Pre-mRNA splicing and human disease

Nuno André Faustino1,3 and Thomas A. Cooper1,2,4

Departments of 1Pathology and 2Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA; 3Graduate Program in Basic and Applied Biology, ICBAS, University of Oporto, Portugal

The precision and complexity of intron removal during pre-mRNA splicing still amazes even 26 years after the discovery that the coding information of metazoan genes is interrupted by introns [Berget et al. 1977; Chow et al. 1977]. Adding to this amazement is the recent realization that most human genes express more than one mRNA by alternative splicing, a process by which functionally diverse protein isoforms can be expressed according to different regulatory programs. Given that the vast majority of human genes contain introns and that most pre-mRNAs undergo alternative splicing, it is not surprising that disruption of normal splicing patterns can cause or modify human disease. The purpose of this review is to highlight the different mechanisms by which disruption of pre-mRNA splicing play a role in human disease. Several excellent reviews provide detailed information on splicing and the regulation of splicing [Burge et al. 1999; Hastings and Krainer 2001; Black 2003]. The potential role of splicing as a modifier of human disease has also recently been reviewed [Nissim-Rafinia and Kerem 2002].

Constitutive splicing and the basal splicing machinery

The typical human gene contains an average of 8 exons. Internal exons average 145 nucleotides [nt] in length, and introns average more than 10 times this size and can be much larger [Lander et al. 2001]. Exons are defined by rather short and degenerate classical splice-site sequences at the intron/exon borders (5′/H11032/H11032 splice site, 3′ splice site, and branch site; Fig. 1A). Components of the basal splicing machinery bind to the classical splice-site sequences and promote assembly of the multicomponent splicing complex known as the spliceosome. The spliceosome performs the two primary functions of splicing: recognition of the intron/exon boundaries and catalysis of the cut-and-paste reactions that remove introns and join exons. The spliceosome is made up of five small nuclear ribonucleoproteins [snRNPs] and >100 proteins. Each snRNP is composed of a single uridine-rich small nuclear RNA [snRNA] and multiple proteins. The U1 snRNP binds the 5′ splice site, and the U2 snRNP binds the branch site via RNA:RNA interactions between the snRNA and the pre-mRNA [Fig. 1B]. Spliceosome assembly is highly dynamic in that complex rearrangements of RNA:RNA, RNA:protein, and protein:protein interactions take place within the spliceosome. Coinciding with these internal rearrangements, both splice sites are recognized multiple times by interactions with different components during the course of spliceosome assembly [for example, see Burge et al. 1999; Du and Rosbash 2002; Lallena et al. 2002; Liu 2002]. The catalytic component is likely to be U6 snRNP, which joins the spliceosome as a U4/U6 · U5 tri-snRNP [Villa et al. 2002].

A splicing error that adds or removes even 1 nt will disrupt the open reading frame of an mRNA; yet exons are correctly spliced from within tens of thousands of intronic nucleotides. This remarkable precision is, in part, built into the mechanism of intron removal because once the spliceosome is assembled, the base-paired snRNAs target specific phosphate bonds for cleavage. The challenge for the spliceosome comes in recognizing the correct splice sites prior to the cut-and-paste reactions. The short and degenerate splice sites contain only half of the information necessary for splice-site recognition [Lim and Burge 2001] because bona fide splice sites must be distinguished from pseudo splice-site sequences that resemble classical splice sites but are never used. Pseudo splice sites can outnumber bona fide splice sites within a pre-mRNA by an order of magnitude [Sun and Chasin 2000]. Auxiliary cis-elements, known as exonic and intronic splicing enhancers [ESEs and ISEs] and exonic and intronic splicing silencers [ESSs and ISSs; Fig. 1B], aid in the recognition of exons [see below].

It is now clear that exon recognition is accomplished by the accumulated recognition of multiple weak signals, resulting in a network of interactions across exons as well as across introns [Fig. 1B, Berget 1995; Reed 1996]. It is also clear that different constitutive exons are recognized by different mechanisms and require different sets of auxiliary elements in addition to the classical splice-site sequences. The significance of these observations is threefold. First, there are a considerable number of disease-causing mutations in exons or introns that disrupt previously unrecognized auxiliary cis-elements as well as the well-known classical splice sites [Fig. 1C]. Second, because exons differ in their requirements for recognition, mutations that disrupt the function of the
splicing machinery will have different effects on different subsets of exons. Third, variability in the basal splicing machinery among different cell types could cause cell-specific sensitivities to individual splicing mutations.

**Alternative splicing**

Alternative splicing is the joining of different 5' and 3' splice sites, allowing individual genes to express multiple mRNAs that encode proteins with diverse and even antagonistic functions. Up to 59% of human genes generate multiple mRNAs by alternative splicing (Lander et al. 2001), and ∼80% of alternative splicing results in changes in the encoded protein (Modrek and Lee 2002), revealing what is likely to be the primary source of human proteomic diversity. Alternative splicing generates segments of mRNA variability that can insert or remove amino acids, shift the reading frame, or introduce a termination codon [Fig. 2]. Alternative splicing also affects gene expression by removing or inserting regulatory elements controlling translation, mRNA stability, or localization.

A large fraction of alternative splicing undergoes cell-specific regulation in which splicing pathways are modulated according to cell type, developmental stage, gender, or in response to external stimuli. In the best characterized models of vertebrate cell-specific alternative splicing, regulation is mediated by intronic repressor and activator elements distinct from the classical splicing sequences. Cell specificity emerges primarily from two features: First, the repression of splicing in the inappropriate cell type is combined with activation of splicing in the appropriate cell type; and, second, combinatorial control is exerted by multiple components involving cooperative assembly of activation and/or repression complexes on the cis-acting elements surrounding the regulated splice sites (Grabowski 1998; Smith and Valcarcel 2000). The straightforward model is that these complexes serve to enhance or inhibit recognition of the classical splice sites by the basal splicing machinery. Activating and repressing activities coexist within cells (Charlet et al. 2002a), and it remains unclear why activation dominates in one cell type whereas repression dominates in another. Importantly, mutations that perturb this balance can result in aberrant regulation of alternative splicing, causing the expression of protein isoforms that are inappropriate for a cell type or developmental stage.

**Human disease caused by disruption of pre-mRNA splicing**

To define the diverse mechanisms by which defects in pre-mRNA splicing result in a primary cause of disease, we have classified splicing mutations into four categories [Fig. 3]. These categories are based on two criteria. First, does the mutation affect expression of a single gene by disrupting a cis-acting element, or does the mutation have an effect in trans on multiple genes by disrupting a component of the splicing machinery or of a splicing regulatory complex? Second, does the mutation cause aberrant splicing (expression of unnatural mRNAs) by creating unnatural splicing patterns or aberrant regulation of splicing (the inappropriate expression of natural mRNAs) by disrupting use of alternatively used splice sites?

Cis-acting mutations can affect the use of constitutive splice sites [Fig. 3A] or alternative splice sites [Fig. 3B]. Disrupted constitutive splicing most often results in loss of gene expression due to aberrant splicing [see below]. On the other hand, a cis-acting mutation that inactivates
(or activates) one of two alternatively used splice sites will force expression of one of the alternative splicing patterns. Although a natural mRNA is expressed, its expression in an inappropriate tissue or developmental stage might result in disease.

Trans-acting splicing mutations can affect the function of the basal splicing machinery [Fig. 3C] or factors that regulate alternative splicing [Fig. 3D]. Mutations that affect the basal splicing machinery have the potential to affect splicing of all pre-mRNAs, whereas mutations that affect a regulator of alternative splicing will affect only the subset of pre-mRNAs that are targets of the regulator. Each of these four categories are described in the remainder of the review.

Cis effects: mutations that disrupt use of constitutive splice sites

The majority of mutations that disrupt splicing are single nucleotide substitutions within the intronic or exonic segments of the classical splice sites [Fig. 1C]. These mutations result in either complete exon skipping, use of a nearby pseudo 3' or 5' splice site, or retention of the mutated intron. Mutations can also introduce a new splice site within an exon or intron. In rare cases, mutations that do not disrupt or create a splice site activate pre-existing pseudo splice sites distal from the mutation [Pagani et al. 2002], consistent with the proposal that introns contain splicing-inhibitory sequences [Fairbrother and Chasin 2000]. In most cases, use of unnatural splice sites or intron retention introduces premature termination codons (PTCs) into the mRNA, typically resulting in degradation by nonsense-mediated decay [see text]. Therefore, introduction of a premature termination codon into an mRNA by alternative splicing can be a mechanism to down-regulate expression of a gene. Alternative terminal exons: The 3' end of an mRNA is determined by a directed cleavage event followed by addition of the poly[A] tail [Proudfoot et al. 2002]. Selection of one of multiple terminal exons [7] results from a competition between cleavage at the upstream poly[A] site or splicing to the downstream 3' splice site. There are also examples of competition between a 5' splice site and a poly[A] site within an upstream terminal exon [8]. Variability at the 3' end of the mRNA produces either proteins with different C termini or mRNAs with different 3'-UTRs.

**Figure 2.** Alternative splicing generates variable segments within mRNAs. Alternative promoters: Selection of one of multiple first exons results in variability at the 5' terminus of the mRNA [1]. The determinative regulatory step is selection of a promoter rather than splice-site selection. The effect on the coding potential depends on the location of the translation initiation codon. If translation initiates in at least one of the first exons, the encoded proteins will contain different N termini. Alternatively, if translation initiates in the common exon, the different mRNAs will contain different 5' untranslated regions but encode identical proteins. Red indicates variable regions within the mRNA and encoded protein. Alternative splicing of internal exons: Alternative splicing patterns for internal exons include cassette [2], alternative 5' splice sites [3], alternative 3' splice sites [4], intron retention [5], and mutually exclusive [6]. The variable segment within the mRNA results from insertion/deletion, or a mutually exclusive swap. The effects on coding potential are an in-frame insertion or deletion, a reading-frame shift, or introduction of a stop codon. mRNAs containing a stop codon >50 nt upstream of the position of the terminal intron are degraded by nonsense-mediated decay [see text]. Therefore, introduction of a premature termination codon into an mRNA by alternative splicing can be a mechanism to down-regulate expression of a gene. Alternative terminal exons: The 3' end of an mRNA is determined by a directed cleavage event followed by addition of the poly[A] tail [Proudfoot et al. 2002]. Selection of one of multiple terminal exons [7] results from a competition between cleavage at the upstream poly[A] site or splicing to the downstream 3' splice site. There are also examples of competition between a 5' splice site and a poly[A] site within an upstream terminal exon [8]. Variability at the 3' end of the mRNA produces either proteins with different C termini or mRNAs with different 3'-UTRs.

**Pre-mRNA splicing and human disease**
sequences, the only splicing elements widely recognized at the time. It is now known that widespread aberrant splicing is also caused by mutations that disrupt exonic splicing elements (ESEs and ESSs; Fig. 1C). Given recent predictions that the majority of human exons contain ESEs (Liuet al. 2001; Fairbrother et al. 2002), one striking realization is that a significant fraction of exonic mutations that cause disease are unrecognized splicing mutations (for review, see Cooper and Mattox 1997; Caceres and Kornblihtt 2002; Cartegni et al. 2002). The identification of disease-causing mutations is based primarily on linkage of the mutation with the disease phenotype. The effect of the mutation on gene expression is generally assumed based in its location. Because exonic mutations are assumed to cause disease by affecting only the coding potential, silent mutations have been ignored as potential causes of disease, missense mutations have been assumed to create a significant alteration in protein function, and nonsense mutations have been assumed to lead to expression of nonfunctional or deleterious truncated proteins or loss of function caused by NMD. In fact, the primary mechanism of disease in a significant fraction of disease-causing exonic mutations is a catastrophic splicing abnormality rather than a direct effect on coding potential (Cartegni et al. 2002).

