Published in final edited form as:

Trends Genet. 2012 April ; 28(4): 147–154. doi:10.1016/j.tig.2012.01.001.

New connections between splicing and human disease

Richard A. Padgett

Department of Molecular Genetics, Cleveland Clinic, Cleveland, Ohio, USA

Abstract

The removal by splicing of introns from the primary transcripts of most mammalian genes is an essential step in gene expression. Splicing is performed by large, complex ribonucleoprotein particles called spliceosomes. Mammals contain two types that splice out mutually exclusive types of introns. However, the role of the minor spliceosome has been poorly studied. Recent reports have now shown that mutations in one minor spliceosomal snRNA, U4atac, are linked to a rare autosomal recessive developmental defect. In addition, very exciting recent results of exome deep sequencing have found that recurrent, somatic, heterozygous mutations of other splicing factors occur at high frequencies in certain cancers and pre-cancerous conditions suggesting that alterations in the core splicing machinery can contribute to tumorigenesis. Missplicing of critical genes may underlie the pathologies of both these diseases. Identifying these genes and understanding the mechanisms involved in their missplicing may lead to advancements in diagnosis and treatment.

Intron removal by the spliceosomes

The removal of introns from pre-messenger RNAs by RNA splicing is an essential function in virtually all eukaryotic organisms. In vertebrates, including humans, most genes have multiple introns and most such genes are spliced in more than one pattern to give rise to considerable numbers of alternative mRNA isoforms. These isoforms can have altered protein coding potential and/or altered regulatory regions and are thought to help generate organismal level complexity from a limited number of genes. As a consequence, mutations that affect either individual splicing signals or the splicing machinery itself have been found to be the underlying defects in a number of diseases. Here I discuss two recent discoveries that link new mutations in components of the splicing machinery to human diseases. In the first case, inherited mutations lead to a form of hereditary dwarfism while, in the second case, somatically acquired mutations are linked to certain cancers and pre-cancerous conditions.

The odd structure of most mammalian genes with their alternating pattern of exons and introns was first recognized in the 1970s [1, 2]. Much subsequent work showed that the splicing machinery and the RNA sequences that specify the sites of splicing were highly conserved across all eukaria. Nevertheless, in the 1990s, a second, much rarer type of pre-mRNA intron was discovered in vertebrates that used different splice site sequences [3, 4]. It was quickly determined that these rare introns were spliced by an alternative spliceosome that worked alongside the previously known spliceosome [5, 6]. With the sequencing of

© 2012 Elsevier Ltd. All rights reserved.

Corresponding Author: Padgett, R.A. (padgetr@ccf.org).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

many eukaryotic genomes, we now recognize that these minor class or U12-dependent introns entered the eukaryotic genome very early in evolution [7] although they have been subsequently lost from a number of organisms [8]. In the human genome, we can computationally recognize about 800 such introns within about the same number of genes [9–11]. Each of these genes also contains introns of the major or U2-dependent class requiring that the two splicing systems cooperate to generate the final mRNAs from these genes.

Introns are recognized in large part through their splice site sequences located at the exon/intron junctions (Figure 1). The consensus splice site sequences for U2-dependent and U12-dependent introns differ so that, in most cases, the spliceosome responsible for their removal can be predicted by computational methods [9–11]. Within the U12-dependent intron group there are two classes that differ in their terminal nucleotides. As shown in Figure 1B, some introns begin with the dinucleotides AT and end with AC while the other class begins with GT and ends with AG.

The splice site sequences of all types of introns help to promote splicing and determine the precise sites of the RNA cleavage and ligation reactions diagramed in Figure 2A. Splicing involves a two step reaction that is initiated by the attack of the 2' hydroxyl group of an adenosine residue near the 3' end of the intron on the phosphodiester bond at the 5' splice site. This yields a free 5' exon and a branched lariat RNA in which the 5' end of the intron is linked to the branch site via a 2'–5' phosphodiester bond. In the second step, the 3' hydroxyl of the 5' exon attacks the phosphodiester bond at the 3' splice site to yield the ligated exons and the excised lariat intron RNA.

