related symptoms [63,138] (Fig 4). These ASO display surprisingly persistent effects and, strikingly, systemic administration by subcutaneous injection is even more effective than intracerebroventricular administration, arguing that SMN function in peripheral tissues...
strongly contributes to disease progression [139]. Such approaches are currently under clinical trials for the treatment of SMA and Duchenne muscular dystrophy (DMD) [140]. The rationale for the treatment of DMD is that the effect of disease-causing mutations in the dystrophin gene can be overcome by inducing skipping of the exon containing the mutation (or additional exons to preserve the reading frame). Given the length of the gene, with 79 exons, and the repetitive nature of some of its domains, shorter versions of the protein can provide sufficient activity to restore muscle fiber function (Fig 4). Other strategies involve antisense oligonucleotides blocking or degrading the CUG repeat expansions in DM [141,142]. Variants of U7 snRNA (normally involved in histone 3 blocking or degrading the CUG repeat expansions in DM [141,142]. Other strategies involve antisense oligonucleotides blocking or degrading the CUG repeat expansions in DM [141,142]. Variants of U7 snRNA (normally involved in histone 3 blocking or degrading the CUG repeat expansions in DM [141,142].)

Trans-splicing
Trans-splicing is a natural process involving splice sites in two different pre-mRNA transcripts and occurs in a variety of organisms, including protozoa, trypanosomes, and nematodes. It has also been observed in Drosophila and mammalian cells, linked to apoptosis, axon guidance, and the maintenance of cell pluripotency [149,150]. Pre-mRNA trans-splicing, trans-splicing ribozymes, and tRNA-splicing endonucleases have been proposed as RNA repair strategies of potential therapeutic value [151]. Spliceosome-mediated RNA trans-splicing (SMArt) approaches have been anticipated for the treatment of several diseases, including cystic fibrosis [152], SMA [153], DMD [154], and RP [155]. The general strategy is to introduce a pre-trans-splicing molecule (PTM) containing the sequences to be replaced, preceded by a targeting sequence complementary to an intron in the target RNA and containing also a 3′ SS. Splicing between the 5′ SS of the target intron and the 3′ SS of the PTM leads to chimeric transcripts that restore correct mRNA expression. The main hurdle for the therapeutic application of these technologies remains to enhance the limited in vivo efficacy of the trans-splicing process.

Small splicing-modifying molecules
Recent findings have raised significant expectations that small molecules targeting the splicing machinery display specific effects on subsets of splice sites and are potentially useful as therapeutic drugs. Such compounds can also shed light on key mechanisms and alterations of RNA processing associated with complex diseases, including cancer. Three families of natural compounds, fermentation products from Pseudomonas and Streptomyces, display antitumoral properties and target the U2 snRNP SF3B complex [113,156] (Fig 4). While structurally diverse, these molecules harbor a common pharmacophore and have been used as backbones for the synthesis of other active compounds, including spliceostatin A, meyamycins, E7107, and sudemycins [113,156,157]. How can small molecules targeting core components of the splicing machinery have cytostatic effects on cancer cells without being generally cytotoxic? One intriguing observation is that cancer cells appear to be particularly sensitive to the molecular effects of these compounds [158]. For example, potential therapeutic effects of spliceostatin A were found in melanoma cells that acquire vemurafenib resistance through mutations that alter splicing of BRAF, eliminating its RAS-binding domain [159], suggesting a special sensitivity of cancer-associated splicing events. Another explanation is that, at concentrations in which splicing inhibitory drugs exert cytostatic, but not general cytotoxic effects, they do not globally inhibit splicing but rather display selective effects on AS, particularly in genes relevant for cell cycle progression and apoptosis [158,160]. These specific effects may be reminiscent of the effects of SF3B1 mutations frequently found in a variety of tumors, discussed above, suggesting that specific splice site recognition can be tuned either by mutations in, or by small molecules binding to, the SF3B complex. The molecular basis for this specificity remains to be understood, but the drugs appear to interfere, in a sequence context-dependent manner, with steps leading to the progression and proofreading of spliceosome assembly [160–162].

Two recent studies demonstrated the potential of high-throughput drug screens to identify molecules that modulate particular splicing events, specifically to promote SMN2 exon 7 inclusion as possible therapy for SMA [163,164]. Using relatively simple splicing-based fluorescent reporters, structurally different compounds were identified that promote SMN2 exon 7 inclusion, functional SMN protein expression as well as motor function and survival in mouse models of the disease. Although further transcriptome and mechanistic studies will be required, the effects were remarkably selective and at least one of the compounds appeared to specifically
stabilize base pairing between U1 snRNA and particular sequence features of the SMN2 exon 7/intron junction. In this regard, it is interesting that a number of compounds, including kinetin [165], cardiac glycosides [166], and RECTAS [167], have been shown to facilitate recognition of the 5′ SS of the IKBKP gene mutated in FD, perhaps by stabilizing particular sets of base-pairing interactions, as proposed for SMN2. Another interesting therapeutic lead would be the use of compounds able to modulate particular secondary pre-mRNA structures, as shown for MAPT exon 10 [168].

It seems likely that a detailed understanding of the mechanisms of splice site recognition and spliceosome assembly, including structural determination of complexes bound to specific substrates in the absence and in the presence of small molecule modulators, will open an almost unexplored territory to regulate gene expression and potentially correct disease-causing splicing defects.

Acknowledgements

We apologize to many colleagues whose work could not be directly referenced because of space constraints. We thank members of our laboratory for comments on the manuscript. Work in JV laboratory is supported by Fundación Botín, Banco de Santander through its Santander Universities Global Division, Consolider RNAREG, Ministerio de Economía y Competitividad, and AGAUR. GD is supported by the Marie Sklodowska Curie Fellowship Program.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Pathogenicity of splicing defects

Elisabeth Daguenet et al


EMBO reports

Pathogenicity of splicing defects

Elisabeth Daguenet et al


Pathogenicity of splicing defects
Elisabeth Daguenet et al