The definitive test of whether a disease-causing mutation affects splicing is by direct analysis of mRNA linear structure for correct splicing and mRNA steady-state levels to detect NMD. Ideally, RNA from the affected tissue should be analyzed because cis-acting splicing mutations can have cell-specific effects [Slaugenhaupt et al. 2001]. Unfortunately, the appropriate tissues are often not available to analyze splicing of endogenous mRNAs. As alternatives, mutations that disrupt ESEs or ESSs have been identified using transient transfection of minigenes or in vitro splicing assays comparing splicing of the mutant and wild-type exons (e.g., see McCarthy and Phillips 1998; D’Souza et al. 1999; Paganiet al. 2000; Cartegni and Krainer 2002).

The ability to identify exonic auxiliary splicing elements based on sequence alone would significantly enhance identification of disease-causing mutations. Bonafide mutations could be distinguished from benign polymorphisms and the missense, and nonsense mutations that disrupt ESEs or ESSs could be recognized. As therapies directed toward reverting aberrant splicing patterns become practical, the relevance of identifying splicing mutations will increase. Two major classes of ESEs have been defined based on nucleotide composition: purine-rich and A/C-rich (Cooper and Mattox 1997). The purine-rich ESEs are recognized by a conserved family of serine/arginine-rich [SR] proteins that recruit spliceosome components [such as U2 auxiliary factor, U2AF] to the splice sites [Fig. 1B; Blencowe 2000]. ESEs can also enhance splicing by inhibiting adjacent ESSs [Kan and Green 1999; Zhu et al. 2001]. The A/C-rich ESEs [ACEs] bind

Figure 3. Four classes of pre-mRNA splicing defects that cause disease. (A) Cis-acting mutations that disrupt use of constitutive splice sites: Mutations that disrupt classical splicing signals of a constitutive exon are the most common cause of human disease due to a primary defect in pre-mRNA splicing. The result is expression of unnatural mRNAs, and most often loss of function of the mutated allele due to nonsense-mediated decay (NMD) or expression of proteins containing internal deletions, a shift in the reading frame, or C-terminal truncations. (B) Cis-acting mutations that disrupt use of alternative splice sites: Cis-acting mutations that cause disease by disrupting alternative splicing have been described for four different genes [see Fig. 4]. (C) Trans-acting mutations that disrupt the basal splicing machinery: Two diseases are known to be caused by mutations that affect the function of the basal splicing machinery. (D) Trans-acting mutations that disrupt splicing regulation: Regulation of alternative splicing is disrupted in several forms of cancer and the trinucleotide repeat disorder, myotonic dystrophy.
the cold-box protein, YB-1, and promote splicing by an undetermined mechanism (Coulter et al. 1997; Stickeler et al. 2001).

Several complementary approaches are being used to identify additional auxiliary splicing elements (for review, see Ladd and Cooper 2002). A recent computational analysis of human genomic sequence identified 10 ESEs, 5 of which are novel, by analyzing hexameric sequences enriched in exons that are flanked by weak splice sites. All 10 ESEs functioned autonomously to enhance splicing of a weak exon in vivo (Fairbrother et al. 2002). In a different approach, preferred ESE targets for four individual SR proteins were identified using functional systematic evolution of ligands by exponential enrichment (SELEX; Liu et al. 1998, 2000). The consensus sequences derived from these experiments were used to develop an ESE prediction program (at http://exon.cshl.org/ESE), which has subsequently been used to identify ESE mutations that cause pathogenic splicing abnormalities in four genes including breast cancer susceptibility genes, BRCA1 and BRCA2 (Liu et al. 2001; Fackenthal et al. 2002), and the SMN2 gene, which plays a role in spinal muscular atrophy (SMA; Cartegni and Krainer 2002; see below). Cartegni and Krainer (2002) predicted that 50% of exonic mutations that cause exon skipping disrupt binding sites for one of the four SR proteins used for functional SELEX. The corollary prediction is that the other 50% disrupt binding sites for other proteins. Analyses of ESEs and the mechanism of ESE-mediated splicing have focused on purine-rich SR protein-binding sites. Given that additional ESEs continue to be identified, it is likely that the diversity as well as the number of ESEs relevant to human disease have been underestimated.

Cis effects: mutations that disrupt use of alternative splice sites

Pre-mRNA mutations that affect the use of an alternative splice site shift the ratio of natural protein isoforms (Fig. 3B) rather than create an aberrant splice with the usual associated loss of function. There are four well-characterized examples of such mutations associated with human disease (Fig. 4).

Familial isolated growth hormone deficiency type II (IGHD II)

Postnatal growth in humans requires secretion of growth hormone (GH) from the anterior pituitary. Familial isolated GH deficiency type II (IGHD II) is a dominantly
inherited disorder caused by mutations in the single GH gene (GH-1), in which the main symptom is short stature (Cogan et al. 1994). GH-1 contains five exons and generates a small amount (5%-10%) of alternatively spliced mRNAs (Lecomte et al. 1987). Full-length GH protein is 22 kD, whereas use of an alternative 3′ splice site that removes the first 45 nt of exon 3 and skipping of exon 3 generate 20-kD and 17.5-kD isoforms, respectively [Fig. 4A]. All IGHD II mutations cause increased alternative splicing of exon 3 by disrupting one of three splicing elements: an ISE, an ESE, or the 5′ splice site [Fig. 4A; Binder et al. 1996; Cogan et al. 1997; Moseley et al. 2002]. The natural functions of the 17.5-kD and 20-kD proteins are unknown, but dominant inheritance is thought to result from a dominant-negative effect of the truncated proteins on secretion (Binder et al. 1996). The ISE was first identified by two independent GH-1 mutations located in the intron downstream from exon 3 [Fig. 4A; Cogan et al. 1997], and analysis of transiently expressed GH-1 minigene constructs demonstrated ISE activity in vivo [McCarthy and Phillips 1998]. One IGHD II mutation is a G → A substitution within one of two adjacent G triplets. The other is a deletion that removes both G triplets. The G triplets disrupted by these mutations are similar to a regulatory element identified in the chicken β-tropomyosin gene [Sirand-Pugnet et al. 1995]. The association of G triplets with 5′ splice sites was identified computationally early on [Nussinov 1988; Engerbret et al. 1992], and these elements have been shown to recruit U1 [McCullough and Berget 2000]. A G → A substitution in the fifth nucleotide of exon 3 was recently linked to disease in an IGHD II family and was shown to disrupt an ESE [Fig. 4A; Moseley et al. 2002]. The ESE mutations caused exon skipping as well as enhanced use of the alternative 3′ splice site within exon 3. Finally, a G → A mutation in the first nucleotide of intron 3 disrupts the 5′ splice site and causes complete exon skipping and expression of the 17.5-kD isoform [Fig. 4A; Cogan et al. 1997].

Frasier syndrome

Inactivation of the Wilms tumor suppressor gene [WT1] is responsible for ~15% of Wilms tumors, a pediatric cancer of the kidney [Call et al. 1990; Gessler et al. 1990]. Three additional disorders are associated with abnormalities in WT1 expression: WAGR [Wilms tumor, aniridia, genitourinary abnormalities, mental retardation], Denys-Drash syndrome [DDS], and Frasier syndrome [FS]. All three diseases are characterized by urogenital defects involving kidney and gonad developmental defects. Consistent with these defects, normal expression patterns of human WT1 during development indicate important roles in kidney and gonad development [Armstrong et al. 1993], and WT1-null mice lack gonads and kidneys [Kreidberg et al. 1993].

The human WT1 pre-mRNA undergoes extensive alternative splicing; however, the only alternative splice conserved among vertebrates is the use of two alternative 5′ splice sites for exon 9 separated by 9 nt that encode lysine–threonine–serine (KTS; Fig. 4B; Miles et al. 1998). The +KTS and −KTS isoforms are expressed at a constant ratio favoring the +KTS isoform in all tissues and developmental stages that express WT1 [Haber et al. 1991]. The majority of individuals with FS were found to have mutations that inactivate the downstream 5′ splice site, resulting in a shift to the −KTS isoform [Fig. 4B; Barbaux et al. 1997; Kohsaka et al. 1999; Melo et al. 2002].

The WT1 protein contains four C2H2 zinc fingers at its C terminus and a proline/glutamine-rich N-terminal region. The variable KTS region is located between the third and fourth zinc fingers. Mouse models that express only the endogenous −KTS or +KTS isoforms provide a striking demonstration of the functional differences between the two nearly identical isoforms [Hammes et al. 2001]. The properties of the two WT1 isoforms also indicate that they perform distinct functions. The −KTS isoform trans-activates transcription of genes involved in early gonad development including S1 and Sry [Hosain and Saunders 2001; Wilhelm and Englert 2002]. In contrast, the +KTS isoform binds DNA only weakly, and is unable to activate targets of the −KTS isoform [Wilhelm and Englert 2002]. The +KTS isoform appears to function in RNA metabolism, perhaps pre-mRNA splicing. Whereas −KTS shows diffuse nuclear localization, +KTS colocalizes in nuclear speckles, which are thought to be storage areas for components of the basal splicing machinery [Larsson et al. 1995; Davies et al. 1998]. In addition, +KTS binds U2AF65, an essential splicing factor involved in the early steps of exon recognition.

Because FS is dominantly inherited, affected individuals have a wild-type allele expressing the normal ratio of +KTS/−KTS. The ratio of the two isoforms is critical [Barbaux et al. 1997].

Frontotemporal dementia and Parkinsonism linked to Chromosome 17 (FTDP-17)

Aggregation of the microtubule-associated protein tau into neuronal cytoplasmic inclusions is associated with several neuropathological conditions characterized by progressive dementia including Alzheimer’s disease, Pick’s disease, and frontotemporal dementia and Parkinsonism linked to Chromosome 17 (FTDP-17; Buée et al. 2000). FTDP-17 is an autosomal dominant disorder caused by mutations in the MAPT gene that encodes tau. Tau is required for microtubule assembly and function and is thought to play a major role in microtubule-dependent transport in axons. Free tau, not bound to microtubules, is proposed to be subject to hyperphosphorylation and aggregation.

Since the initial discovery in 1998 that MAPT mutations cause FTDP-17, at least 16 mutations have been identified in 50 FTDP-17 families. MAPT mutations fall into two mechanistic classes. One class includes mutations that alter the biochemical properties of the protein.
In vitro analysis of these mutant proteins demonstrated either altered ability to modulate microtubule polymerization or enhanced self-aggregation into filaments that resemble neurofibrillary tangles. A second class of disease-causing mutations that affected splicing was revealed by mutations clustered in and around the alternatively spliced exon 10 (Fig. 4C). A primary role for splicing defects was indicated by the discovery of 5′ splice site mutations and the observation that not all exon 10 missense mutations altered tau function in vitro. Subsequently, silent mutations in exon 10 were linked to FTDP-17, ruling out expression of a mutated protein as the pathogenic event (Hong et al. 1998; D’Souza et al. 1999). Exon 10 encodes the last of four microtubule-binding domains, and exon 10 inclusion is determinative for the ratio of the 4R-tau and 3R-tau protein isoforms (4R and 3R designate four and three microtubule-binding domains, respectively). The normal 4R/3R ratio is 1, and some FTDP-17 mutations alter this ratio by as little as twofold, which indicates that a strict balance is required for either normal tau function or to prevent tau aggregation.

