Splicing regulation: From a parts list of regulatory elements to an integrated splicing code

ZEFENG WANG[1] and CHRISTOPHER B. BURGE[2]

[1]Department of Pharmacology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

ABSTRACT

Alternative splicing of pre-mRNAs is a major contributor to both proteomic diversity and control of gene expression levels. Splicing is tightly regulated in different tissues and developmental stages, and its disruption can lead to a wide range of human diseases. An important long-term goal in the splicing field is to determine a set of rules or "code" for splicing that will enable prediction of the splicing pattern of any primary transcript from its sequence. Outside of the core splice site motifs, the bulk of the information required for splicing is thought to be contained in exonic and intronic *cis*-regulatory elements that function by recruitment of sequence-specific RNA-binding protein factors that either activate or repress the use of adjacent splice sites. Here, we summarize the current state of knowledge of splicing *cis*-regulatory elements and their context-dependent effects on splicing, emphasizing recent global/genome-wide studies and open questions.

Keywords: context dependence; pre-mRNA splicing; splicing code; splicing factor; splicing regulation

PRE-MRNA SPLICING

Because human genes typically contain multiple introns, the process of pre-mRNA splicing is an essential step in the expression of most genes. A majority of human genes undergo alternative splicing (AS), generating multiple splicing isoforms containing different combinations of exons (Johnson et al. 2003). The effects of AS on protein products can be dramatic, e.g., producing soluble versus membrane-bound forms of the Fas receptor that have opposing effects on apoptosis (Cascino et al. 1995), or producing isoforms of the *Drosophila* fruitless protein that act to specify sexual orientation (Demir and Dickson 2005). The major forms of AS are summarized in Figure 1A. Splicing regulation has been comprehensively described in several recent reviews (Black 2003; Konarska and Query 2005; Matlin et al. 2005; Blencowe 2006; House and Lynch 2008). Here, we briefly summarize some aspects of splicing specificity and regulation before turning to our main topic of splicing regulatory elements and the rules governing their activity.

The sequential phosphodiester transfer reactions involved in splicing are catalyzed by large ribonucleoprotein complexes known as spliceosomes. Containing more than 100 core proteins and five small nuclear RNAs (snRNAs U1, U2, U4, U5, and U6), spliceosomes may be the most complex machines in the cell (Zhou et al. 2002; Jurica and Moore 2003; Nilsen 2003). In addition to these core factors, additional regulatory proteins participate in the splicing of particular pre-mRNAs. Splicing of most introns is thought to occur cotranscriptionally with fairly extensive interactions between splicing factors and the core transcription machinery (Das et al. 2006; Hicks et al. 2006). Splicing mechanisms have also been well reviewed recently (Hertel and Graveley 2005; Konarska and Query 2005).

CORE SPLICING SIGNALS

Three sites, the 5' splice site (5'ss), the 3' splice site (3'ss), and the branch point sequence (BPS), participate in the splicing reaction and are present in every intron, and thus are known as the core splicing signals. These signals are recognized multiple times during spliceosome assembly, with the 5'ss interacting initially with U1 and later with U6 snRNP, and the BPS interacting with SF1/mBBP and later with U2 snRNP (Fig. 1B; Will and Lührmann 2007). Because the 5'ss and 3'ss can be mapped precisely by aligning cDNA or expressed sequence tag (EST) sequences to the genome, large datasets of accurately annotated exon/
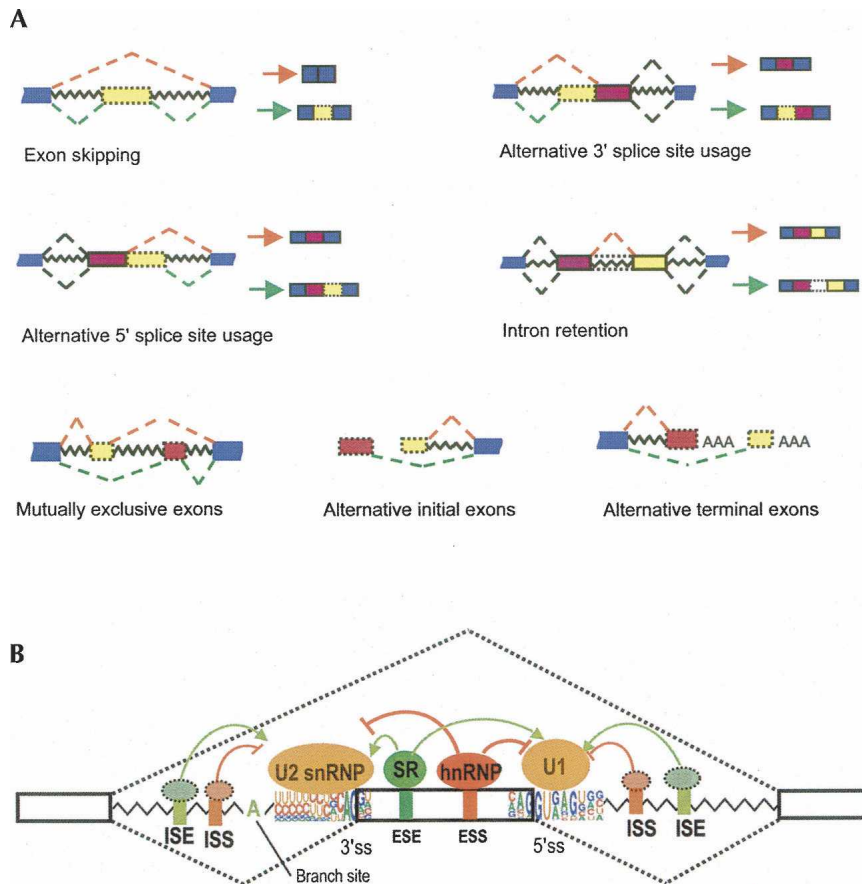
FIGURE 1. (*A*) Major forms of alternative splicing. In many cases, these common forms can be combined to generate more complicated alternative splicing events. (*B*) A schematic of regulated splicing. (Open boxes) Exons, (jagged lines) introns, (brackets) splice sites (ss). The consensus motifs of ss are shown in pictogram, and the branch point adenosine is indicated. (Dashed lines) Two alternative splicing pathways, with the middle exon either included or excluded. Splicing is regulated by *cis*-elements (ESE, ESS, ISS, and ISE) and *trans*-acting splicing factors (SR proteins, hnRNP, and unknown factors).