The recognition of the splice sites and the chemical reactions of splicing take place within one of the most complex molecular machines in the cell called the spliceosome (reviewed in [12, 13]). The two types of spliceosomes are each formed from five small nuclear RNAs (snRNAs), a number of associated proteins that form small nuclear ribonucleoprotein particles called snRNPs as well as a large number of additional non-snRNP proteins. Proteomic studies of *in vitro* assembled spliceosomes have counted 150–300 individual polypeptides (reviewed in [12, 14, 15]). The U2- and U12-dependent spliceosomes differ mainly in their snRNAs and snRNPs. The major U2-dependent spliceosome requires the function of snRNAs known as U1, U2, U4, U5 and U6 while the minor U12-dependent spliceosome requires U11, U12, U5, U4atac and U6atac (Figure 2B). Thus, four of the snRNAs are unique to each spliceosome while the U5 snRNA is used by both. There is also considerable overlap in the proteins required for the two spliceosomes [16–18].

Each spliceosome is assembled around an intron in a complex multistep process (reviewed in [12, 17]). For our purpose, I have presented a simplified version that focuses on the factors and processes recently implicated in diseases. The early steps in intron recognition are diagrammed in Figure 2 for the major U2-dependent spliceosome. Less is known about these steps in the minor U12-dependent spliceosome pathway. This part of the assembly pathway is where splice sites are recognized, a process which culminates in the base pairing of U1 snRNA to the 5' splice site and U2 snRNA to the branch site sequence. As shown in Figure 2, a class of RNA binding proteins called SR proteins, represented here by SRSF2, bind to exonic splicing enhancer elements [19]. These SR proteins help to recruit the U1 snRNP to the 5' splice site and factors such as the U2AF65/U2AF35 complex to the 3' splice site either directly or through other proteins such as ZRSR2. The U2AF complex then recruits SF1 to the branch site region. U2 snRNP then displaces SF1 at the branch site where it base pairs assisted by additional RNA interactions through proteins such as SF3B1 and SF3A1. The complexes forming on the 5' and 3' splice sites are linked to each other by additional proteins such as PRPF40B. The splice sites of U12-dependent introns are

recognized by a similar pathway culminating in the base pairing of U11 snRNA to the 5' splice site and of U12 snRNA to the branch site [17].

Following these early splice site recognition events, the spliceosomes are assembled in a step-wise fashion that is diagrammed in simplified form in Figure 3A and B for the two different spliceosomes [13]. In both spliceosomes, the splice site sequences in the pre-mRNA and the snRNAs participate in a number of base pairing interactions shown by yellow boxes. The left side of Figure 3 shows the 5' splice site and branch site regions base paired to U1 and U2 for the U2-dependent intron and U11 and U12 for the U12-dependent intron. In the next series of steps, a large tri-snRNP complex containing three snRNAs (U4, U5 and U6 for the U2-dependent spliceosome and U4atac, U5 and U6atac for the U12-dependent spliceosome) associates with the initial complex leading to a rearrangement of the spliceosome and the loss of U1 and U4 or U11 and U4atac as shown. The remaining U2 and U6 or U12 and U6atac snRNAs base pair with each other and the splice sites and, in conjunction with U5 snRNA and a large number of protein factors, catalyze the splicing reactions. A point that will become important later is that within the tri-snRNP complexes, U4 and U6 snRNAs, as well as their analogs U4atac and U6atac snRNAs, are extensively base paired to each other. As part of the activation process of the spliceosomes, these base pairings are unwound and U4 (or U4atac) is lost while the newly exposed regions of U6 (or U6atac) snRNA form several critical RNA-RNA interactions [13, 17].

RNA splicing and disease

The fact that genes are composed of multiple segments has a variety of functional consequences, some positive and some negative. On the positive side, it is clear that the vast majority of mammalian genes are spliced and otherwise processed in different ways allowing for the expression of many different mRNA isoforms and protein products from a relatively small number of nuclear genes [20]. This diversity of gene products is central to the complex processes of growth, development and differentiation. The dark side of this diversity is that mutations can disrupt these processes in a variety of ways often leading to pathological conditions [21]. Indeed, frank mutations of the consensus splice site sequences found at the junctions between exons and introns are estimated to account for about 10% of human heritable disorders [22]. Mutations outside of these splice site sequences that affect splicing regulatory sequences (found in both exons and introns) may account for another significant fraction, perhaps as high as 25% [23, 24].