The 4R/3R ratio is maintained by a complex set of intronic and exonic splicing elements surrounding and within exon 10 including ESEs, ESSs, ISSs, and a putative hairpin structure that sequesters the 5′ splice site [Fig. 4C]. The vast majority of FTDP-17 mutations affect these regulatory elements and cause disease by increasing inclusion of exon 10. As expected, the 4R-tau protein isoform predominates in the insoluble tau aggregates in individuals with FTDP-17 (Hutton et al. 1998; Spillantini et al. 1998). However, not all FTDP-17 mutations are expected to increase the 4R/3R ratio. One mutation, a 3-nt deletion within the 5′ ESE (ΔK280), results in complete exon skipping in a minigene construct, presumably because it weakens an ESE [D’Souza et al. 1999]. This mutation also decreases the 4R-tau protein function in vitro, although the biochemical properties of the recombinant 4R protein are irrelevant if the exon is completely skipped in affected individuals. Unfortunately, the level of exon 10 inclusion in these individuals is unknown because tissue samples are not available.

**Atypical cystic fibrosis**

Cystic fibrosis [CF] is an autosomal recessive disorder caused by loss of function of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. CFTR encodes a cAMP-dependent transmembrane chloride channel that is expressed in secretory epithelium. In the USA, more than two-thirds of individuals affected with CF carry the devastating ΔF508 mutation, which causes a failure of the protein to localize to the apical plasma membrane. Fifty percent of affected individuals are homozygous for this allele, resulting in severe pulmonary and pancreatic disease. However, less frequent, “milder” mutations that retain residual CFTR function are responsible for a range of CF-related disorders including late onset or less severe pulmonary disease, male infertility due to congenital bilateral absence of the vas deferens [CBAVD], and chronic idiopathic pancreatitis [Noone and Knowles 2001].

Two polymorphisms in the CFTR gene that contribute to atypical CF phenotypes are located at the 3′ end of intron 8 and directly affect splicing of exon 9 [Fig. 4D]. One is a variant polyuridine tract containing 5, 7, or 9 uridines within the polypyrimidine tract of intron 8. The second is a polymorphic poly[UG] tract immediately upstream of the [U] tract. Both polymorphisms are located between the presumptive branch site for intron 8 and the AG-terminal dinucleotide. Nearly all individuals express a small fraction of CFTR mRNAs that lack exon 9 and express a nonfunctional protein [Delaney et al. 1993; Strong et al. 1993]. It is unclear whether this alternative splice serves a purpose.

The shortest [U] allele, 5U, can be associated with a high level of exon skipping in respiratory epithelial cells compared with the 7U and 9U alleles. The frequency of 5U carriers is estimated to be 10% worldwide [Kiesewetter et al. 1993]. Some individuals homozygous for the 5U allele skip exon 9 in >95% of CFTR mRNAs in lung epithelium [Chu et al. 1992]. This polymorphism is rarely sufficiently penetrant to be associated with a severe CF phenotype [Noone et al. 2000]; however, many individuals affected with CBAVD are compound heterozygotes for the 5U allele with a severe CFTR mutation. Some individuals with CBAVD are 5U homozygotes [Chillon et al. 1995], indicating that the 5U allele alone can cause disease and the disease in these individuals correlates with the level of exon 9 skipping and the subsequent loss of CFTR function [Larriba et al. 1998].

On the other hand, the identification of healthy 5U homozygotes demonstrated that the penetration of the 5U allele is quite variable. Variable penetrance is explained in part by the second polymorphic [UG]n tract which ranges in size from [UG]7 to [UG]13. Longer UG tracts are associated with higher disease penetrance and increased skipping of exon 9 in individuals with CBAVD. In fact, healthy fathers of individuals affected with CBAVD have been shown to contain shorter [UG]n polymorphisms and exhibit less exon 9 skipping than their affected sons, explaining the variable penetrance within some families [Cuppens et al. 1998]. Transient transfection analysis of CFTR minigenes directly demonstrated that the longer [UG]n tract correlates with increased exon 9 skipping, but only when combined with the 5U allele [Niksic et al. 1999]. A protein of unknown function, TDP-43, binds to the [UG]n tract and inhibits exon 9 inclusion [Buratti et al. 2001]. The prediction, thus far untested, is that the polypyrimidine tract of the 5U allele binds U2AF65 poorly compared with the 7U and 9U alleles and that this interaction is negatively affected by binding of TDP-43 to the upstream [UG]n tract [Buratti et al. 2001].

**Trans effects: mutations that affect the basal splicing machinery**

There are several genetic diseases in which a mutation disrupts the machinery of splicing, either the constitu-
Retinitis pigmentosa

Retinitis pigmentosa (RP) is a heterogeneous disease affecting 1 in 4000 individuals characterized by progressive retinal degeneration, night blindness, loss of peripheral vision, and ultimately total blindness. The disease results from the specific loss of rod photoreceptor cells. RP can be inherited as an autosomal dominant, autosomal recessive, or X-linked disorder. More than 30 different RP genes and loci have been identified, most of which have retina-specific functions. However, within the past year and a half, three genes responsible for autosomal dominant RP [PRPF31, HPRP3, and PRPC8] have been identified as the human orthologs of the yeast genes PRP31, PRP3, and PRP8, respectively (McKie et al. 2001; Vithana et al. 2001; Chakarova et al. 2002). All three yeast genes are involved in the function of the U4/U6 · U5 tri-snRNP, the spliceosome component required for the transition to a catalytically active state. All three human proteins were found in isolated functional spliceosomal snRNPs.

Pathogenic mutations in PRPF31 have been identified in four RP families and three sporadic cases of autosomal dominant RP [Vithana et al. 2001]. The mutations include insertions, deletions, missense mutations, and splice-site mutations. It is likely that in at least some mutant alleles, the function of PRPF31 is severely affected, if not completely eliminated. Therefore, PRPF31 mutations are likely to cause autosomal dominant RP due to haploinsufficiency, although a dominant-negative effect for alleles expressing truncated proteins cannot be ruled out.

Of the three splicing-factor genes that cause RP, the function of PRPF31 is best defined. Prp31p is an essential splicing factor in both Saccharomyces cerevisiae (60% similarity to human) and Schizosaccharomyces pombe (68% similarity to human; Weidenhammer et al. 1996; Bishop et al. 2000). Human PRPF31 is a U4/U6 snRNP-associated protein that promotes association between U4/U6 snRNP and U5 snRNP by direct interactions with a 102-kD U5-specific protein. In vitro splicing using HeLa cell nuclear extracts immunodepleted of PRPF31 showed accumulation of the prespliceosome complex [containing U1 and U2 snRNPs] by preventing association of the U4/U6 · U5 tri-snRNP and assembly of the active spliceosome. Addition of recombinant PRPF31 reversed this inhibition, demonstrating that the deficiency of PRPF31 was responsible for the block (Makarova et al. 2002). Although several mutations in snRNP proteins inhibit the spliceosome-to-spliceosome transition, PRPF31 is unique in that it directly mediates formation of the U4/U6 · U5 tri-snRNP rather than direct interactions between the tri-snRNP and the prespliceosome. The ability to deplete PRPF31 and then reconstitute PRPF31-dependent splicing provides a powerful in vitro assay to test the intrinsic activities of PRPF31-disease-causing mutations.

Mutations in HPRP3 have been shown to cause RP in three families and two sporadic cases (Chakarova et al. 2002). All five examples are caused by one of two missense mutations in two highly conserved adjacent codons in exon 11. This protein domain is unique in the database, and its specific function is unknown. Like PRPF31p, HPRP3 is a component of the U4/U6 snRNP (Wang et al. 1997). In mammals, HPRP3 is thought to recruit HPRP4 to the U4/U6 snRNP (Gonzalez-Santos et al. 2002). HPRP3 and HPRP4 are homologs of the yeast U4/U6-snRNP-specific proteins. In yeast, PRP3 and PRP4 genetically interact, and physical interactions between Prp3p and Prp4p proteins are required for association of Prp3p and Prp4p with U4/U6 (Ayadi et al. 1998).

Mutations in PRPC8 cause a severe form of RP [McKie et al. 2001]. Seven different mutations have been identified in three RP families and four individuals with a history of autosomal dominant RP. All of these mutations cluster in a highly conserved 14-amino-acid region in the last exon. PRPC8 encodes PRP8, a 220-kD core component of the U5 snRNP. The PRP8 protein is highly conserved, being 62% identical in human and S. cerevisiae throughout its ~2300 residues. PRP8 is known to be an integral component of the spliceosome catalytic core and makes direct contact with both the 5’- and 3’-splice sites and U6 as well as U5 snRNAs (Wyatt et al. 1992; Teigelkamp et al. 1995; Vidal et al. 1999). PRP8 is thought to provide overall structural support for the catalytic core and to modulate the RNA helicase activities that control the extensive RNA:RNA base-pairing rearrangements required to activate the spliceosome (Collins and Guthrie 2000). The remarkable clustering of the mutations identifies a specific functional domain, but it remains to be determined whether these mutations inactivate the allele or create a protein with dominant-negative function. The possible basis for the striking cell-specific effects of the RP mutations is discussed below.

Spinal muscular atrophy

Spinal muscular atrophy (SMA) is an autosomal recessive disorder that is one of the most common genetic causes of childhood mortality. The main characteristic of the disease is progressive loss of spinal cord motor function, resulting in a failure to develop the forces needed to support the head and trunk. SMA is caused by mutations in the survival motor neuron (SMN) gene, which encodes a protein required for the survival of motor neurons. The disease is characterized by progressive weakness and wasting of the voluntary muscles, leading to respiratory failure and death in untreated cases. The severity of the disease ranges from mild to severe, with type I SMA being the most severe form and type II SMA being the mildest form. The onset of symptoms typically occurs within the first few months of life, and the disease progresses rapidly in type I SMA, often leading to death within the first few years. In type II SMA, the disease usually becomes apparent between 6 and 18 months of age and progresses more slowly, allowing some affected individuals to live into adulthood. The prevalence of SMA is estimated to be around 1 in 6000 live births, with a carrier frequency of about 1 in 40 in the general population. The disease is caused by a variety of genetic mutations in the SMN gene, including deletions, duplications, and point mutations. The natural history of SMA is characterized by a progressive decline in muscle strength and function, leading to significant morbidity and mortality. The disease affects multiple organ systems, including the muscles, respiratory system, and gastrointestinal tract. The mainstay of treatment for SMA is supportive care, including respiratory therapy and nutritional support. There is no cure for SMA, but advances in genetic testing and counseling have helped to improve understanding of the disease and support affected families.
neurons, resulting in skeletal muscle denervation with subsequent weakness, atrophy, and paralysis of voluntary muscles. The SMA locus maps to a complex inverted repeat of ~500 kb on Chromosome 5q13 that contains several genes. The cause of SMA in 96% of cases is homozygous loss of the telomeric copy of the survivor of motor neuron gene (SMN1) located within the inverted repeat [Wirth 2000]. A duplicated gene within the centromeric copy of the inverted repeat (SMN2) is also transcribed and contains only a few nucleotide substitutions, none of which alters the protein coding sequence. Despite the potential to encode the identical protein, the SMN2 gene does not completely compensate for loss of SMN1 function because one of the nucleotide substitutions disrupts an ESE in exon 7 that causes the exon to be skipped in the majority of SMN2 mRNAs [Cartegni and Krainer 2002].

The resulting SMN2 ΔE7 mRNA encodes a truncated protein missing the C-terminal 16 residues and is thought to be nonfunctional [Cifuentes-Díaz et al. 2001]. SMN is a ubiquitously expressed 294-amino-acid protein that is essential in S. pombe [Owen et al. 2000] and is required for cell viability in vertebrates [Wang and Dreyfuss 2001]. The specific functions of SMN are unknown, but it is in a complex [Baccon et al. 2002] that interacts with components of several RNP complexes with diverse functions, which suggests that SMN acts as a “master assembler” of RNP complexes [Terns and Terns 2001]. The best characterized role for the SMN complex is in the assembly of U1, U2, U4, and U5 snRNPs, which contain a common set of seven Sm proteins as well as sets of proteins unique to each snRNP. SnRNP assembly begins with export of nascent snRNA to the cytoplasm, where Sm proteins assemble as a ring around a 9-nt Sm-binding site on each snRNA, forming the so-called core snRNPs. The assembled Sm proteins plus a trimethylguanosine (m3G) cap at the snRNA 5′ end serve as a bipartite nuclear localization signal. Once in the nucleus, snRNP-specific proteins are added to the core snRNPs to form active snRNPs [Will and Lührmann 2001].