base pairs or more) that on average account for >90% of the primary transcript. This transcript geometry, and the predominant exon-skipping phenotype of splice site mutations, are consistent with the idea that in mammals splice sites are predominantly recognized in pairs across the exon through "exon definition" (Robberson et al. 1990; Nakai and Sakamoto 1994; Sterner et al. 1996). Exon definition involves initial interaction across the exon between factors recognizing the 5′ss and the upstream 3′ss, whereas in the alternative model, intron definition, interactions occur first across the intron between factors recognizing the 5′ss and the downstream 3′ss (for review, see Berget 1995). Recent analyses of the coevolution of the 5′ss and 3′ss have detected predominant cross-exon interactions in human and mouse, but cross-intron interactions in invertebrates, plants, and fungi, with pufferfishes representing an intermediate state, supporting the primacy of exon definition in mammals and intron definition in most other metazoans (Xiao et al. 2007). Because exon definition occurs early during splicing and may lead to commitment of the exon to splicing, this step is critical for splicing regulation and specificity. For example, polypyrimidine tract binding protein (PTB/hnRNP I) can inhibit exon definition complex formation by binding to an ESS sequence causing skipping of Fas exon 6 (Izquierdo et al. 2005); this same factor can also inhibit the spliceosome assembly across introns (intron definition) in repressing splicing of the c-src N1 exon (Sharma et al. 2005). Following initial splice site recognition in exon definition, a series of sequential structural rearrangements is required to activate the spliceosome, and commitment to alternative splice site pairing may occur after initial splice site recognition and E complex formation (Lim and Hertel 2004).

intron structures can be obtained, permitting development of statistical models of 5′ss and 3′ss motifs that capture second-order statistical interactions between positions to more accurately predict splice site locations (Yeo and Burge 2004). However, only a few dozen mammalian BPSs have been mapped. The limited size of the available BPS data set and the low information content of this motif make it difficult to derive a reliable sequence model to predict BPSs in introns. A recent study used comparative genomics to improve BPS prediction (Kol et al. 2005), and more human BPS sequences were identified recently by sequencing lariat RT-PCR products (GAO et al. 2008).

EXON DEFINITION AS A KEY STEP IN MAMMALIAN SPLICING REGULATION

A typical human gene contains relatively short exons (typically, 50–250 base pairs [bp] in length) separated by much larger introns (typically, hundreds to thousands of

SPLICING REGULATORY ELEMENTS AND ASSOCIATED FACTORS

Monte Carlo simulations inserting artificial motifs with varying information content into transcripts in place of natural splice sites have been used to estimate that the core human splice site motifs contain only about half of the information required to accurately define exon/intron boundaries, even considering only short introns (Lim and

Burge 2001). Large human introns typically contain numerous "decoy" splice sites: sequences with similar score/degree of consensus matching as authentic splice sites. Not infrequently, decoy splice sites occur in pairs as "pseudoexons," which resemble authentic exons in terms of length and splice site strength, but are very rarely if ever spliced (Sun and Chasin 2000). Despite the large potential for errors, the splicing process appears to occur with very high fidelity, implying the widespread involvement of transcript features besides the core splice signals in splice site selection. This additional information is thought to derive in large part from the presence of numerous *cis*-regulatory elements that serve as either splicing enhancers or silencers. These elements are conventionally classified as exonic splicing enhancers (ESEs) or silencers (ESSs) if from an exonic location they function to promote or inhibit inclusion of the exon they reside in, and as intronic splicing enhancers (ISEs) or silencers (ISSs) if they enhance or inhibit usage of adjacent splice sites or exons from an intronic location. In general, these splicing regulatory elements (SREs) function by recruiting *trans*-acting splicing factors that activate or suppress splice site recognition or spliceosome assembly by various mechanisms (Matlin et al. 2005; Chasin 2007).

SREs have most often been identified by directed mutagenesis of alternatively spliced genes and by analysis of naturally occurring mutations that perturb splicing of disease genes. ESEs are very abundant in constitutive exons, but there is ample evidence that silencers play at least as important a role in splicing regulation as enhancers (Fairbrother et al. 2002; Wang et al. 2004; Zhang and Chasin 2004). Based on the selective constraints on constitutive exons for efficient and accurate splicing and on AS exons for regulated splicing, enhancing elements are expected to play predominant roles in constitutive splicing, with silencers being relatively more prominent in control of AS.

EXONIC SPLICING ENHANCERS AND SILENCERS

It is now well established that ESEs include a diverse range of sequences, and many if not all exons contain internal ESE sequences (Schaal and Maniatis 1999; Fairbrother et al. 2002; Cartegni et al. 2003). Most ESEs function by recruiting members of the SR protein family (for review, see Graveley 2000). These factors usually regulate splicing by binding ESEs through their N-terminal RRM domains and mediating protein–protein interactions that facilitate spliceosome assembly through C-terminal RS domains (Graveley and Maniatis 1998). However, at least for in vitro splicing, the RS domains may not always be required for ESE-dependent splicing activation (Shaw et al. 2007), and RS domains can interact with RNA at the BPS and 5′ss during splicing complex assembly, at least in some cases (Shen et al. 2004; Shen and Green 2006).

ESSs are often bound by splicing repressors of the hnRNP class, a diverse group of proteins containing one or more RNA-binding domains and sometimes splicing inhibitory domains such as glycine-rich motifs (Pozzoli and Sironi 2005). hnRNPs function by a wide variety of mechanisms. For example, PTB (hnRNP I) can block essential interactions between U1 and U2 snRNPs (Izquierdo et al. 2005; Sharma et al. 2005), whereas hnRNP A1 can inhibit splicing by binding on either side and "looping out" exons or by directly displacing snRNP binding (Zhu et al. 2001; Nasim et al. 2002). Other splicing inhibitory mechanisms may also be used by these or other repressors.

Beyond molecular genetics, global approaches, including both computational and experimental methods, have been developed to identify ESEs and ESSs on a large scale (for review, see Chasin 2007). ESEs have been identified experimentally by in vitro and in vivo SELEX approaches (Tian and Kole 1995; Coulter et al. 1997; Liu et al. 1998, 2000). ESEs have also been computationally identified based on their enrichment in authentic exons versus introns and in exons with weak splice sites (Fairbrother et al. 2002), and by their enrichment in authentic exons versus pseudoexons and 5′ UTRs of intronless genes (Zhang and Chasin 2004). These studies have generated comprehensive lists of ESE oligonucleotides that help to predict the splicing phenotypes of exonic mutations (Cartegni et al. 2002; Fairbrother et al. 2002; Pfarr et al. 2005). ESSs have also been predicted based on the assumption that ESSs are depleted from authentic exons (Zhang and Chasin 2004). We developed a cell-based fluorescence-activated screen (FAS) and used it to identify 133 ESS decanucleotides active in human cells from a library of random sequences (Wang et al. 2004). Further analyses of these FAS–ESS sequences suggested that ESSs play an important role in distinguishing authentic exons from pseudoexons and in regulating alternative splice site usage and intron retention (Wang et al. 2004, 2006).