In addition to mutations in splicing signals of individual genes, mutations of the splicing machinery itself give rise to human diseases (recently reviewed in [23]). Two notable examples are spinal muscular atrophy (SMA), which is caused by mutations in the *SMN2* gene [25], and autosomal dominant retinitis pigmentosa (adRP), which is caused by mutations of several spliceosomal proteins. The *SMN2* gene product is required for proper maturation of the snRNPs used by the spliceosomes [26]. In SMA, a reduced level of this protein leads to the tissue specific death of spinal motor neurons. In adRP, about 5% of cases have been linked to mutations in five genes (*PRPF31*, *PRPF8*, *HPRP3*, *PAP1* and *SNRNP200*) all of which code for protein components of the tri-snRNP complexes (U4, U5, U6 and U4atac, U5, U6atac) that join the spliceosomes in the step shown in Figure 3 [27]. These mutations lead to the very tissue specific death of retinal photoreceptors cells seen in this disorder. Why these various mutations should all produce the same very limited spectrum of pathology seen in adRP is unclear. However, each of these genes is involved in the structure and/or function of the tri-snRNP complex suggesting that a common defect in the spliceosome assembly pathway may underlie the adRP pathology. The important message from these two cases is that mutations in broadly expressed, essential genes can give rise to highly tissue specific pathologies with no apparent effects on other cell types.

Below I will examine two newly identified types of splicing factor mutations that lead to human disease.

Mutations in a spliceosomal RNA cause disease

Microcephalic Osteodysplastic Primordial Dwarfism type I (MOPD I also known as Taybi-Linder Syndrome) is an autosomal recessive genetic disorder that is rare in most populations but is quite common among the Amish in Ohio [28]. This condition is diagnosed on the basis of characteristic neurological and skeletal abnormalities as well as slow intra-uterine growth and greatly reduced size at term. Life spans of MOPD I patients are quite variable from only a few months to close to 13 years [29]. The cause of death is not clear in many cases but appears to be most often associated with infections suggesting the possibility of immune system defects as well.

Two research groups localized the MOPD I mutations by deep sequencing to a small region within an intron of the *CLASP1* gene on chromosome 2 [28, 30]. Remarkably, this intron contains the gene for the U12-dependent spliceosomal snRNA U4atac that is transcribed in the opposite direction from the *CLASP1* gene. In total, the two groups identified eight different single nucleotide point mutations in the U4atac snRNA gene (*RNU4ATAC*) that were found in MOPD I patients as either homozygotes or compound heterozygotes. All but one of these mutations were located within a single double stranded stem in the U4atac RNA secondary structure (Figure 4). All of the disease associated mutations disrupted presumptive base pairing interactions as shown. The secondary structure of U4atac is conserved in widely divergent organisms and is highly similar to the conserved structure of the U2-dependent spliceosomal U4 snRNA suggesting that this region is likely to be critical for function.

The analogous 5' stem loop element of U4 snRNA has been shown to bind two proteins (NHP2L1 and PRP31) that are essential for the formation of the tri-snRNP complexes used in the assembly of the spliceosomes [31]. Both of these proteins are also known to bind to wild type U4atac snRNA [32]. It is likely, but not yet proven, that the mutations in this region of U4atac snRNA found in MOPD I patients disrupt the binding of one or both of these proteins. The level of U4atac snRNA in one patient cell line was reported to be similar to wild type suggesting that the mutant snRNA itself is not unstable [30]. Functional studies of mutant U4atac snRNAs carried out in an *in vivo* splicing reporter system showed that all of the MOPD I associated mutations tested reduced U4atac activity in U12-dependent splicing to less than 10% of wild type [28]. Both groups reported that patient derived cell lines showed defects in U12-dependent intron splicing while U2-dependent intron splicing was relatively unaffected [28, 30]. In addition, restoration of a wild type U4atac snRNA gene to patient cells caused a rescue of the U12-dependent splicing defect [28]. All of these molecular studies support the proposition that the MOPD I phenotype is due to defective U12-dependent splicing.