SMN is required for the cytoplasmic assembly of the core snRNPs. Immunodepletion of SMN plus a tightly associated integral component of the SMN complex, Gemin2, prevented U1 snRNP assembly in Xenopus oocyte extracts despite the presence of abundant Sm proteins. Assembly was restored by adding back purified SMN complex [Meister et al. 2001]. Overexpression of an N-terminal truncation mutant of SMN (SMNΔN27) with dominant-negative activity resulted in cytoplasmic coaccumulation of Sm proteins, SMN, and U snRNAs. The snRNA did not contain the m3G cap, suggesting that these accumulations result from arrested snRNP maturation [Pellizzoni et al. 1998]. Native SMN complexes purified from cells have recently been shown to be necessary and sufficient to promote ATP-dependent assembly of core snRNPs in vitro. Furthermore, under the conditions used, the SMN complex was required to prevent binding of Sm proteins to non-U snRNAs [Pellizzoni et al. 2002]. Although purified Sm proteins assemble on snRNAs in an ordered pathway in vitro in the absence of non-snRNP factors [Raker et al. 1996], the SMN complex could be required for efficient core assembly in the complex cellular environment or to regulate snRNP assembly in response to cellular metabolism.

Four clinical types of SMA have been defined based on age of onset and disease severity, which ranges from intrauterine demise to mild symptoms in older individuals. Results from individuals affected with SMA and SMA mouse models demonstrate that there is a clear correlation between SMN protein levels, loss of motor neurons, and disease severity [Coovert et al. 1997; Lelebvre et al. 1997; Jablonka et al. 2000]. Because both copies of SMN1 are missing in most individuals with SMA, the only source of full-length SMN protein is the small fraction of SMN2 mRNAs that include exon 7. Quantification of SMN2 gene number using real-time PCR showed that individuals with the less severe type III typically have multiple copies of the SMN2 gene through gene replacement and duplication [Feldkötter et al. 2002]. Therefore, the effects of a primary loss of SMN1 are ameliorated by the small amount of full-length SMN protein encoded by each copy of the SMN2 gene.

A strong correlation between the loss of motor neurons and the reduction of nuclear staining for SM-containing snRNPs in mouse models of SMA strongly suggests that the SMN deficiency causes disease by a defect in pre-mRNA splicing. Unlike humans, mice have only one Smn gene. Smn−/− mice die at the blastocyst stage (Schrank et al. 1997), and Smn−/− mice develop symptoms strikingly similar to SMA [Jablonka et al. 2002]. The diffuse staining of cytoplasmic SMN was reduced in spinal neurons of Smn−/− mice, and nuclear anti-Sm immunofluorescence (for nuclear snRNPs) was reduced by 39% [Jablonka et al. 2000]. In addition, Smn−/− and Gemin2−/− double heterozygotes had a 61% reduction in nuclear Sm staining correlating with substantially increased motor neuron loss compared with Smn−/− mice [Jablonka et al. 2002].

Smn−/− mice are normal at birth but develop SMA-like symptoms within days owing to a normal developmentally regulated decline in which SMN protein levels in the spinal cord drop to <50% of fetal levels, primarily between postnatal days 5 and 15 [Hsieh-Li et al. 2000; Jablonka et al. 2000; Monani et al. 2000]. This downregulation also occurs in humans [Burl et al. 1998], and individuals with type III SMA display a worsening of symptoms that correlates with this down-regulation of SMN protein. The drop in SMN protein occurs in several tissues that are unaffected in SMA despite the fact that SMN protein levels are lower in these than in the spinal cord [Lelebvre et al. 1997; Burl et al. 1998]. A muscle-specific knockout of Smn induces severe muscular dystrophy, indicating that substantial reduction of SMN will induce intrinsic muscle disease [Cifuentes-Díaz et al. 2001]. These results indicate that postnatal motor neurons require higher steady-state levels of SMN protein than other metabolically active tissues.
What is the basis for cell-specificity in RP and SMA?

For both RP and SMA, the primary defect appears to be a loss of function of essential splicing factors, although dominant-negative function for some RP alleles cannot be ruled out. How can the loss of ubiquitous functions result in such remarkable cell-specific sensitivity? Because exons are diverse units of recognition, different exons are likely to exhibit a wide range of sensitivities to deficiencies of essential splicing factors. Perhaps only a subset of pre-mRNAs (or even one pre-mRNA) required for rod cell or motor neuron viability is affected by deficiencies in the U4/U6·U5 tri-snRNP or SMN function, respectively. It can also be argued that cell-specific pre-mRNAs are more likely to be affected by a deficiency of a basal splicing factor than pre-mRNAs that are widely expressed. In contrast to cell-specific pre-mRNAs, widely expressed pre-mRNAs must have the ability to undergo efficient splicing in a variety of nuclear environments and presumably contain information in cis for more robust splicing. The few essential splicing factors that have been examined in vertebrates show surprisingly variable levels of expression among different tissues that do not correlate with tissue metabolic activity. For example, SF1, a spliceosome component involved in the initial recognition of the branch site, is barely detectable in pancreas, kidney, and lung, whereas PRP8 is barely detectable in liver (Luo et al. 1999; Vervoort et al. 2000).

Even in yeast, where intron recognition is highly homogenous, loss-of-function phenotypes for PRP2 and CEF1 are due to defective removal of single introns (Chen et al. 1998; Burns et al. 2002). For example, a screen for mutants that disrupt transport of secretory proteins from the endoplasmic reticulum [ER] to the Golgi identified a well-characterized essential splicing factor, PRP2. PRP2 is an RNA-dependent ATPase required for the first transesterification reaction. The protein secretion defect of the PRP2 mutant was found to be caused by inefficient splicing of the intron of the SAR1 gene, which encodes a small GTPase required for ER vesicle formation. The PRP2 protein secretion defect was suppressed by overexpressing the SAR1 cDNA or by removing the SAR1 gene intron (Chen et al. 1998). Similarly, Cef1p [CDC5 in S. pombe] was genetically identified as a cell cycle regulator (Ohk et al. 1994), and a role in pre-mRNA splicing was subsequently found by genetic and biochemical analyses (for review, see Burns et al. 2002). Global analysis of splicing in Cef1p mutants using an oligonucleotide array (Clark et al. 2002) demonstrated a significant general splicing defect. Despite this, the G2/M block was relieved by replacing the α-tubulin gene with the α-tubulin cDNA, demonstrating that failure to remove the single α-tubulin intron is primarily responsible for the CEF1 loss-of-function phenotype (Burns et al. 2002).

The opsin pre-mRNA is one potential target of the presumed tri-snRNP deficiency in RP. The opsin protein binds covalently to a chromophore to form the photopigment rhodopsin, which undergoes a conformational change in response to photons that initiates the photodetection cascade (Bessant et al. 2001). Rhodopsin is embedded in the extensive array of membranous discs present in each rod cell. The discs undergo daily renewal just prior to waking (Korenbrot and Fernald 1989), putting considerable demand on the splicing machinery to produce huge amounts of opsin mRNA. Insufficient production of rhodopsin caused by opsin gene mutations also causes dominantly inherited RP, consistent with the proposal that PRPF31, HPRP3, and PRPC8 mutations result in a rhodopsin deficiency secondary to a splicing defect.

Analogous potential targets required for motor neuron viability in SMA are less obvious. It is possible that one or only a few pre-mRNAs are affected by the sequence of events resulting from reduced assembly of core snRNPs. It is also possible, however, that the pathogenic mechanism in SMA is more complex than a loss of snRNPs. If an SMN deficiency results in promiscuous association of Sm proteins with inappropriate RNAs in vivo as it does in vitro (Pellizzoni et al. 2002), a loss of function of those RNAs or a gain of function of the aberrant complex could contribute to pathogenesis.

Trans effects: mutations that affect regulators of alternative splicing

Three mouse models illustrate the deleterious effects resulting from the loss of factors that regulate splicing (SC35, QKI-5, and Nova-1; Jensen et al. 2000; Wang et al. 2001; Wu et al. 2002). All three examples illustrate that inactivation of a splicing regulator in mice specifically affects its natural pre-mRNA targets. The same specificity is expected in human diseases caused by disrupted function of alternative splicing regulators.

Myotonic dystrophy

Myotonic dystrophy [DM] is the one human disease in which disease phenotype has been directly linked to disrupted regulation of alternative splicing (Fig. 3D). DM is an autosomal dominant disorder and the most common form of adult-onset muscular dystrophy, with a worldwide incidence of 1 in 8000. DM is unusual because of its phenotypic variability even within families and the diversity of tissues affected. Symptoms include skeletal muscle hyperexcitability [myotonia], progressive muscle wasting, cardiac conduction defects, cataracts, smooth muscle dysfunction, testicular atrophy, an unusual form of insulin resistance, and neuropsychiatric and cognitive disturbances (Harper 2001). Two types of DM have been identified. The most common form is type 1 [DM1], which is caused by a CTG expansion in the 3′ untranslated region [UTR] of the DM protein kinase [DMPK] gene located on Chromosome 19q13.3. Disease severity and age of onset correlate with repeat length, which ranges from 80 to thousands of repeats. Unaffected individuals have fewer than ~40 repeats. DM type 2 [DM2] is caused by a large CCTG expansion in intron 1 of the ZNF9 gene on Chromosome 3q21 [Liquori et al. 2001].
Several independent lines of evidence indicate that the predominant mechanism for DM pathogenesis is a gain of function for RNA transcribed from the expanded alleles. First, no point mutants or deletions within the DM1 or DM2 loci cause DM, indicating that the repeats are deterministic for these diseases rather than a loss of function associated with the DM1 or DM2 loci. Second, the fact that two different loci containing similar expanded repeats cause strikingly similar diseases strongly suggests that DM1 and DM2 share a common pathogenic mechanism that is independent of a loss of function for the affected locus. Third, RNAs containing long tracks of CUG or CCUG repeats are transcribed from the expanded DMPK and ZNF9 alleles, and both repeat-containing RNAs accumulate in discrete nuclear foci [Taneja et al. 1995; Davis et al. 1997; Liquori et al. 2001].

Fourth, transgenic mice [HSA14] expressing 250 CUG repeats in the 3′-UTR of the human skeletal α-actin gene reproduced myotonia and the histopathological features observed in DM1 muscle [Mankodi et al. 2000], demonstrating that expression of CUG repeats independent of the DM1 locus is sufficient to induce major features of the disease.

According to the RNA gain-of-function hypothesis, DM pathogenesis results from disrupted RNA processing secondary to disrupted function of RNA-binding proteins by the expanded RNA repeats [Wang et al. 1995]. Consistent with this hypothesis, five pre-mRNAs have been shown to undergo aberrantly regulated splicing in DM1 tissues and/or mouse models: cardiac troponin T (cTNT), insulin receptor [IR], muscle-specific chloride channel (ClC-1), tau, and myotubularin-related 1 [Philips et al. 1998; Savkur et al. 2001; Seznec et al. 2001; Buj-Bello et al. 2002; Charlet et al. 2002b]. Misregulated splicing of IR and ClC-1 pre-mRNAs is likely to directly cause two common symptoms in individuals affected with DM1. The IR splicing switch observed in DM1 skeletal muscle results in expression of a lower signaling IR isoform directly correlating with the unusual form of insulin resistance observed in individuals with DM1 [Savkur et al. 2001]. Similarly, loss of ClC-1 function secondary to aberrantly regulated splicing is sufficient to account for myotonia, the delayed muscle relaxation following voluntary contraction caused by repeated firing of action potentials. Recent results from individuals with DM1 and HSA14 mice demonstrate that aberrantly regulated splicing of ClC-1 pre-mRNAs introduces PTCs resulting in NMD of the ClC-1 mRNA and ultimately loss of ClC-1 function [Charlet et al. 2002b; Mankodi et al. 2002].