Additional exonic SREs have been predicted based on sequence conservation (Goren et al. 2006), but the activities of these elements as either splicing enhancers or silencers or neutral sequences were observed to depend heavily on their exonic context. It is not clear whether the context dependence observed was due to the flexible activity of the sequences tested, or resulted from alterations to endogenous SREs in the reporters used. For example, all of the sequences inserted into the SXN minigene promoted exon skipping, which could be explained by disruption of an unknown ESE in the site where the foreign sequences were inserted (Goren et al. 2006). Other studies of ESSs have found much greater consistency of function in different exonic contexts (Wang et al. 2004, 2006).

INTRONIC SPLICING ENHANCERS AND SILENCERS

A number of intronic SREs are also known (for reviews, see Ladd and Cooper 2002; Zheng 2004), but fewer large-scale screens have been conducted for intronic elements, and

many more intronic elements likely remain to be identified. One well characterized ISE is the G triplet (GGG) or G run (G_n; $n \geq 3$), which often occur in clusters and can enhance recognition of adjacent 5'ss or 3'ss (McCullough and Berget 1997, 2000). This ISE is common in GC-rich introns and is conserved between human and mouse (Yeo et al. 2004). Intronic CA repeats in several cases can enhance splicing of upstream exons, probably through binding of hnRNP L (Hui et al. 2005; Hung et al. 2007). UGCAUG hexanucleotides or slight variations often occur downstream of neuron-specific exons and function as ISEs by binding to the brain- and muscle-specific splicing factors Fox-1/Fox-2 (Brudno et al. 2001; Jin et al. 2003; Minovitsky et al. 2005; Nakahata and Kawamoto 2005; Underwood et al. 2005). Pairs of YCAY motifs (Y = C or U) are recognized by the neuron-specific Nova family of splicing factors to regulate a large number of splicing events in the brain (Jensen et al. 2000; Ule et al. 2003). Interestingly, depending on their relative location in pre-mRNA, YCAY pairs can also function as either ESSs or ISSs (Hui et al. 2005; Ule et al. 2006). Such context dependence will be discussed in more detail below.

Characterized ISSs include binding sites for the splicing repressors PTB and hnRNP A1 (for reviews, see Zheng 2004; Matlin et al. 2005), CA-rich sequences bound by hnRNP L (Hui et al. 2005), specific octamers flanking exon IIIb of the *FGFR2* gene (Wagner et al. 2005), and two elements in intron 7 of human *SMN2*, a therapeutic target for spinal muscular atropy (Singh et al. 2006; Kashima et al. 2007).

Intronic elements (ISS and ISE) are likely of primary importance in regulating AS events, as the intronic regions surrounding alternative exons are far more conserved in mammals than those surrounding constitutive exons, out to a distance of 150 bp or more (Sorek and Ast 2003). Such increased conservation has been used to predict unannotated alternative exons (Sorek et al. 2004; Yeo et al. 2005), and to predict intronic SREs (Voelker and Berglund 2007; Yeo et al. 2007). In the predictions of intronic SREs, both groups identified the Fox-1/2 binding motif UGCAUG as the most conserved sequence near AS exons, and the elements identified by Yeo and colleagues included most known binding motifs associated with tissue-specific splicing factors. Remarkably, not only the Fox-1/Fox-2 splicing factor, but also its highly specific binding motif UGCAUG, are conserved from nematodes to mammals (Kabat et al. 2006), indicating an ancient role in splicing regulation. The analyses by Voelker and colleagues also suggested that the AU-rich motifs are strongly associated with constitutive splicing and may function as ISEs (Voelker and Berglund 2007); rigorous controls for GC content effects and experimental tests will be needed to firmly establish this result.

Splicing enhancers and silencers often function additively, with additional copies increasing their effect on splicing regulation (e.g., Huh and Hynes 1994; McCullough and Berget 1997; Chou et al. 2000; Wang et al. 2004; Zhang

and Chasin 2004), either because they increase the affinity of the associated factor (Dominguez and Allain 2006) or because they increase the copy numbers of the factor that are recruited, sometimes in synergistic fashion. Different SREs may also function cooperatively to regulate alternative splicing. For example, exonic UAGG motifs and intronic GGGG motifs overlapping the 5'ss can function cooperatively to silence the brain-specific CI cassette exon (exon 19) of the glutamate NMDA R1 receptor gene (Han et al. 2005); the purine-rich ESE in exon N1 of the c-src gene can cooperate with the downstream G-run ISE to increase exon N1 inclusion (Modafferi and Black 1999)

ELEMENTS THAT REGULATE INTRON RETENTION

Among the major types of alternative splicing, intron retention is unique in that it does not involve a choice between competing pairs of splice sites. Instead, it involves the choice between using a pair of splice sites to excise an intron and bypassing splicing to export intron-containing mRNA to the cytoplasm. Therefore, intron retention is probably regulated both by factors generally involved in splicing—some of which may also play a role in mRNA export—and by dedicated mRNA export factors (for review, see Reed and Cheng 2005).

Many SREs that affect other types of AS can regulate intron retention when properly situated (Sakabe and de Souza 2007). For example, G-run ISEs can promote splicing of a retained intron in human thrombopoietin (Marcucci et al. 2007), and a subset of FAS–ESS elements can activate splicing of a retained intron (Wang et al. 2006). In addition, the CA-rich motifs that bind to hnRNP L can enhance the splicing of multiple retained introns as revealed by splicing microarray analyses following RNAi against hnRNP L (Hung et al. 2007). Another intensely studied case is the *Drosophila* P-element third intron (INV3), which is retained in somatic cells but is fully spliced in germ cells (Rio 1991). One element that promotes intron retention is a 5'ss-like sequence in the upstream exon (Siebel et al. 1992); similar 5'ss-like sequences have ESS activity in the ATM gene (Pagani et al. 2002) and were identified in the FAS–ESS screen (Wang et al. 2004). Some protein factors that regulate INV3 retention are homologous with vertebrate proteins that regulate splicing by binding to splicing enhancers or silencers (Siebel et al. 1994, 1995; Min et al. 1997). Therefore, it is reasonable to speculate that most general SREs can also regulate intron retention since they usually have a direct effect on splice site activity.