This discovery represents the first case of a human disease caused by mutations in a spliceosomal snRNA. It is also the first case of a functional defect linked to the minor spliceosome. While relatively few genes contain U12-dependent introns, many genes that do are essential for basic cellular functions. The larger number of cell types affected in MOPD I compared to SMA or adRP suggests that more target genes are being affected by the U4atac snRNA mutations, however, this remains to be shown.

A surprising feature of the MOPD I U4atac snRNA mutations is that so many different mutations were found in a rather small number of patients from several different human populations (Figure 4). This raises the possibility that there may be an, unrecognized pool of mutations in this gene present at a low level throughout many populations. We can also see

in the Ohio Amish population the more typical case of a single recessive mutant allele that manifests mainly in small consanguineous groups. In the Ohio Amish, for example, the MOPD I-associated 51G>A mutation is found in 8% of the members [28]. The modern members of this group are all descended from a small number of founding families. Through many generations of intra-group marriage, most modern members are related to each other in multiple ways. It seems likely that the 51G>A mutation was present in one of the founding families of the Ohio Amish. Interestingly, the Pennsylvania Amish, founded by a different group of families, do not appear to carry this mutation [28].

The U4atac snRNA mutations seen in MOPD I have been shown to reduce the function of U4atac snRNA by over 90%. U4atac snRNA is also the least abundant spliceosomal snRNA in most cell types [33] which might suggest that it could be limiting in normal cells. However, as its role in splicing seems to be limited to the delivery of U6atac snRNA to the forming spliceosome, its function may be more transient than other snRNAs allowing a lower level to be sufficient. Since the heterozygous parents of homozygous MOPD I children appear to have no reported phenotype, we may conclude that the loss of close to 50% of U4atac snRNA function is well tolerated. It is only in homozygous mutant individuals where the loss of 90–95% of function causes severe disease.

Current questions raised by the MOPD I mutations

Among the immediate questions raised by these findings are first, what is the molecular defect in the mutant U4atac snRNAs i.e. at what point in the assembly of the snRNP complexes or their function in splicing do they fail? This problem could be investigated using *in vitro* biochemical studies of NHP2L1 and PRP31 protein binding to the mutant snRNA to assay for defects in direct binding. Since MOPD I patient derived cell lines are available, the ability of the mutant U4atac snRNA to form U4atac/U6atac di-snRNP and U4atac/U6atac/U5 tri-snRNP complexes could be examined.

A second, and more difficult question is what is/are the important downstream gene or genes that are affected by the splicing defect? We can speculate that either the mRNA level or the spliced isoforms of one or more U12-dependent intron containing gene is reduced or altered in the mutant cells leading to the specific pathologies seen in MOPD I. Detailed analyses of the transcriptomes of patient cells may begin to reveal such genes.

MOPD I is one of several inherited forms of primordial dwarfism [34]. While the precise clinical presentations of these syndromes vary, they all share the feature of extreme global growth failure both pre- and post-natally. Ten genes have now identified in these disorders including U4atac snRNA. The other genes are involved in other fundamental cell processes including DNA replication [35–37], DNA damage response [38, 39] and centrosome function [40, 41].

There are some interesting connections between these functions and U12-dependent intron-containing genes. For example, mutations have been found in three of the *ORC* genes (*ORC1*, *ORC4* and *ORC6*), which are part of a complex (ORCs 1–6) involved in activating DNA replication origins, in patients with Meier-Gorlin syndrome [35–37]. The gene for another component, *ORC3*, contains a U12-dependent intron making this a candidate pathway for the growth retardation effect in MOPD I. Similarly, an important downstream target of intracellular signalling pathways leading to cell proliferation is the E2F family of transcription factors. All *E2F* genes in mammals contain a U12-dependent intron. Again, poor U12-dependent splicing of one or more of these genes could lead to decreased levels of E2F proteins which could, in turn, lead to growth retardation. The determination of the downstream target genes responsible for the defects in growth and development in MOPD I

may provide additional insights into the still largely unknown pathways that regulate organism size and cellular differentiation.