The mechanism by which CUG-repeat RNA induces disease is likely to involve CUG-repeat-binding proteins. Several CUG-repeat-binding proteins have been identified including muscleblind, CUG-binding protein (CUG-BP), elav-type RNA binding protein 3 [ETR-3], which is 78% identical to CUG-BP, and protein kinase R (PKR; Timchenko et al. 1996; Lu et al. 1999; Miller et al. 2000; Tian et al. 2000). The proteins from three human muscleblind genes [Fardaei et al. 2002] are homologs of a protein required for development of muscle and photo-receptor cells in Drosophila [Begemann et al. 1997; Artero et al. 1998]. These proteins contain two Cys-His-type zinc finger domains found in RNA processing and transcription factors. CUG-repeat RNA with more than ∼20 repeats forms double-stranded RNA containing U-U mismatches [Napiera et al. 1997; Michalowski et al. 1999], and muscleblind has a strong affinity for double-stranded CUG-repeat RNA in vitro and colocalizes with the nuclear foci containing CUG- and CCUG-repeat RNA in DM cells [Michalowski et al. 1999; Miller et al. 2000; Fardaei et al. 2001]. Because the function of muscleblind is unknown, the consequences of muscleblind colocalization with CUG-repeat RNA for splicing and DM pathogenesis remain to be determined.

CUG-BP was identified as a protein that bound to a single-stranded synthetic [CUG]6 RNA. In contrast to muscleblind, CUG-BP does not bind double-stranded CUG-repeat RNA [Miller et al. 2000] and does not colocalize with the CUG-repeat RNA in nuclear foci [Michalowski et al. 1999; Fardaei et al. 2001]. Although the physical evidence links muscleblind and not CUG-BP with the nuclear foci of CUG-repeat RNA, functional analyses indicate that increased activity of CUG-BP is responsible for the aberrant regulation of cTNT, IR, and ClC-1 alternative splicing observed in DM1. First, CUG-BP is a well-characterized alternative splicing regulator [Ladd et al. 2001]. It is one of six paralogs called CUG-BP and ETR-3 Like Factors (CELF; Ladd et al. 2001) or Bruno-like [Bruno; Good et al. 2000] proteins. Second, the steady-state levels of CUG-BP protein are elevated in DM1 striated muscle tissues where aberrantly regulated splicing has been demonstrated [Savkur et al. 2001; Timchenko et al. 2001]. Third, cTNT, IR, and ClC-1 pre-mRNAs are known targets for CUG-BP regulation [Philips et al. 1998; Savkur et al. 2001; Charlet et al. 2002b]. For all three pre-mRNAs, CUG-BP has been shown to bind to U/G-rich motifs in introns adjacent to the regulated splice sites. Furthermore, overexpression of CUG-BP with cTNT, IR, and ClC-1 minigenes in normal cells induces the splicing patterns observed in DM1 striated muscle, which are different for the different pre-mRNAs (cTNT exon 5 inclusion, IR exon 11 skipping, and ClC-1 intron 2 retention), consistent with the increased steady-state levels observed in DM1 striated muscle. Pre-mRNAs containing mutated CUG-BP-binding sites are no longer regulated by CUG-BP overexpression. The effects of elevated CUG-BP appear to be limited to its natural targets because the ratio of alternatively spliced isoforms of hnRNP A1 is unaffected in DM1 [Philips et al. 1998]. Fourth, overexpression of CUG-repeat RNA with cotransfected cTNT and IR minigenes induced the aberrant splicing patterns observed in DM1 striated muscle [Philips et al. 1998; Savkur et al. 2001]. Minigene pre-mRNAs containing mutated CUG-BP-binding sites did not respond to coexpressed CUG-repeat RNA, demonstrating that CUG-BP, or another protein that binds to the CUG-BP-binding site, mediates the splicing switch induced by CUG-repeat RNA. Fifth, a cTNT minigene expressed in DM1 muscle cultures reproduced the aberrant splicing pattern of endogenous
cTNT, whereas a minigene containing a mutated CUG-BP-binding site was not aberrantly regulated in DM1 muscle cultures [Philips et al. 1998]. Taken together, these results indicate that the aberrant regulation of these targets observed in DM1 skeletal muscle is mediated by CUG-BP or other CELF proteins such as ETR-3 that bind to the intronic regulatory elements.

A general model for the pathogenic mechanism of DM is that expression of CUG- or CCUG-repeat RNA induces overexpression of CUG-BP, resulting in misregulated splicing of its target pre-mRNAs [Fig. 5]. The mechanism by which CUG-repeat RNA induces CUG-BP expression is unknown and could be dependent or independent of binding of muscleblind to CUG-repeat RNA. When CUG-repeat RNA was expressed in COS cells, the half-life of endogenous CUG-BP protein increased greater than twofold [Timchenko et al. 2001], consistent with the increased steady-state levels observed in DM1 striated muscle tissue. The half-life of CUG-BP in DM1 cells remains to be determined. In addition, CUG-BP phosphorylation and nuclear:cytoplasmic distribution are altered in DM1 striated muscle tissues [Roberts et al. 1997], however, the relationship between these changes and the aberrantly regulated splicing observed in these tissues has not yet been established.

For all five pre-mRNAs whose regulated splicing is known to be affected, the splicing patterns switch to an embryonic pattern. The splicing changes observed in skeletal muscle occur without signs of regeneration and therefore are not due to recapitulation of the embryonic program by muscle satellite cells [Savkur et al. 2001]. Instead, there appears to be a programmatic switch that specifically reverts alternative splicing regulation to an embryonic state. An understanding of the process that has gone awry is likely to give information regarding the network of alternative splicing regulation that occurs during development.

Changes in splicing regulation associated with neoplasia and malignancy

The transition from normal cell growth to neoplasia and then to malignancy (which includes anaplasia, invasion, and metastasis) represents a multistep selection for the most aggressive cells. For many genes, dramatic changes in alternative splicing patterns are associated with neoplasia and metastasis [Philips and Cooper 2000; Nissim-Rafinia and Kerem 2002]. Many of the genes affected are those likely to be involved in neoplasia such as regulators of apoptosis, hormones, and receptors mediating cell–cell and cell–matrix interactions. The splicing changes are due to changes in trans-acting regulatory factors because mutations are not detected in the affected genes [Fig. 3D]. The frequent association between splicing regulation and neoplasia or metastasis has raised several questions. First, do alterations in the abundance or activities of splicing regulators contribute to malignant transformation? Second, if so, what genes undergo altered splicing, and what properties of the aberrantly expressed isoforms contribute to malignancy? Third, what regulatory factors have altered activities? Fourth, can alternative splicing be used to identify and subclassify cancer to predict natural outcomes and evaluate treatment responsiveness? Fifth, are either the splicing regulators or pre-mRNAs that are misregulated potential targets of therapeutic intervention?

In 1991, Gunthert et al. [1991] demonstrated a landmark association between an alternative splicing switch of CD44 and the acquisition of metastatic potential in a rat pancreatic adenocarcinoma cell line. CD44 functions in cell adhesion, migration, and cell–matrix interactions, roles well suited for affecting metastatic potential. The CD44 gene expresses a family of cell-surface glycoproteins by alternative splicing of nine internal exons [exons 7–15]. CD44 mRNAs lacking the variable regions (known as CD44 standard or CD44s) predominate in healthy human adult tissues, whereas isoforms containing the variable regions (known collectively as CD44v or CD44v) are expressed in several tissues during development and during T-cell activation [Sneath and Mangham 1998]. Specific CD44 variants, particularly CD44v5, CD44v6, and CD44v7 [variant regions 5, 6, and 7, containing exons 10, 11, 12, respectively], have been associated with many human malignancies [Sneath and Mangham 1998]. Gunthert et al. [1991] identified CD44v6 as a cell-surface marker found in metastatic cell lines but not nonmetastatic cell-line derivatives. Strikingly, overexpression of CD44v6 bestowed metastatic potential to a nonmetastatic cell line [Gunthert et al. 1991]. Monoclonal antibodies to CD44s inhibited metas-
tasis of a human melanoma cell line put into nude mice without affecting the primary tumor, demonstrating selective inhibition of metastatic spread [Guo et al. 1994]. Despite the strength of these initial observations and the subsequent demonstration of CD44v expression in a variety of malignancies, a direct cause–effect relationship between the expression of specific CD44v isoforms and neoplastic transformation or acquisition of metastatic potential has been difficult to establish. Malignant transformation is often associated with increased expression of CD44 as well as alternative splicing transitions, thus it is not clear whether a qualitative or quantitative effect is most relevant to neoplasia and metastasis. In addition, many different CD44v isoforms are associated with malignancy, yet the specific functions of these isoforms are not known.

New tools are emerging to allow rapid and direct analysis of large numbers of alternative splicing events that will expedite identification of those transitions that consistently correlate with malignancy [Hu et al. 2001; Yeakley et al. 2002]. This information can be used to identify alternative splicing switches that have a cause–effect relationship with neoplasia and malignancy. These assays also provide diagnostic and prognostic tools to rapidly identify cancer subtypes and to correlate these subtypes with the most effective treatments following the paradigm recently established by microarray analysis of mRNA steady-state levels [Macgregor and Squire 2002]. Analysis of pre-mRNAs with similarly regulated splicing programs can be used to identify the regulatory networks that are affected and, in the future, the regulatory factors that play a role in neoplasia.

Three recent analyses have established correlations between the expression of splicing regulatory factors and alternative splicing transitions associated with malignant transformation. In one, the relative abundance of specific SR protein family members was shown to increase during progression from neoplasia to metastasis in a well-characterized model of mouse mammary gland tumorigenesis [Stickeler et al. 1999]. Increased SR protein expression also correlated with increased complexity of CD44 isoforms. Although changes in SR protein expression were a marker for neoplasia [the potential of nonneoplastic cells to progress to tumors], they were not predictive for tumor incidence or invasiveness, indicating that additional cellular changes are required. Therefore, an alteration in SR protein function could be one of the multiple changes required for neoplasia and malignancy.

Several genes that are determinative for apoptosis express antagonistic pro- and antiapoptotic isoforms determined by alternative splicing [Jiang and Wu 1999]. Recent studies suggest a pathway for induction of apoptosis by the second messenger, ceramide, that involves regulation of SR protein phosphorylation. Phosphorylation of SR proteins modulates their RNA-binding specificity, protein:protein interactions, intrinsic splicing activity, and nuclear:cytoplasmic localization [Tacke et al. 1997; Xiao and Manley 1997; Caceres et al. 1998]. Ceramide was shown to activate protein phosphatase 1 [PP1], causing dephosphorylation of SR proteins. SR protein dephosphorylation correlated with an alternative splicing switch to proapoptotic isoforms of Bcl-x(s) and caspase 9 [Challant et al. 2001, 2002], which suggests a role for SR protein posttranslational modification in an apoptosis signal transduction pathway. However, regulation of Bcl-x(s) and caspase-9 alternative splicing by the SR proteins remains to be demonstrated.