Perhaps the best-studied systems involving regulated intron retention occur in retroviruses (e.g., HIV) that need to balance between splicing of mRNA and transport of unspliced genomic RNA to the cytoplasm for packaging (Fischer et al. 1999). Retroviruses use many general SREs to regulate splicing of their RNA (for review, see Stoltzfus and

Madsen 2006) and also use RNA elements to regulate nuclear export of unspliced RNA (for reviews, see Pollard and Malim 1998; Cullen 2003). How HIV RNAs balance between the two pathways is not completely clear, but some common elements might serve as a functional bridge between the two pathways by participating in the regulation of both splicing and RNA export (e.g., an ESS recognized by hnRNP A1 [Asai et al. 2003] and the Rev response elements [Pongoski et al. 2002]).

CONTEXT DEPENDENCE OF SRES

It was realized early in the study of splicing that the activities of SREs may depend on the relative locations of the elements in pre-mRNAs. For example, G triplets commonly enhance splicing from intronic locations (McCullough and Berget 1997), but they function as splicing silencers when located in exons (Chen et al. 1999). A more subtle effect is that the activity of SR proteins to promote splicing depends on the distance between the ESE and the adjacent splice site (Graveley et al. 1998). In some cases, the SREs identified in one AS event fail to regulate splicing when located in a heterologous exon/intron context. These phenomena may be collectively described as the "context dependence" of SREs and can be considered to fall into two categories: (1) location-dependent activity, in which activity varies with relative positions in the pre-mRNA (Fig. 2A); and (2) gene-dependent activity, in which activity observed in one gene is lost when the SRE is moved to another (Fig. 2B).
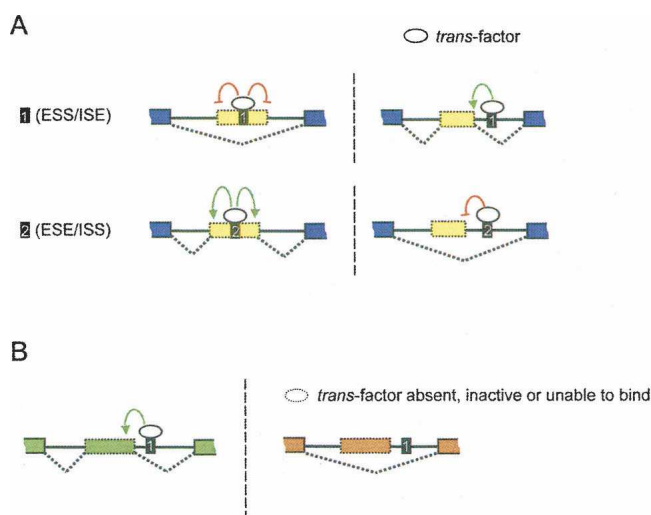

FIGURE 2. Schematic of two types of context dependence for SREs. (*A*) Location-dependent activity of SREs, in which activity varies with pre-mRNA position. For example, G runs can function as both ESSs and ISEs (*upper* panel), and some SR protein binding sites can function as both ESEs and ISSs (*lower* panel). (Yellow) Alternative exons; (blue) constitutive exons. (*B*) Gene-dependent activity of SREs, in which activity observed in one gene is lost when the SRE is moved to another. Different genes are shown in different colors.

Location-dependent activity reflects the flexibility of splicing regulatory factors in their interactions with core splicing machinery. Given the size and complexity of spliceosomes and their multistep assembly, it is perhaps not surprising that the activities of factors that interact with core spliceosome components would vary depending on their locations relative to these components. For example, the G-run-binding factor hnRNP H can participate in a splicing enhancer complex when G runs are located downstream of the 5′ss (Chou et al. 1999; Caputi and Zahler 2001; Hastings et al. 2001; Schaub et al. 2007) but can inhibit splicing when similar sequences are located in an exon (Chen et al. 1999; Caputi and Zahler 2001).

Although the degree of context dependence of SREs may initially appear chaotic or confusing, there are usually patterns and rules that summarize this activity. For example, the dual activity of G runs as ISEs and ESSs can be conceptually interpreted as defining a region of pre-mRNA as intron rather than exon. We have observed that most ESSs inhibit the intron-proximal 5′ss or 3′ss when located between competing alternative splice sites, and that some ESSs, including G runs, can promote splicing of retained introns when located inside the intron (Wang et al. 2006). The underlying mechanism for these activities of G runs could involve inhibition by hnRNP F/H of exon definition complex formation "across" the site of binding. As noted above, clusters of YCAY motifs that bind the Nova family of neuron-specific splicing factors can function as ESSs, ISEs, or ISSs, depending on their position relative to the regulated exon. An "RNA map" describing these context-dependent activities has been generated that predicts the direction of Nova-dependent splicing regulation with very high accuracy (Ule et al. 2006). This study, including also a dissection of the mechanisms of splicing activation and repression by Nova, represents an important demonstration that the set of splicing regulatory activities of a factor can be rationalized.

Analogous to the common overlap between ESSs and ISEs, some SR proteins can promote splicing when bound to sites (ESEs) in exons, and also inhibit splicing when bound to intronic (ISS) sites (Kanopka et al. 1996; Ibrahim el et al. 2005; Buratti et al. 2007). Results of a FAS-based screen for ISSs suggest that overlap between ISS and ESE activity could be a general phenomenon (Z. Wang, J. Zhang, X. Xiao, and C.B. Burge, unpubl.). A recent report suggested the counterintuitive result that some elements (designated "ESRs") can function as both ESSs and ESEs depending on exonic context (Goren et al. 2006). However, in our own analyses, large sets of ESEs and ESSs identified using entirely different methods (statistically-based and experimentally-based) were nonoverlapping, suggesting that a single sequence is unlikely to commonly function as both an ESS and an ESE (Fairbrother et al. 2002; Wang et al. 2004).

For intronic elements, the situation is much more complex. For example, Nova binding sequences commonly function as ISEs when located near the splice sites of the intron downstream of the regulated exon and as ISSs when located near the 5′ss of the upstream intron (Ule et al. 2006). Similarly, the Fox-1/2 binding site (UGCAUG) can function as an ISS when located upstream of an exon (Jin et al. 2003) and as an ISE when located downstream of an alternative exon (Huh and Hynes 1994; Lim and Sharp 1998; Jin et al. 2003). CA repeats bound by hnRNP L can function either as ISEs or ISSs depending on their distance to the upstream exon (Hui et al. 2005).