In a somewhat analogous example, analysis of a *Drosophila melanogaster* strain carrying a defective U6atac snRNA gene revealed significant changes in gene expression of a large number of genes in spite of having very few U12-dependent introns in its genome. The effects were proposed to ultimately stem from the reduced expression of a single nuclear encoded mitochondrial protein with a U12-dependent intron in its gene [42].

In the case of the MOPD I mutations in U4atac snRNA, it is somewhat surprising that such a drastic reduction in function of an essential spliceosomal component is still compatible with growth and development of most tissues albeit with a reduced rate of growth. Many human genes essential to cellular viability contain U12-dependent introns [10, 11] so that a complete loss of minor spliceosomal function is likely to be lethal at the cellular level. Nevertheless, cells can survive with only 10% of normal U4atac snRNA activity [28, 33]. In contrast, the average effect on splicing efficiency of U12-dependent introns in patient-derived cell lines was only about two fold [28, 30]. Perhaps only the rate of splicing of most introns is affected when U4atac snRNA is limiting. However, evidence suggests that changes in transcription and splicing rates can alter the balance of alternatively spliced mRNA isoforms [43] which could, in turn, lead to pathology.

A third question is are there diseases linked to mutations of the other three U12-dependent spliceosomal snRNAs? Should we expect to see cases of MOPD I or similar syndromes caused by mutations in these genes? As genome sequencing becomes less expensive and more common, we may see more cases and perhaps more diseases being linked to these genes. However, it is worth noting that the popular exome sequencing technique does not currently include the sequencing of snRNA genes. The recent MOPD I studies instead relied on partial deep sequencing of known candidate chromosomal regions to identify the U4atac snRNA mutations [28, 30]. Finally, we can ask if there is a specific phenotype of people with less severe loss of function mutations of U4atac snRNA such that they fall between the 90–95% defect in MOPD I and the 50% loss in the phenotypically normal parents of MOPD I patients? In conclusion, although MOPD I is a very rare disease, it provides us with evidence of the importance of the U12-dependent splicing system and provides a window into several important issues in gene expression as well as cellular growth and development pathways.

Splicing factor mutations in myeloid neoplasms and leukemias

All of the above examples are linked to inherited mutations in splicing components. Recently, a striking example of the effects of acquired somatic mutations in splicing factors has been described. The sequencing of the DNA from abnormal blood cells from patients with several types of leukemia and pre-leukemic syndromes has shown that a high proportion of these cases are associated with somatic mutations in spliceosomal proteins [44–47]. These diseases include myelodysplastic syndrome (MDS), which often progresses to acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL) and chronic myelomonocytic leukemia (CMML). The spliceosomal proteins that have been found to be mutated are those shown in Figure 2B involved in the early steps of U2-dependent splice site recognition. In particular, heterozygous, single amino acid substitution mutations of the SF3B1 protein are seen in up to 65% of a subset of MDS patients with ring sideroblasts, 5% of AML and CMML patients and 15% of CLL patients. Similar mutations of SRSF2 were detected in 28% of CMML patients and mutations of U2AF1 (U2AF35) in about 10% of MDS, CMML and AML patients. Other splicing factors are also frequently mutated in myeloid neoplasms including ZRSR2 (URP), SF1, PRPF40B, U2AF1 (U2AF65) and SF3A1

[44–47]. These mutations occurred in a mutually exclusive pattern and also, with the exception of ZRSR2, were only amino acid substitution mutations often with a high recurrence of individual missense mutations. In the case of ZRSR2, missense, nonsense and frameshift mutations were all detected. The lack of nonsense and frameshift mutations and the high recurrence of particular missense mutations in many of these factors suggests that disease phenotype is not caused by decreased levels of these factors but by an alteration in protein function [48]. In the absence of functional studies on these mutated proteins, the mechanism is not known but it could involve either a gain of function (e.g. recognition of incorrect splice sites) or dominant negative function (e.g. blockage of splice site recognition). Strikingly, all of these factors are involved in splice site recognition often in the context of alternatively spliced genes [49–52]. This suggests that the mutated factor in each case may be perturbing the balance of one or more sets of critical alternative isoforms. Thus, as with the case of MOPD I, the identification of the key target genes affected by these mutations will be an important next step.