In a third example, an alternative splicing switch of fibroblast growth factor receptor 1 (FGFR1) from a lower to a higher affinity receptor is proposed to provide a growth advantage during malignant progression of glial cells to glioblastoma [Yamaguchi et al. 1994]. The splicing switch, which involves increased skipping of the α exon, is reproduced using an FGFR1 minigene in cultured glioblastoma cells. Two intronic elements adjacent to the regulated exon were shown to be required for the switch [Jin et al. 1999]. The splicing regulator PTB binds to these elements and promotes exon skipping when coexpressed with the FGFR1 minigene. Furthermore, expression of PTB in malignant glioblastomas is increased relative to glial cells, correlating with skipping of the FGFR1 α exon [Jin et al. 2000].

There is likely to be a determinative role for alternative splicing regulation in the progression of some neoplasia and malignancies. However, identification of the relevant correlations will require a systematic approach. Cancer is an extremely heterologous disease that differs according to tumor type and tissue of origin. Cancer cells put into culture lose and gain a variety of properties, particularly those most relevant to neoplasia and malignancy. It is therefore difficult to generalize results between different neoplasias and even between primary tumors and their derived cell lines. Correlations observed in one tumor type might not apply to others or may be lost or even artifactually created in cell culture. The analysis requires very well-defined tissue samples such as those in which metastatic cells are compared with their nonmetastatic parental line or in vivo models of tumor progression. Alternative splicing arrays used in combination with quantitative analysis can be used to screen for alternative splicing changes relevant to neoplasia and distinguish the contributions of alternative splicing from those associated with changes in expression levels.

Conclusions

Novel therapeutic strategies directed toward correcting or circumventing splicing abnormalities are now emerging. Approaches include overexpression of proteins that alter splicing of the affected exon [Hofmann et al. 2000; Nissim-Rafinia et al. 2000], use of oligonucleotides to block use of aberrant splice sites and force use of beneficial splice sites [Kalbfuss et al. 2001; Mercatante and Kole 2002], use of compounds that affect phosphorylation of splicing factors [Pilch et al. 2001] or stabilize putative secondary structures [Varani et al. 2000], high-throughput screens to identify compounds that influence splicing efficiencies of target pre-mRNAs [Andre-
assi et al. 2001), and a trans-splicing approach to replace mutated exons with wild-type exons (Liu et al. 2002). The limitations that need to be addressed include [1] the efficiency with which the treatment corrects the splicing defect; (2) the “side effects” on splicing of pre-mRNAs other than the target; and [3] effective and long-lasting delivery to the appropriate cells. Altered splicing patterns can also serve as markers of the altered cellular state associated with disease even when they are not in the primary pathway of the disease mechanism and have the potential to provide diagnostic and prognostic information. We can look forward to intriguing biology and increasing utility as we come to understand the diverse mechanisms by which disrupted splicing and splicing regulation contribute to human disease.

Acknowledgments

We thank Jo Ann Wise, Miles Wilkinson, and Gerard Schellenberg for helpful discussions and Utz Fischer, Juan Valcarcel, Adrian Krainer, Nicolas Charlet-B., Andrea Ladd, and other members of the Cooper laboratory for comments on the manuscript. Work in our laboratory is supported by the National Heart, Lung, and Blood Institute and the National Institute of Arthritis, Musculoskeletal, and Skin Diseases. N.A.F. is supported by the Portuguese Foundation for Science and Technology.

References

Andreason, C., Jarecki, J., Zhou, J., Coovert, D.D., Monani, U.R., Chen, X., Whitney, M., Pollok, B., Zhang, M., Androphy, E., et al. 2001. Aclacinomycin A treatment restores SMN levels to cells derived from type I spinal muscular atrophy patients. *Hum. Mol. Genet.* 10: 2841–2819.

Armstrong, J.F., Pritchard-Jones, K., Bickmore, W.A., Hastie, N.D., and Bard, J.B. 1993. The expression of the Wilms' tumour gene, WT1, in the developing mammary embryo. *Mech. Dev.* 40: 85–97.

Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Milodzik, M., Perez-Alonso, M., and Baylies, M.K. 1998. The muscleblind gene participates in the organization of Z-bands and epidermal attachments of *Drosophila* muscles and is regulated by Dmef2. *Dev. Biol.* 195: 131–143.

Ayadi, L., Callebaut, I., Saguez, C., Villa, T., Morrow, J.P., and Banroques, J. 1998. Functional and structural characterization of the prp3 binding domain of the yeast prp4 splicing factor. *J. Mol. Biol.* 284: 673–687.

Baccon, J., Pellizzoni, L., Rappaport, J., Mann, M., and Dreyfuss, G. 2002. Identification and characterization of Gemin7, a novel component of the splicing motor neuron complex. *J. Biol. Chem.* 277: 31957–31962.

Barbaux, S., Niaudet, P., Gubler, M.C., Grunfeld, J.P., Jaubert, F., Kuttken, F., Fekete, C.N., Souleyreau-Therville, N., Thibaud, E., Fellous, M., et al. 1997. Donor splice-site mutations in WT1 are responsible for Frasier syndrome. *Nat. Genet.* 17: 467–470.

Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez-Alonso, M., and Milodzik, M. 1997. Muscleblind, a gene required for photoreceptor differentiation in *Drosophila*, encodes novel nuclear CysHis-type zinc-finger-containing proteins. *Development* 124: 4321–4331.

Berget, S.M. 1995. Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270: 2411–2414.

Berget, S.M., Moore, C., and Sharp, P.A. 1977. Spliced segments at the 5’ terminus of adenosvirus 2 late mRNA. *Proc. Natl. Acad. Sci.* 74: 3171–3175.

Bessant, D.A., Ali, R.R., and Bhattacharya, S.S. 2001. Molecular genetics and prospects for therapy of the inherited retinal dystrophies. *Curr. Opin. Genet. Dev.* 11: 307–316.

Binder, G., Brown, M., and Parks, J.S. 1996. Mechanisms responsible for dominant expression of human growth hormone gene mutations. *J. Clin. Endocrinol. Metab.* 81: 4047–4050.

Bishop, D.T., McDonald, W.H., Gould, K.L., and Forshburg, S.L. 2000. Isolation of an essential Schizosaccharomyces pombe gene, pp31*, that links splicing and meiosis. *Nucleic Acids Res.* 28: 2214–2220.

Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* [in press].

Blencowe, B.J. 2000. Exonic splicing enhancers: Mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* 25: 106–110.

Buse, L., Bussiere, T., Buce-Scherrer, V., Delacourte, A., and Hof, P.R. 2000. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res. Brain Res. Rev.* 33: 95–130.

Buj-Bello, A., Furling, D., Tronchere, H., Laporte, J., Lerouge, T., Butler-Browne, G.S., and Mandel, J.L. 2002. Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in DM1 muscle cells. *Hum. Mol. Genet.* 11: 2297–2307.

Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M., and Baralle, F.E. 2001. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.* 20: 1774–1784.

Burge, C.B., Tuschl, T., and Sharp, P.A. 1999. Splicing of pre-mRNAs by the spliceosomes. In *The RNA world* [eds. R.F. Gesteland et al.], pp. 525–560. Cold Spring Harbor Press, Cold Spring Harbor, NY.

Burlet, P., Huber, C., Bertrand, S., Ludsky, M.A., Zwaenepoel, I., Clermont, O., Roume, J., Delezoide, A.L., Cartaud, J., Munnich, A., et al. 1998. The distribution of SMN protein complex in human fetal tissues and its alteration in spinal muscular atrophy. *Hum. Mol. Genet.* 7: 1927–1933.

Burns, C.G., Ohi, R., Mehta, S., O’Toole, E.T., Winey, M., Clark, T.A., Sugnet, C.W., Ares Jr., M., and Gould, K.L. 2002. Removal of a single α-tubulin gene intron suppresses cell cycle arrest phenotypes of splicing factor mutations in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 22: 801–815.

Caceres, J.F. and Kornblith, A.R. 2002. Alternative splicing: Multiple control mechanisms and involvement in human disease. *Trends Genet.* 18: 186–193.

Caceres, J.F., Screaton, G.R., and Krainer, A.R. 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & Dev.* 12: 55–66.

Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., Lewis, W.H., et al. 1990. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms’ tumor locus. *Cell* 60: 509–520.

Cartegni, L. and Krainer, A.R. 2002. Mutation of an essential SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.* 30: 377–384.

Cartegni, L., Chew, S.L., and Krainer, A.R. 2002. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* 3: 285–298.

Chakarova, C.F., Hims, M.M., Bolz, H., Abu-Safieh, L., Patel,
R.J., Papaioannou, M.G., Inglehearn, C.F., Keen, T.J., Willis, C., Moore, A.T., et al. 2002. Mutations in HPRT3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* 11: 87–92.

Chaffant, C.E., Ogretmen, B., Galadari, S., Kroezen, B.J., Pettus, B.J., and Hannun, Y.A. 2001. FAS activation induces dephosphorylation of SR proteins, dependence on the de novo generation of ceramide and activation of protein phosphatase 1. *J. Biol. Chem.* 276: 44848–44855.

Chaffant, C.E., Rathman, K., Pinkerman, R.L., Wood, R.E., Obeid, L.M., Ogretmen, B., and Hannun, Y.A. 2002. De novo ceramide regulates the alternative splicing of caspase 9 and Bcl-x in A549 lung adenocarcinoma cells. Dependence on protein phosphatase-1. *J. Biol. Chem.* 277: 12587–12595.

Charlet-B., N., Logan, P., Singh, G., and Cooper, T.A. 2002a. Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol. Cell* 9: 649–658.

Chaffant, B., N., Savkur, R.S., Singh, G., Philips, A.V., Grice, E.A., and Cooper, T.A. 2002b. Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell* 10: 45–53.

Chen, E.J., Frad, A.R., Chittouras, E., and Kaiser, C.A. 1998. A link between secretion and pre-mRNA processing defects in *Saccharomyces cerevisiae* and the identification of a novel splicing gene, RSE1. *Mol. Cell Biol.* 18: 7139–7146.

Chillon, M., Casals, T., Mercier, B., Bassas, L., Lissens, W., Silber, S., Romey, M.C., Ruiz-Romero, J., Verlingue, C., Claustres, M., et al. 1995. Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. *N. Engl. J. Med.* 332: 1475–1480.

Chow, L.T., Gelinis, R.E., Broker, T.R., and Roberts, R.J. 1977. An amazing sequence arrangement at the 5’ ends of adenovirus 2 messenger RNA. *Cell* 12: 1–8.

Chu, C.S., Trapnell, B.C., Currin, S.M., Cutting, G.R., and Crystal, R.G. 1992. Extensive posttranscriptional deletion of the coding sequences for part of nucleotide-binding fold 1 in respiratory epithelial mRNA transcripts of the cystic fibrosis transmembrane conductance regulator gene is not associated with the clinical manifestations of cystic fibrosis. *J. Clin. Invest.* 90: 785–790.

Cifuentes-Diaz, C., Frugier, T., Tiziano, F.D., Lacene, E., Robet, N., Joshi, V., Moreau, M.H., and Melki, J. 2001. Deletion of murine SMN exon 7 directed to skeletal muscle leads to severe muscular dystrophy. *J. Cell Biol.* 152: 1107–1114.

Clark, T.A., Sagnit, C.W., and Ares Jr., M. 2002. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296: 907–910.

Cogan, J.D., Phillips III, J.A., Schenken, S.S., Milner, R.D., and Sakati, N. 1994. Familial growth hormone deficiency: A metabolic basis of inherited disease. In *nature and mechanisms of human gene mutation.* In *The metabolic basis of inherited disease.* Eds. C.R. Scriver et al., pp. 259–291. McGraw-Hill, New York.

Cogan, J.D., Prince, M.A., Lekhakula, S., Bundey, S., Futrakul, A., McCarthy, E.M., and Phillips III, J.A. 1997. A novel mechanism of aberrant pre-mRNA splicing in humans. *Hum. Mol. Genet.* 6: 909–912.

Collins, A.C. and Guthrie, C. 2000. The question remains: Is the spliceosome a ribozyme? *Nat. Struct. Biol.* 7: 850–854.