Gene-dependent activity of SREs may often result from factors that determine whether a *cis*-element can be recognized by the associated factor. Because most splicing regulatory *cis*-elements are short RNA oligonucleotides, they tend to be very abundant in pre-mRNA. Based on a recent comparative analysis of exonic splicing elements, it was estimated that about 1000 hexamers can function as either ESSs or ESEs (Stadler et al. 2006). Therefore, every pre-mRNA will contain many potential splicing regulatory motifs, only a fraction of which may be recognized by *trans*-factors. This is analogous to transcriptional regulation in that a search near a promoter region will predict a large number of transcription factor binding sites, only a small fraction of which are functional. Determining which *cis*-elements will be recognized remains an open problem. One obvious factor is the local structure of the pre-mRNA, which could affect *cis*-element accessibility. For example, RNA structure in the fibronectin EDA exon appears to affect recognition of an ESE by SR proteins, and this structure accounts for much of the difference between the splicing behavior of the orthologous exons in human and mouse (Buratti et al. 2004). A recent analysis of experimentally determined SREs suggested that these SRE motifs are significantly enriched in single-stranded regions of pre-mRNA (Hiller et al. 2007), suggesting that pre-mRNA secondary structure may play a general role in determining SRE function. Some factors, including Nova and muscleblind-like splicing factors, recognize their cognate motifs in specific structural contexts (Jensen et al. 2000; Warf and Berglund 2007).

In addition to affecting SRE accessibility, secondary structures may also regulate splicing by directly affecting splice site accessibility. For example, a stem–loop structure at the 5′ss of exon 10 of the human *tau* gene directly affects the activity of the 5′ss, with stabilization of this structure decreasing exon 10 inclusion and destabilization of this stem–loop increasing exon 10 inclusion (Donahue et al. 2006). Another remarkable example was found in exon 6 of the *Drosophila* Dscam gene, where the intronic secondary structure ensures that inclusion of 48 alternative exons occurs in a mutually exclusive fashion (Graveley 2005). However, it is unclear whether examples like this represent unusual cases or are a general rule, and even for tau exon 10 the role of structure in splicing regulation is debated (D'Souza and Schellenberg 2002). Spliceosomes contain multiple RNA helicase components that can unwind RNA structures and remodel RNA/protein complexes (for review, see Bleichert and Baserga 2007). Although the primary function of these spliceosome-associated helicases appears to be rearrangement of snRNA/snRNA, snRNA/pre-mRNA, and snRNA/protein interactions in the spliceosome, at least one appears to function at an earlier stage, influencing the alternative splicing of the CD44 pre-mRNA, possibly by remodeling its structure and/or associated protein complexes (Honig et al. 2002; Lee 2002). By analogy to translation, where ribosome-associated helicases alter mRNA structure to facilitate translocation, it is tempting to speculate that spliceosome-associated helicases may also disrupt splicing-inhibitory structure in pre-mRNAs. As the general roles of structure in splicing regulation are still not clearly defined, large-scale measurements of RNA structures should be very valuable. Some recently developed methods like selective 2′-hydroxyl acylation and primer extension (SHAPE) provide new possibilities for high-throughput measurements of RNA structures (Merino et al. 2005), but high-throughput methods capable of assessing structure in vivo still need to be developed.

The activities of SREs are of course dependent on the presence and activity of the associated *trans*-factors, and this dependence is probably responsible for most of the tissue- or cell-type-specific splicing. SRE activity can also respond to external signals that alter the expression or activity of specific *trans*-factors (for review, see Shin and Manley 2004). The mechanisms by which inducible splicing responds to different stimuli can be very diverse, including, but not limited to, neuronal depolarization through CaM kinase (Xie and Black 2001; Lee et al. 2007), heat shock response through SR protein dephosphorylation (Shin and Manley 2002; Shin et al. 2004), response to T cell activation through an inducible ESS bound by hnRNP L (Rothrock et al. 2003, 2005), and response to neuronal excitation through a UAGG motif recognized by hnRNP A1 (Han et al. 2005).

GLOBAL ANALYSES OF SPLICING REGULATION

A number of new technologies have been developed recently for genome-wide analysis of AS. Several microarray platforms have been designed to distinguish between different splicing isoforms and to detect AS at a genomic scale. "Exon junction arrays" represent one early design, with high-density oligonucleotide probes targeted to the junctions between consecutive exons (Johnson et al. 2003). Other designs include the use of probe sets to target bodies and junctions of constitutive and alternative exons (Clark et al. 2002; Pan et al. 2004; Blanchette et al. 2005; Sugnet et al. 2006), as well as the use of bead-based fiber-optic microarray platforms with high detection sensitivity (Yeakley et al. 2002). These designs have facilitated analyses

of genome-wide alternative splicing in human, mouse, and chimp (Srinivasan et al. 2005; Boutz et al. 2007; Calarco et al. 2007; Ip et al. 2007; Ni et al. 2007), as well as detection of the global impact of specific splicing factors or environmental stimuli on splicing regulation (Park et al. 2004; Hung et al. 2007; Makeyev et al. 2007; Pleiss et al. 2007a,b). A very high-density "exon array" containing probes in essentially all known and predicted human exons was recently developed by Affymetrix (Gardina et al. 2006). Generally speaking, these array platforms appear able to identify a subset of AS events with high confidence, but have an unknown and probably substantial rate of false negatives.

The systematic identification of RNA targets for different *trans*-factors can also be achieved by some new approaches such as a cross-linking/immunoprecipitation (CLIP) (Ule et al. 2003), RNP immunoprecipitation (RIP) (Keene et al. 2006; Townley-Tilson et al. 2006), and genomic SELEX (Lorenz et al. 2006). These analyses have the potential to identify regulatory targets of a factor and can be applied genome-wide when coupled with microarray or high-throughput sequencing technologies. Analysis of the target sequences will help to define the sequence determinants of binding, and may also help to identify cooperative or antagonistic relationships between different factors. Since only a subset of binding events confers regulatory activity, it is important to also have evidence of regulation. Such evidence can be obtained from knockout/knockdown or overexpression of the factor, which can be also applied on a genome-wide scale when coupled with isoform-specific microarrays (Blanchette et al. 2005), or with high-throughput sequencing.

based only on the information that is accessible and known (or strongly implicated) to be recognized by the nuclear pre-mRNA splicing machinery. Thus, in addition to predictive accuracy, faithfulness to the in vivo mechanism is important in splicing simulation, and an important application of splicing simulation is to evaluate how different sequence features contribute to the determination of splicing specificity. This philosophy is different from that of most gene prediction/gene finding algorithms, whose goal is generally to make the most accurate prediction of exon and gene locations using whatever information is available. For example, most current gene finders use reading frame consistency of exons and/or cross-species conservation in their predictions, information that is inaccessible to the spliceosome.