These results suggest that certain types of cancer are driven by alterations in the cellular splicing machinery. The abnormal cells seen in MDS must have a growth advantage over normal bone marrow cells and this may push them along the pathway to leukemia. From a clinical perspective, these new results also provide prognostic markers for the course of disease. For example, in MDS, mutations in SF3B1 are associated with longer survival while the same mutations in CLL appear to be associated with poorer prognosis [44, 46]. Knowing the precise mutations present in an individual patient's tumor may be able to guide the choice of therapy and improve outcomes. In addition, mechanistic analyses of the mutant proteins and their downstream targets may provide additional drug targets for the treatment of these diseases.

Concluding remarks

These recent results emphasize the role of mutations in the essential machinery of pre-mRNA splicing in the realm of human health and disease. While the minor U12-dependent spliceosome has been relatively little studied, it is now clear that defects in this machine can have substantial effects on growth and development. Future studies that identify the target genes have the potential to uncover important pathways in the various cell types affected in MOPD I. Furthermore, the linkage of somatic splicing factor mutations to pre-neoplastic conditions and leukemias suggests that this may be an important pathway in many types of cancer. The current sequencing of cancer exomes will be most revealing in terms of identifying additional factors or additional types of cancer in which these factors are mutated. The high frequency and site specificity of some of these mutations also make the affected factors attractive targets for drug development especially if the mutations lead to proteins with altered functions.

References

1. Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A*. 1977; 74:3171–3175. [PubMed: 269380]
2. Chow LT, Gelinis RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*. 1977; 12:1–8. [PubMed: 902310]
3. Hall SL, Padgett RA. Conserved sequences in a class of rare eukaryotic introns with non-consensus splice sites. *J. Mol. Biol.* 1994; 239:357–365. [PubMed: 8201617]
4. Jackson IJ. A reappraisal of non-consensus mRNA splice sites. *Nucl. Acids Res.* 1991; 19:3795–3798. [PubMed: 1713664]
5. Hall SL, Padgett RA. Requirement of U12 snRNA for the in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science*. 1996; 271:1716–1718. [PubMed: 8596930]

6. Tarn W-Y, Steitz JA. A novel spliceosome containing U11, U12 and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell*. 1996; 84:801–811. [PubMed: 8625417]
7. Russell AG, Charette JM, Spencer DF, Gray MW. An early evolutionary origin for the minor spliceosome. *Nature*. 2006; 443:863–866. [PubMed: 17051219]
8. Bartschat S, Samuelsson T. U12 type introns were lost at multiple occasions during evolution. *BMC Genomics*. 2010; 11:106. [PubMed: 20149226]
9. Alioto TS. U12DB: a database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res*. 2007; 35:D110–D115. [PubMed: 17082203]
10. Burge CB, Padgett RA, Sharp PA. Evolutionary fates and origins of U12-type introns. *Mol. Cell*. 1998; 2:773–785. [PubMed: 9885565]
11. Levine A, Durbin R. A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res*. 2001; 29:4006–4013. [PubMed: 11574683]
12. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*. 2009; 136:701–718. [PubMed: 19239890]
13. Will, CL.; Luhrmann, R. Spliceosome structure and function. In: Gestland, RF.; Cech, TR.; Atkins, JF., editors. *The RNA World, Third Edition*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2006. p. 369-400.
14. Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*. 2003; 12:5–14. [PubMed: 12887888]
15. Valadkhan S, Jaladat Y. The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine. *Proteomics*. 2010; 10:4128–4141. [PubMed: 21080498]
16. Schneider C, Will CL, Makarova OV, Makarov EM, Luhrmann R. Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Mol Cell Biol*. 2002; 22:3219–3229. [PubMed: 11971955]
17. Will CL, Luhrmann R. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem*. 2005; 386:713–724. [PubMed: 16201866]
18. Will CL, Schneider C, Reed R, Luhrmann R. Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science*. 1999; 284:2003–2005. [PubMed: 10373121]
19. Lin S, Fu XD. SR proteins and related factors in alternative splicing. *Adv Exp Med Biol*. 2007; 623:107–122. [PubMed: 18380343]
20. Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456:470–476. [PubMed: 18978772]
21. Ward AJ, Cooper TA. The pathobiology of splicing. *J Pathol*. 2010; 220:152–163. [PubMed: 19918805]
22. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet*. 2007; 8:749–761. [PubMed: 17726481]
23. Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell*. 2009; 136:777–793. [PubMed: 19239895]
24. Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res*. 2011; 21:1563–1571. [PubMed: 21750108]
25. Lefebvre S, et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*. 1995; 80:155–165. [PubMed: 7813012]
26. Neuenkirchen N, Chari A, Fischer U. Deciphering the assembly pathway of Sm-class U snRNPs. *FEBS Lett*. 2008; 582:1997–2003. [PubMed: 18348870]
27. Mordes D, Luo X, Kar A, Kuo D, Xu L, Fushimi K, Yu G, Sternberg P Jr, Wu JY. Pre-mRNA splicing and retinitis pigmentosa. *Mol Vis*. 2006; 12:1259–1271. [PubMed: 17110909]
28. He H, et al. Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science*. 2011; 332:238–240. [PubMed: 21474760]
29. Nagy R, Wang H, Albrecht B, Wieczorek D, Gillesen-Kaesbach G, Haan E, Meinecke P, de la Chapelle A, Westman JA. Microcephalic Osteodysplastic Primordial Dwarfism type I with biallelic mutations in the RNU4ATAC gene. *Clin Genet*. 2011