Cooper, D.N., Krawczak, M., and Antonarakis, S.E. 1995. The nature and mechanisms of human gene mutation. In *The metabolic basis of inherited disease.* Eds. C.R. Scriver et al., pp. 259–291. McGraw-Hill, New York.

Cooper, T.A. and Mattox, W. 1997. The regulation of splice site selection and its role in human disease. *Am. J. Hum. Genet.* 61: 259–266.

Coovert, D.D., Le, T.T., McAndrew, P.E., Strasswimmer, J., Crawford, T.O., Mendell, J.R., Coulson, S.E., Androphy, E.J., Prior, T.W., and Burghes, A.H. 1997. The survival motor neuron protein in spinal muscular atrophy. *Hum. Mol. Genet.* 6: 1205–1214.

Coulter, L.R., Landree, M.A., and Cooper, T.A. 1997. Identification of a new class of exonic splicing enhancers by in vivo selection. *Mol. Cell. Biol.* 17: 2143–2150.

Cuppens, H., Lin, W., Jaspers, M., Costes, B., Teng, H., Vankeerberghen, A., Jorissen, M., Droogmans, G., Reynaert, I., Goossens, M., et al. 1998. Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes. The polymorphic (Tg)_S locus explains the partial penetrance of the T5 polymorphism as a disease mutation. *J. Clin. Invest.* 101: 487–496.

Davies, R.C., Calvio, C., Brett, E., Larsson, S.H., Lamond, A.I., and Hastie, N.D. 1998. WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. *Genes & Dev.* 12: 3217–3225.

Davis, B.M., McCurrach, M.E., Taneja, K.L., Singer, R.H., and Housman, D.E. 1997. Expansion of a CUG trinucleotide repeat in the 3’ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl. Acad. Sci.* 94: 7388–7393.

Delaney, S.J., Rich, D.P., Thomson, S.A., Hargrave, M.R., Lovelock, P.K., Welsh, M.J., and Wainwright, B.J. 1993. Cystic fibrosis transmembrane conductance regulator splice variants are not conserved and fail to produce chloride channels. *Nat. Genet.* 4: 426–431.

D’Souza, I., Poorkaj, P., Hong, M., Novchin, D., Lee, V.M., Bird, T.D., and Schellenberg, G.G. 1999. Missense and silent tau gene mutations cause frontotemporal dementia with Parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. *Proc. Natl. Acad. Sci.* 96: 5598–5603.

Du, H. and Rosbash, M. 2002. The U1 snRNP protein U1C recognizes the 5’ splice site in the absence of base pairing. *Nature* 419: 86–90.

Engelbrecht, J., Knudsen, S., and Brunak, S. 1992. G + C-rich tract in 5’ end of human introns. *J. Mol. Biol.* 227: 108–113.

Fackenthal, J.D., Cartegni, L., Krainer, A.R., and Olopade, O.I. 2002. BRCA2 T2722R is a deleterious allele that causes exon skipping. *Am. J. Hum. Genet.* 71: 625–631.

Fairbrother, W.G. and Chasin, L.A. 2000. Human genomic sequences that inhibit splicing. *Mol. Cell. Biol.* 20: 6816–6825.

Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* 297: 1007–1013.

Fardaei, M., Larkin, K., Brook, J.D., and Hamshere, M.G. 2001. In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res.* 29: 2766–2771.

Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S., and Brook, J.D. 2002. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* 11: 805–814.

Feldkotter, M., Schwarz, V., Wirth, R., Wienker, T.F., and Wirth, B. 2002. Quantitative analyses of SMN1 and SMN2 based on real-time light cycle PCR: Fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. *Am. J. Hum. Genet.* 70: 358–368.

Gessler, M., Poustka, A., Cavenee, W., Neve, R.L., Orkin, S.H., and Bruns, G.A. 1990. Homozygous deletion in Wilms tumor. *Science* 249: 774–778.
Faustino and Cooper

E., Chen, W., Burgess, S., Haldi, M., Artzt, K., Farrington, S., et al. 2002. Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. Nat. Genet. 31: 135–140.

Gonzalez-Santos, J.M., Wang, A., Jones, J., Ushida, C., Liu, J., and Hu, J. 2002. Central region of the human splicing factor Hprp3p interacts with Hprp4p. J. Biol. Chem. 277: 23764–23772.

Good, P.J., Chen, Q., Warner, S.J., and Herring, D.C. 2000. A family of human RNA-binding proteins related to the Drosofila Bruno translational regulator. J. Biol. Chem. 275: 28583–28592.

Grabowski, P.J. 1998. Splicing regulation in neurons: Tinkering with cell-specific control. Cell 92: 709–712.

Gunthert, U., Hofmann, M., Rudy, W., Reber, S., Zoller, M., Haufusmann, I., Matzku, S., Wenzel, A., Ponta, H., and Herrlich, P. 1991. A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. Cell 65: 13–25.

Guo, Y., Ma, J., Wang, J., Che, X., Narula, J., Bigby, M., Wu, M., and Sy, M.S. 1994. Inhibition of human melanoma growth and metastasis in vivo by anti-CD44 monoclonal antibody. Cancer Res. 54: 1561–1565.

Haber, D.A., Sohn, R.L., Buckler, A.J., Pelletier, J., Call, K.M., and Houseman, D.E. 1991. Alternative splicing and genomic structure of the Wilms tumor gene WT1. Proc. Natl. Acad. Sci. 88: 9618–9622.

Hammes, A., Guo, J.K., Lutsch, G., Lecheste, J.R., Landrock, D., Ziegler, U., Gubler, M.C., and Schedl, A. 2001. Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. Cell 106: 319–329.

Harper, P.S. 2001. Myotonic dystrophy. W.B. Saunders, London.

Hastings, M.L. and Krainer, A.R. 2001. Pre-mRNA splicing in the new millennium. Curr. Opin. Cell Biol. 13: 302–309.

Hentze, M.W. and Kulozik, A.E. 1999. A perfect message: RNA splicing and polypyrimidine tract-binding protein in glialbloma multiforme tumors. Cancer Res. 60: 1221–1224.

Kalbfuss, B., Mabon, S.A., and Misteli, T. 2001. Correction of alternative splicing of tau in frontotemporal dementia and Parkisonism linked to chromosome 17. J. Biol. Chem. 276: 42896–42923.

Kan, J.L. and Green, M.R. 1999. Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. Genes & Dev. 13: 462–471.

Kiesewetter, S., Macek Jr., M., Davis, C., Currstin, S.M., Chu, C.S., Graham, C., Shironim, A.E., Cashman, S.M., Tsui, L.C., Mickle, J., et al. 1993. A mutation in CFTR produces different phenotypes depending on chromosomal background. Nat. Genet. 5: 274–278.

Kohsaka, T., Tagawa, M., Takekoshi, Y., Yanagisawa, H., Tadokoro, K., and Yamada, M. 1999. Exon 9 mutations in the WT1 gene, without influencing KTS splice isoforms, are also responsible for Frasier syndrome. Hum. Mutat. 14: 466–470.

Korenblot, J.I. and Fernald, R.D. 1989. Circadian rhythm and light regulate opsin mRNA in rod photoreceptors. Nature 337: 454–457.

Krawczak, M., Reiss, J., and Cooper, D.N. 1992. The mutational spectrum of single base-pair substitutions in messenger RNA splice junctions of human genes—Causes and consequences. Hum. Genet. 90: 41–54.

Kreidberg, J.A., Sariola, H., Loring, J.M., Maca, M., Pelletier, J., Housman, D., and Jaenisch, R. 1993. WT-1 is required for early kidney development. Cell 74: 679–691.

Ladd, A.N. and Cooper, T.A. 2002. Finding signals that regulate alternative splicing in the post-genomic era. Genome Biol. 3: 8.1–8.16.

Ladd, A.N., Charlet-B., N., and Cooper, T.A. 2001. The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol. Cell. Biol. 21: 1285–1296.

Lalena, M.J., Chalmers, K.J., Llamazares, S., Lamond, A.I., and Varacelle, J. 2002. Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SpF45. Cell 109: 285–296.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Larriba, S., Bassas, L., Gimenez, J., Ramos, M.D., Segura, A., Nunes, V., Estivill, X., and Casals, T. 1998. Testicular CFTR splice variants in patients with congenital absence of the vas deferens. J. Biol. Chem. 273: 1237–1245.
Pre-mRNA splicing and human disease

Maquat, L.E. and Carmichael, G.G. 2001. Quality control of mRNA function. Cell 104: 173–176.

McCarthy, E.M. and Phillips III, J.A. 1998. Characterization of an intron splice enhancer that regulates alternative splicing of human GH pre-mRNA. Hum. Mol. Genet. 7: 1491–1496.

McCullough, A.J. and Berget, S.M. 2000. An intronic splicing enhancer binds U1 snRNPs to enhance splicing and select 5′ splice sites. Mol. Cell. Biol. 20: 9225–9235.

Mckie, A.B., McHale, J.C., Keen, T.J., Tarttell, E.E., Goliath, R., van Lith-Verhoeven, J.J., Greenberg, J., Ramesar, R.S., Hoyng, C.B., Cremers, F.P., et al. 2001. Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa [RP13]. Hum. Mol. Genet. 10: 1555–1562.

Meister, G., Buhler, D., Pillai, R., Lortschpeich, F., and Fischer, U. 2001. A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nat. Cell Biol. 3: 945–949.

Melo, K.F., Martin, R.M., Costa, E.M., Carvalho, F.M., Jorge, A.A., Arnhold, I.J., and Mendonca, B.B. 2002. A new phenotype of Frasier syndrome due to I599 +4C → T mutation in the WT1 gene: Predominantly male ambiguous genitalia and absence of gonadal dysgenesis. J. Clin. Endocrinol. Metab. 87: 2500–2505.

Mercatante, D.R. and Kole, R. 2002. Control of alternative splicing by antisense oligonucleotides as a potential chemotherpay: Effects on gene expression. Biochim. Biophys. Acta 1587: 126–132.

Michalowski, S., Miller, J.W., Urbiniati, C.R., Paliouras, M., Swanson, M.S., and Griffith, J. 1999. Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. Nucleic Acids Res. 27: 3534–3542.

Miles, C., Elgar, G., Coles, E., Kleinjan, D.J., van Heyningen, V., and Hastie, N. 1998. Complete sequencing of the Fugu WAGR region from WT1 to PAX6: Dramatic compaction and conservation of synteny with human chromosome 11p13. Proc. Natl. Acad. Sci. 95: 13068–13072.

Miller, J.W., Urbiniati, C.R., Teng-Umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A., and Swanson, M.S. 2000. Recruitment of human muscleblind proteins to (CUG)_n expansions with myotonic dystrophy. EMBO J. 19: 4439–4448.

Modrek, B. and Lee, C. 2002. A genomic view of alternative splicing. Nat. Genet. 30: 13–19.

Monani, U.R., Sendtner, M., Coover, D.D., Parsons, D.W., Andressasi, C., Le, T.T., Jablonka, S., Schrank, B., Rossol, W., Prior, T.W., et al. 2000. The human centromeric survival motor neuron gene (SMN2) rescues embryonic lethality in Smn−/− mice and results in a mouse with spinal muscular atrophy. Hum. Mol. Genet. 9: 333–339.

Moseley, C.T., Mullis, P.E., Prince, M.A., and Phillips III, J.A. 2002. An exon splice enhancer mutation causes autosomal dominant GH deficiency. J. Clin. Endocrinol. Metab. 87: 847–852.

Nakai, K. and Sakamoto, H. 1994. Construction of a novel database containing aberrant splicing mutations of mammalian genes. Gene 141: 171–177.