The ExonScan algorithm, illustrated in Figure 3, involves scanning the pre-mRNA for pairs of nearby potential 3′ss and 5′ss, and scoring these sites and nearby exonic and intronic SREs. Log-odds scoring is used, which rewards elements based on their statistical enrichment in the relevant location relative to their background frequency in the genome and naturally assigns positive scores to enhancers and negative scores to silencers. The results generated with this algorithm provided important clues to the in vivo function of ESSs (Wang et al. 2004). Such simulation should be improved by use of more refined sets of sequence elements and context-dependent scoring schemes, and by consideration of the functional interactions between different elements, as inferred from experimental assays or from patterns of coevolution (Xiao et al. 2007).
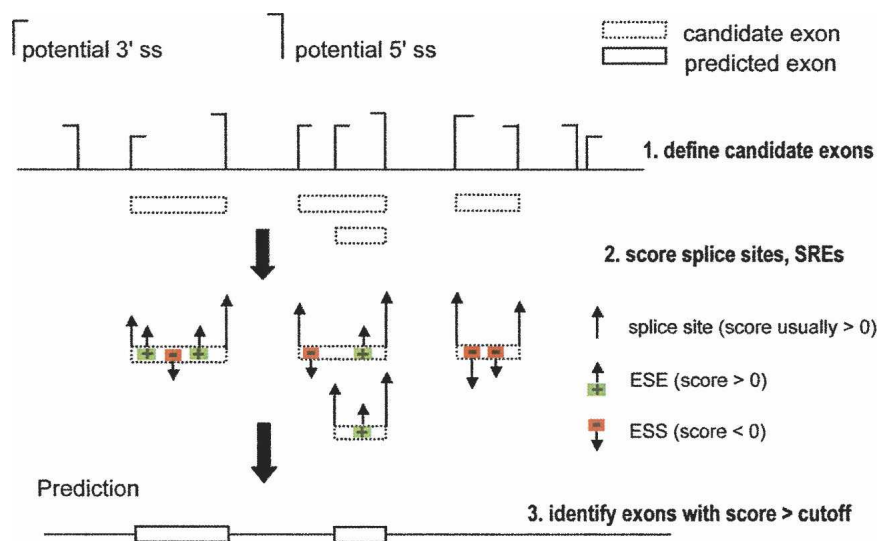
SPLICING SIMULATION: PUTTING THE PIECES TOGETHER

An important long-term goal in the splicing field is to determine a "splicing code:" a set of rules that can predict the splicing pattern of any primary transcript sequence (Fu 2004; Matlin et al. 2005). As large "parts lists" of splicing regulatory *cis*-elements are identified, and identities and functions of associated *trans*-acting factors are determined, a natural next step is to assemble the available information into a predictive framework to simulate the recognition of exons and introns that occurs during splicing. An initial approach to this very challenging problem was the development of the general splicing simulation algorithm, ExonScan (Fig. 3; Wang et al. 2004). The goal of splicing simulation is to predict the splicing patterns of transcripts

FIGURE 3. Schematic diagram of a general splicing simulator, ExonScan. (Green boxes) Splicing enhancers, (red boxes) silencers. For simplicity, only ESSs and ESEs are indicated. Step 1, Identify the exon candidate as a pair of splice site (shown as brackets) that is 50–250 bp apart. Step 2, Score the splice sites and splicing enhancers and silencers, and sum up for the score of each exon candidate. Step 3, Determine the threshold and make the final prediction.

SPLICING REGULATORY NETWORK AS A SUBNETWORK OF GENE REGULATION

The splicing regulatory network is part of a larger network of gene regulation with which it is linked both physically and functionally. Most introns are thought to be spliced cotranscriptionally (de la Mata et al. 2003; Das et al. 2006; Hicks et al. 2006), and SR proteins associate with the C-terminal domain of RNA polymerase II (de la Mata and Kornblihtt 2006; Das et al. 2007). As a downstream event of transcription, AS can be regulated by factors that affect transcription. For example, a mutated RNA polymerase II with a slow elongation rate had been shown to affect the AS of endogenous genes (de la Mata et al. 2003), and the insertion of a transcription-pausing element, MAZ4, into a minigene construct can influence splicing of an alternative exon (Robson-Dixon and Garcia-Blanco 2004). Therefore, it is tempting to speculate that some of the identified SREs could act through effects on transcription, e.g., functioning as a transcription pause site to promote the inclusion of a weak upstream exon. In addition, the whole process of mRNA biosynthesis and processing, including transcription, 5′ capping, splicing, polyadenylation, and transport is extensively coupled, with a number of factors involved in more than one step of this process (Hieronymus and Silver 2004; Kornblihtt et al. 2004). The splicing regulation network should therefore be viewed as a specialized subnetwork of a more general gene regulatory netwok.

PERSPECTIVE

Sorting out the complex network that controls constitutive and alternative splicing represents a major challenge for post-genomic biology. A promising route to such global understanding is a "bottom-up" approach, involving systematic identification of the *cis*-regulatory components of the network, determination of rules—context dependence or otherwise—for their activity and functional interactions with other elements, and integration of these rules into simulation algorithms representing successive approximations to the splicing code. Beyond the challenges discussed above, another important challenge will be to learn how to generalize the *cis*-elements, factors, and rules identified in one system (e.g., in one cell type or developmental stage) to others. This will undoubtedly require new experimental approaches that can be readily adapted to a variety of systems.

Given the complexity of splicing regulation, the splicing code will not have the simple tabular form of the genetic code. Instead, it may look more like the U.S. tax code (i.e., IRS form 1040), with a variety of tables and subtables that are applicable in different circumstances, and a certain amount of arithmetic required to obtain the correct answer.

REFERENCES

Asai, K., Platt, C., and Cochrane, A. 2003. Control of HIV-1 env RNA splicing and transport: Investigating the role of hnRNP A1 in exon splicing silencer (ESS3a) function. *Virology* **314:** 229–242.

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

Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72:** 291–336.

Blanchette, M., Green, R.E., Brenner, S.E., and Rio, D.C. 2005. Global analysis of positive and negative pre-mRNA splicing regulators in *Drosophila*. *Genes & Dev.* **19:** 1306–1314.

Bleichert, F. and Baserga, S.J. 2007. The long unwinding road of RNA helicases. *Mol. Cell* **27:** 339–352.