30. Edery P, et al. Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science*. 2011; 332:240–243. [PubMed: 21474761]
31. Liu S, Li P, Dybkov O, Nottrott S, Hartmuth K, Luhrmann R, Carlomagno T, Wahl MC. Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science*. 2007; 316:115–120. [PubMed: 17412961]
32. Schultz A, Nottrott S, Hartmuth K, Luhrmann R. RNA structural requirements for the association of the spliceosomal hPrp31 protein with the U4 and U4atac small nuclear ribonucleoproteins. *J Biol Chem*. 2006; 281:28278–28286. [PubMed: 16857676]
33. Pessa HK, Ruokolainen A, Frilander MJ. The abundance of the spliceosomal snRNPs is not limiting the splicing of U12-type introns. *Rna*. 2006; 12:1883–1892. [PubMed: 16957280]
34. Klingseisen A, Jackson AP. Mechanisms and pathways of growth failure in primordial dwarfism. *Genes Dev*. 2011; 25:2011–2024. [PubMed: 21979914]
35. Bicknell LS, et al. Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nat Genet*. 2011; 43:356–359. [PubMed: 21358632]
36. Bicknell LS, et al. Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. *Nat Genet*. 2011; 43:350–355. [PubMed: 21358633]
37. Guernsey DL, et al. Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. *Nat Genet*. 2011; 43:360–364. [PubMed: 21358631]
38. Alderton GK, Joenje H, Varon R, Borglum AD, Jeggo PA, O'Driscoll M. Seckel syndrome exhibits cellular features demonstrating defects in the ATR-signalling pathway. *Hum Mol Genet*. 2004; 13:3127–3138. [PubMed: 15496423]
39. O'Driscoll M, Ruiz-Perez VL, Woods CG, Jeggo PA, Goodship JA. A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. *Nat Genet*. 2003; 33:497–501. [PubMed: 12640452]
40. Griffith E, et al. Mutations in pericentrin cause Seckel syndrome with defective ATR-dependent DNA damage signaling. *Nat Genet*. 2008; 40:232–236. [PubMed: 18157127]
41. Rauch A, et al. Mutations in the pericentrin (PCNT) gene cause primordial dwarfism. *Science*. 2008; 319:816–819. [PubMed: 18174396]
42. Pessa HK, Greco D, Kvist J, Wahlstrom G, Heino TI, Auvinen P, Frilander MJ. Gene expression profiling of u12-type spliceosome mutant *Drosophila* reveals widespread changes in metabolic pathways. *PLoS One*. 2010; 5
43. Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. Multiple links between transcription and splicing. *Rna*. 2004; 10:1489–1498. [PubMed: 15383674]
44. Papaemmanuil E, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*. 2011; 365:1384–1395. [PubMed: 21995386]
45. Visconte V, et al. SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. *Leukemia*. 2011
46. Wang L, et al. SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. *N Engl J Med*. 2011
47. Yoshida K, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011; 478:64–69. [PubMed: 21909114]
48. Abdel-Wahab O, Levine R. The spliceosome as an indicted conspirator in myeloid malignancies. *Cancer Cell*. 2011; 20:420–423. [PubMed: 22014568]
49. Pacheco TR, Gomes AQ, Barbosa-Morais NL, Benes V, Ansorge W, Wollerton M, Smith CW, Valcarcel J, Carmo-Fonseca M. Diversity of vertebrate splicing factor U2AF35: identification of alternatively spliced U2AF1 mRNAs. *J Biol Chem*. 2004; 279:27039–27049. [PubMed: 15096518]
50. Corrionero A, Minana B, Valcarcel J. Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes Dev*. 2011; 25:445–459. [PubMed: 21363963]
51. Wang J, Manley JL. Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing in vivo in diverse ways. *RNA*. 1995; 1:335–346. [PubMed: 7489505]