Napieraa, M. and Krzyysiak, W.J. 1997. CUG repeats present in myotonia kinase RNA form metastable \‘slippery\’ hairpins. J. Biol. Chem. 272: 31079–31085.

Niksic, M., Romano, M., Buratti, E., Pagani, F., and Baralle, F.E. 1999. Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. Hum. Mol. Genet. 8: 2339–2349.

Nissim-Rafinia, M. and Kerem, B. 2002. Splicing regulation as a
potential genetic modifier. Trends Genet. 18: 123–127.
Nissim-Rafinia, M., Chiba-Falek, O., Sharon, G., Boss, A., and Kerem, B. 2000. Cellular and viral splicing factors can modify the splicing pattern of CFTR transcripts carrying splicing mutations. Hum. Mol. Genet. 9: 1771–1778.
Noone, P.G. and Knowles, M.R. 2001. ‘CFTR-opathies’: Disease phenotypes associated with cystic fibrosis transmembrane regulator gene mutations. Respir. Rev. 2: 328–332.
Noone, P.G., Pue, C.A., Zhou, Z., Friedman, K.J., Wakeling, E.L., Ganeshanathan, M., Simon, R.H., Silverman, L.M., and Knowles, M.R. 2000. Lung disease associated with the IVS8 5′ allele of the CFTR gene. Am. J. Respir. Crit. Care Med. 162: 1919–1924.
Nussinov, R. 1988. Conserved quartets near 5′ intron junctions in primate nuclear pre-mRNA. J. Theor. Biol. 133: 73–84.
Ohi, R., McCollum, D., Hirani, B., Den Haese, G.J., Zhang, X., Burke, J.D., Turner, K., and Gould, K.L. 1994. The Schizosaccharomyces pombe cds5 gene encodes an essential protein with homology to c-Myb. EMBO J. 13: 471–483.
Owen, N., Doe, C.L., Mellor, J., and Davies, K.E. 2000. Characterization of the Schizosaccharomyces pombe orthologue of the human survival motor neuron (SMN) protein. Hum. Mol. Genet. 9: 675–684.
Pagani, F., Buratti, E., Stuani, C., Romano, M., Zuccato, E., Niksic, M., Giglio, L., Faraguna, D., and Baralle, F.E. 2000. Splicing factors induce cytic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. J. Biol. Chem. 275: 21041–21047.
Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dörk, T., and Baralle, F.E. 2002. A new type of mutation causes a splicing defect in ATM. Nat. Genet. 30: 426–429.
Pellizzoni, L., Fuxy, J., and Dreyfuss, G. 1998. A novel function for SMN: the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell 95: 615–624.
Pellizzoni, L., Yong, J., and Dreyfuss, G. 2002. Essential role for the SMN complex in the specificity of snRNP assembly. Science 298: 1775–1779.
Philips, A.V. and Cooper, T.A. 2000. RNA processing and human disease. Cell Mol. Life Sci. 57: 235–249.
Philips, A.V., Timchenko, L.T., and Cooper, T.A. 1998. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science 280: 737–741.
Pilch, B., Allemand, E., Facombre, M., Bailly, C., Riou, J.F., Soret, J., and Tazi, J. 2001. Specific inhibition of serine- and arginine-rich splicing factors phosphorylation, spliceosome assembly, and splicing by the antitumor drug NB-506. Cancer Res. 61: 6876–6884.
Proudfoot, N.J., Furger, A., and Dye, M.J. 2002. Integrating mRNA processing with transcription. Cell 108: 501–512.
Raker, V.A., Plessel, G., and Luhmann, R. 1996. The snRNP core complex pathway: Identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. EMBO J. 15: 2256–2269.
Reed, R. 1996. Initial splice-site recognition and pairing during pre-mRNA splicing. Curr. Opin. Gen. Dev. 6: 215–220.
Roberts, R., Timchenko, N.A., Miller, J.W., Reddy, S., Caskey, C.T., Swanson, M.S., and Timchenko, L.T. 1997. Altered phosphorylation and intracellular distribution of a [CUG], triple repeat RNA-binding protein in patients with myotonic dystrophy and in myotonin protein kinase knockout mice. Proc. Natl. Acad. Sci. 94: 13221–13226.
Rogan, P.K., Faux, B.M., and Schneider, T.D. 1998. Information analysis of human splice site mutations. Hum. Mutat. 12: 153–171.
Savkur, R.S., Philips, A.V., and Cooper, T.A. 2001. aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myot dystrophy. Nat. Genet. 29: 40–47.
Schrank, B., Gotz, R., Gungnerson, J.M., Ure, J.M., Toyka, K.V., Smith, A.G., and Sendtner, M. 1997. Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. Proc. Natl. Acad. Sci. 94: 9920–9925.
Seznec, H., Agbulut, O., Sergeant, N., Savouret, C., Ghedest, A., Tabti, N., Willer, J.C., Ourth, L., Duros, C., Brisson, E., et al. 2001. Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. Hum. Mol. Genet. 10: 2717–2226.
Sirard-Pugnet, P., Durosay, P., Brody, E., and Marie, J. 1995. An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken β-tropomyosin pre-mRNA. Nucleic Acids Res. 23: 3501–3507.
Slaugenhaupt, S.A., Blumenfeld, A., Gill, S.P., Leyne, M., Mull, J., Cua-jungco, M.P., Liebert, C.B., Chadwick, B., Idelson, M., Reznik, L., et al. 2001. Tissue-specific expression of a splicing mutation in the IXKBKAP gene causes familial dysautonomia. Am. J. Hum. Genet. 68: 598–605.
Smith, C.W. and Valcarcel, J. 2000. Alternative pre-mRNA splicing: The logic of combinatorial control. Trends Biochem. Sci. 25: 381–388.
Sneath, R.J. and Mangham, D.C. 1998. The normal structure and function of CD44 and its role in neoplasia. Mol. Pathol. 51: 191–200.
Spillantini, M.G., Murrell, J.R., Goedert, M., Farlow, M.R., Klug, A., and Ghe-tti, B. 1998. Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. Proc. Natl. Acad. Sci. 95: 7773–7774.
Stickler, E., Kittrell, F., Medina, D., and Berget, S.M. 1999. Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. Oncogene 18: 3574–3582.
Stickler, E., Fraser, S.D., Honig, A., Chen, A.L., Berget, S.M., and Cooper, T.A. 2001. The RNA binding protein Y1B-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. EMBO J. 20: 3821–3830.
Strong, T.V., Wilkinson, D.J., Mansoura, M.K., Devor, D.C., Henze, K., Yang, Y.P., Wilson, J.M., Cohn, J.A., Dawson, D.C., Frizzell, R.A., et al. 1993. Expression of an abundant alternatively spliced form of the cytic fibrosis transmembrane conductance regulator (CFTR) gene is not associated with a CAMP-activated chloride conductance. Hum. Mol. Genet. 2: 225–230.
Sun, H. and Chasin, L.A. 2000. Multiple splicing defects in an intronic false exon. Mol. Cell. Biol. 20: 6414–6425.
Tacke, R., Chen, Y., and Manley, J.L. 1997. Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: Creation of an SRp40-specific splicing enhancer. Proc. Natl. Acad. Sci. 94: 1148–1153.
Taneja, K.L., McCurrach, M., Schalling, M., Housman, D., and Singer, R.H. 1995. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. J. Cell Biol. 128: 995–1002.
Tarn, W.Y. and Steitz, J.A. 1997. Pre-mRNA splicing: The discovery of a new spliceosome doubles the challenge. Trends Biochem. Sci. 22: 132–137.
Teigelkamp, S., Newman, A.J., and Beggs, J.D. 1995. Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing suggest a role in stabilization of exon alignment by US snRNAs. EMBO J. 14: 2602–2612.
Terns, M.P. and Terns, R.M. 2001. Macromolecular complexes: SMN—the master assembler. Curr. Biol. 11: R862–R864.
Pre-mRNA splicing and human disease

M.B., and Thornton, C.A. 2000. Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. RNA 6: 79–87.

Timechenko, L.T., Miller, J.W., Timechenko, N.A., Devore, D.R., Datar, K.V., Lin, L.J., Roberts, R., Caskey, C.T., and Swanson, M.S. 1996. Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. Nucleic Acids Res. 24: 4407–4414.

Timechenko, N.A., Cai, Z.J., Welm, A.L., Reddy, S., Ashizawa, T., and Timchenko, L.T. 2001. RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. J. Biol. Chem. 276: 7820–7826.

Varani, L., Spillantini, M.G., Goedert, M., and Varani, G. 2000. Structural basis for recognition of the RNA major groove in the tau exon 10 splicing regulatory element by aminoglycoside antibiotics. Nucleic Acids Res. 28: 710–719.

Vervoort, R., Lennon, A., Bird, A.C., Tulloch, B., Axton, R., Timchenko, L.T., Miller, J.W., Timchenko, N.A., Devore, D.R., and Varani, G. 2000. Identification of a (CUG)n triplet repeat RNA interactions and is necessary for splicing. Genes & Dev. 11: 334–344.

Wyatt, J.R., Sontheimer, E.J., and Steitz, J.A. 1992. Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. Genes & Dev. 6: 2542–2553.

Xiao, S.H. and Manley, J.L. 1997. Phosphorylation of the ASF/SF2 RS domain affects both protein–protein and protein–RNA interactions and is necessary for splicing. Genes & Dev. 11: 182–185.

Zhu, J., Mayeda, A., and Krainer, A.R. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol. Cell 8: 1351–1361.

Vidal, V.P., Verdone, L., Mayes, A.E., and Beggs, J.D. 1999. Characterization of U6 snRNA–protein interactions. RNA 5: 1470–1481.

Villa, T., Pleiss, J.A., and Guthrie, C. 2002. Spliceosomal snRNAs: Mg2+-dependent chemistry at the catalytic core? Cell 109: 149–152.

Wang, A., Forman-Kay, J., Luo, Y., Luo, M., Chow, Y.H., Plumb, J., Friesen, J.D., Tsui, L.C., Heng, H.H., Woolford Jr., J.L., et al. 1997. Identification and characterization of human genes encoding Hprp3p and Hprp4p, interacting components of the spliceosome. Hum. Mol. Genet. 6: 375–381.

Wang, H.Y., Xu, X., Ding, J.H., Bermingham Jr., J.R., and Fu, X.D. 2001. SC35 plays a role in T cell development and alternative splicing of CD45. Mol. Cell 7: 331–342.

Wang, J. and Dreyfuss, G. 2001. A cell system with targeted disruption of the SMN gene: Functional conservation of the SMN protein and dependence of Gemin2 on SMN. J. Biol. Chem. 276: 9599–9605.

Weidenhammer, E.M., Singh, M., Ruiz-Noriega, M., and Woolford Jr., J.L. 1996. The PRP31 gene encodes a novel protein required for pre-mRNA splicing in Saccharomyces cerevisiae. Nucleic Acids Res. 24: 1164–1170.

Wilhelm, D. and Englert, C. 2002. The Wilms tumor suppressor WT1 regulates early gonad development by activation of SF1. Genes & Dev. 16: 1839–1851.

Will, C.L. and Luhrmann, R. 2001. Spliceosomal UsnRNP biogenesis, structure and function. Curr. Opin. Cell Biol. 13: 290–301.

Wirth, B. 2000. An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). Hum. Mutat. 15: 228–237.

Wu, J.L., Reed, R.B., Grabowski, P.J., and Artzt, K. 2002. Function of quaking in myelination: Regulation of alternative splicing. Proc. Natl. Acad. Sci. 99: 4233–4238.
Pre-mRNA splicing and human disease
Nuno André Faustino and Thomas A. Cooper

Genes Dev. 2003. 17:
Access the most recent version at doi:10.1101/gad.1048803

References
This article cites 178 articles, 66 of which can be accessed free at:
http://genesdev.cshlp.org/content/17/4/419.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.