Blencowe, B.J. 2006. Alternative splicing: New insights from global analyses. *Cell* **126:** 37–47.

Boutz, P.L., Stoilov, P., Li, Q., Lin, C.H., Chawla, G., Ostrow, K., Shiue, L., Ares Jr., M., and Black, D.L. 2007. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes & Dev.* **21:** 1636–1652.

Brudno, M., Gelfand, M.S., Spengler, S., Zorn, M., Dubchak, I., and Conboy, J.G. 2001. Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Res.* **29:** 2338–2348. doi: 10.1093/nar/29.11.2338.

Buratti, E., Muro, A.F., Giombi, M., Gherbassi, D., Iaconcig, A., and Baralle, F.E. 2004. RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol. Cell. Biol.* **24:** 1387–1400.

Buratti, E., Stuani, C., De Prato, G., and Baralle, F.E. 2007. SR protein-mediated inhibition of CFTR exon 9 inclusion: Molecular characterization of the intronic splicing silencer. *Nucleic Acids Res.* **35:** 4359–4368. doi: 10.1093/nar/gkm444.

Calarco, J.A., Xing, Y., Caceres, M., Calarco, J.P., Xiao, X., Pan, Q., Lee, C., Preuss, T.M., and Blencowe, B.J. 2007. Global analysis of alternative splicing differences between humans and chimpanzees. *Genes & Dev.* **21:** 2963–2975.

Caputi, M. and Zahler, A.M. 2001. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H′/F/2H9 family. *J. Biol. Chem.* **276:** 43850–43859.

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.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31:** 3568–3571. doi: 10.1093/nar/gkg616.

Cascino, I., Fiucci, G., Papoff, G., and Ruberti, G. 1995. Three functional soluble forms of the human apoptosis-inducing Fas molecule are produced by alternative splicing. *J. Immunol.* **154:** 2706–2713.

Chasin, L. 2007. Searching for splicing motifs. In *Alternative splicing in the postgenomic era* (eds. B.J. Blencowe and B.R. Graveley), pp. 85–106. Landes Biosciences, Austin, TX.

Chen, C.D., Kobayashi, R., and Helfman, D.M. 1999. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat β-tropomyosin gene. *Genes & Dev.* **13:** 593–606.

Chou, M.Y., Rooke, N., Turck, C.W., and Black, D.L. 1999. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell. Biol.* **19:** 69–77.

Chou, M.Y., Underwood, J.G., Nikolic, J., Luu, M.H., and Black, D.L. 2000. Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol. Cell* **5:** 949–957.

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

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.

Cullen, B.R. 2003. Nuclear mRNA export: Insights from virology. *Trends Biochem. Sci.* **28:** 419–424.

D'Souza, I. and Schellenberg, G.D. 2002. Tau exon 10 expression involves a bipartite intron 10 regulatory sequence and weak 5′ and 3′ splice sites. *J. Biol. Chem.* **277:** 26587–26599.

Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., and Reed, R. 2006. Functional coupling of RNAP II transcription to spliceosome assembly. *Genes & Dev.* **20:** 1100–1109.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. 2007. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* **26:** 867–881.

de la Mata, M. and Kornblihtt, A.R. 2006. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat. Struct. Mol. Biol.* **13:** 973–980.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. 2003. A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* **12:** 525–532.

Demir, E. and Dickson, B.J. 2005. fruitless splicing specifies male courtship behavior in *Drosophila*. *Cell* **121:** 785–794.

Dominguez, C. and Allain, F.H. 2006. NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: A novel mode of RNA recognition. *Nucleic Acids Res.* **34:** 3634–3645. doi: 10.1093/nar/gkl488.

Donahue, C.P., Muratore, C., Wu, J.Y., Kosik, K.S., and Wolfe, M.S. 2006. Stabilization of the tau exon 10 stem–loop alters pre-mRNA splicing. *J. Biol. Chem.* **281:** 23302–23306.

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.

Fischer, U., Pollard, V.W., Luhrmann, R., Teufel, M., Michael, M.W., Dreyfuss, G., and Malim, M.H. 1999. Rev-mediated nuclear export of RNA is dominant over nuclear retention and is coupled to the Ran-GTPase cycle. *Nucleic Acids Res.* **27:** 4128–4134.

Fu, X.D. 2004. Towards a splicing code. *Cell* **119:** 736–738.

Gao, K., Masuda, A., Matsuura, T., and Ohno, K. 2008. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.* PMID: 18285363. doi: 10.1093/nar/gkn073.

Gardina, P.J., Clark, T.A., Shimada, B., Staples, M.K., Yang, Q., Veitch, J., Schweitzer, A., Awad, T., Sugnet, C., Dee, S., et al. 2006. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **7:** 325. doi: 10.1186/1471-2164-7-325.

Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., Pupko, T., and Ast, G. 2006. Comparative analysis identifies exonic splicing regulatory sequences—The complex definition of enhancers and silencers. *Mol. Cell* **22:** 769–781.

Graveley, B.R. 2000. Sorting out the complexity of SR protein functions. *RNA* **6:** 1197–1211.

Graveley, B.R. 2005. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123:** 65–73.

Graveley, B.R. and Maniatis, T. 1998. Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol. Cell* **1:** 765–771.

Graveley, B.R., Hertel, K.J., and Maniatis, T. 1998. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J.* **17:** 6747–6756.

Han, K., Yeo, G., An, P., Burge, C.B., and Grabowski, P.J. 2005. A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol.* **3:** e158. doi: 10.1371/journal.pbio.0030158.

Hastings, M.L., Wilson, C.M., and Munroe, S.H. 2001. A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA. *RNA* **7:** 859–874.

Hertel, K.J. and Graveley, B.R. 2005. RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem. Sci.* **30:** 115–118.

Hicks, M.J., Yang, C.R., Kotlajich, M.V., and Hertel, K.J. 2006. Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. *PLoS Biol.* **4:** e147. doi: 10.1371/journal.pbio.0040147.

Hieronymus, H. and Silver, P.A. 2004. A systems view of mRNP biology. *Genes & Dev.* **18:** 2845–2860.

Hiller, M., Zhang, Z., Backofen, R., and Stamm, S. 2007. Pre-mRNA secondary structures influence exon recognition. *PLoS Genet.* **3:** e204. doi: 10.1371/journal.pgen.0030204.

Honig, A., Auboeuf, D., Parker, M.M., O'Malley, B.W., and Berget, S.M. 2002. Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. *Mol. Cell. Biol.* **22:** 5698–5707.