52. Shen H, Zheng X, Luecke S, Green MR. The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev.* 2010; 24:2389–2394. [PubMed: 21041408]

A **B**

Figure 1. Splice site consensus sequences for U2-dependent (a) and U12-dependent introns (b)
 The two classes of introns found in human genes can be distinguished by their splice site sequences. The boxes show graphical representations of the consensus sequences in which the size of each letter represents the frequency of each base at each position over all introns. The bases are ordered by frequency from top to bottom. U2-dependent introns almost always begin with the dinucleotide GT and end with AG while U12-dependent introns can have either AT and AC termini or GT and AG termini.

Figure 2. The splicing reaction and interactions in the early phase of intron recognition and spliceosome formation

A. The splicing reaction catalysed by the spliceosome occurs in two steps. In the first step, the phosphodiester bond at the 5' splice site is broken and the 5' end of the intron is joined to the 2' hydroxyl group of the branch site adenosine residue. In the second step, the phosphodiester bond at the 3' splice site is broken and joined to the free 3' hydroxyl group of the 5' exon. The products are the ligated exons and the intron in the form of a lariat RNA.

B. During the initial steps in splicing, the splice sites and adjacent RNA sequences are bound by a network of interacting factors. A subset of these factors is shown here of which several have been implicated in myeloid malignancies as discussed in the text. Some of the initial interactions include the binding of SR family proteins such as SRSF2 to splicing enhancer elements located within exons which recruit the U2AF65/U2AF35 heterodimer to the 3' splice site either directly or through additional factors such as ZRSR2. The U2AF65 subunit binds to the pyrimidine rich region of the 3' splice site while the U2AF35 subunit binds to the AG dinucleotide at the splice junction. SF1 binds to the branch site A residue while U1 snRNP binds to the 5' splice site through base pairing interactions. The 5' and 3' splice site complexes are joined together by protein-protein interactions mediated by factors such as PRPF40B. Subsequent to these steps, the U2 snRNP is recruited by U2AF to the branch site where it base pairs to the intron RNA. U2 snRNP binding is also stabilized by binding of the SF3b complex (which includes the SF3B1 protein) to the RNA upstream of the branch site.

Figure 3. Formation of the spliceosomes

The early steps of spliceosome formation culminate in the base pairing of U1 or U11 and U2 or U12 to the 5' and 3' splice sites of U2-dependent or U12-dependent introns respectively (base pairings are indicated by the yellow bars). In the next phase of assembly, tri-snRNP complexes composed of U4, U5 and U6 or U4atac, U5 and U6atac are joined to the forming spliceosome. The base pairs connecting U4 and U6 or U4atac and U6atac are unwound and new pairings are made between U6 and U2 or U6atac and U12 leading to the release of U4 or U4atac, U6 or U6atac also form base pairs to the 5' splice site displacing U1 or U11 from the complex. U5 interacts with the exons to hold the RNAs in place during the splicing reactions.

The sequence and predicted secondary structure of U4atac snRNA (in black) is shown in a base pairing configuration with U6atac snRNA (in gray). The sites of MOPD I mutations are boxed and the base changes seen in the patients are indicated. Each patient was homozygous for a single base change or heterozygous for different point mutations on the two alleles.