House, AE. and Lynch, KW. 2008. Regulation of alternative splicing: More than just the ABCs. *J. Biol. Chem.* **283:** 1217–1221.

Huh, G.S. and Hynes, R.O. 1994. Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element. *Genes & Dev.* **8:** 1561–1574.

Hui, J., Hung, L.H., Heiner, M., Schreiner, S., Neumuller, N., Reither, G., Haas, S.A., and Bindereif, A. 2005. Intronic CA-repeat and CA-rich elements: A new class of regulators of mammalian alternative splicing. *EMBO J.* **24:** 1988–1998.

Hung, L.H., Heiner, M., Hui, J., Schreiner, S., Benes, V., and Bindereif, A. 2007. Diverse roles of hnRNP L in mammalian mRNA processing: A combined microarray and RNAi analysis. *RNA* **14:** 284–296.

Ibrahim el, C., Schaal, T.D., Hertel, K.J., Reed, R., and Maniatis, T. 2005. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc. Natl. Acad. Sci.* **102:** 5002–5007.

Ip, J.Y., Tong, A., Pan, Q., Topp, J.D., Blencowe, B.J., and Lynch, K.W. 2007. Global analysis of alternative splicing during T-cell activation. *RNA* **13:** 563–572.

Izquierdo, J.M., Majos, N., Bonnal, S., Martinez, C., Castelo, R., Guigo, R., Bilbao, D., and Valcarcel, J. 2005. Regulation of fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol. Cell* **19:** 475–484.

Jensen, K.B., Musunuru, K., Lewis, H.A., Burley, S.K., and Darnell, R.B. 2000. The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. *Proc. Natl. Acad. Sci.* **97:** 5740–5745.

Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K., and Inoue, K. 2003. A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.* **22:** 905–912.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302:** 2141–2144.

Jurica, M.S. and Moore, M.J. 2003. Pre-mRNA splicing: Awash in a sea of proteins. *Mol. Cell* **12:** 5–14.

Kabat, J.L., Barberan-Soler, S., McKenna, P., Clawson, H., Farrer, T., and Zahler, A.M. 2006. Intronic alternative splicing regulators identified by comparative genomics in nematodes. *PLoS Comput. Biol.* **2:** e86. doi: 10.1371/journal.pcbi.0020086.

Kanopka, A., Muhlemann, O., and Akusjarvi, G. 1996. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381:** 535–538.

Kashima, T., Rao, N., and Manley, J.L. 2007. An intronic element contributes to splicing repression in spinal muscular atrophy. *Proc. Natl. Acad. Sci.* **104:** 3426–3431.

Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. 2006. RIP-Chip: The isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat. Protoc.* **1:** 302–307.

Kol, G., Lev-Maor, G., and Ast, G. 2005. Human–mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. *Hum. Mol. Genet.* **14:** 1559–1568.

Konarska, M.M. and Query, C.C. 2005. Insights into the mechanisms of splicing: More lessons from the ribosome. *Genes & Dev.* **19:** 2255–2260.

Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J., and Nogues, G. 2004. Multiple links between transcription and splicing. *RNA* **10:** 1489–1498.

Ladd, A.N. and Cooper, T.A. 2002. Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol.* **3:** reviews0008.1–0008.16. doi: 10.1186/gb-2002-3-11-reviews 0008.

Lee, C.G. 2002. RH70, a bidirectional RNA helicase, co-purifies with U1snRNP. *J. Biol. Chem.* **277:** 39679–39683.

Lee, J.A., Xing, Y., Nguyen, D., Xie, J., Lee, C.J., and Black, D.L. 2007. Depolarization and CaM kinase IV modulate NMDA receptor splicing through two essential RNA elements. *PLoS Biol.* **5:** e40. doi: 10.1371/journal.pbio.0050040.

Lim, L.P. and Burge, C.B. 2001. A computational analysis of sequence features involved in recognition of short introns. *Proc. Natl. Acad. Sci.* **98:** 11193–11198.

Lim, S.R. and Hertel, K.J. 2004. Commitment to splice site pairing coincides with A complex formation. *Mol. Cell* **15:** 477–483.

Lim, L.P. and Sharp, P.A. 1998. Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. *Mol. Cell. Biol.* **18:** 3900–3906.

Liu, H.X., Zhang, M., and Krainer, A.R. 1998. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes & Dev.* **12:** 1998–2012.

Liu, H.X., Chew, S.L., Cartegni, L., Zhang, M.Q., and Krainer, A.R. 2000. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol. Cell. Biol.* **20:** 1063–1071.

Lorenz, C., von Pelchrzim, F., and Schroeder, R. 2006. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat. Protoc.* **1:** 2204–2212.

Makeyev, E.V., Zhang, J., Carrasco, M.A., and Maniatis, T. 2007. The microRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27:** 435–448.

Marcucci, R., Baralle, F.E., and Romano, M. 2007. Complex splicing control of the human Thrombopoietin gene by intronic G runs. *Nucleic Acids Res.* **35:** 132–142. doi: 10.1093/nar/gkl965.

Matlin, A.J., Clark, F., and Smith, C.W. 2005. Understanding alternative splicing: Toward a cellular code. *Nat. Rev. Mol. Cell Biol.* **6:** 386–398.

McCullough, A.J. and Berget, S.M. 1997. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol. Cell. Biol.* **17:** 4562–4571.

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.

Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. 2005. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127:** 4223–4231.

Min, H., Turck, C.W., Nikolic, J.M., and Black, D.L. 1997. A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes & Dev.* **11:** 1023–1036.

Minovitsky, S., Gee, S.L., Schokrpur, S., Dubchak, I., and Conboy, J.G. 2005. The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *Nucleic Acids Res.* **33:** 714–724. doi: 10.1093/nar/gki210.

Modafferi, E.F. and Black, D.L. 1999. Combinatorial control of a neuron-specific exon. *RNA* **5:** 687–706.

Nakahata, S. and Kawamoto, S. 2005. Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res.* **33:** 2078–2089. doi: 10.1093/nar/gki338.

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

Nasim, F.U., Hutchison, S., Cordeau, M., and Chabot, B. 2002. High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5′ splice site selection in support of a common looping out and repression mechanism. *RNA* **8:** 1078–1089.

Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E., and Ares Jr., M. 2007. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & Dev.* **21:** 708–718.

Nilsen, T.W. 2003. The spliceosome: The most complex macromolecular machine in the cell? *Bioessays* **25:** 1147–1149.

Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dork, T., and Baralle, F.E. 2002. A new type of mutation causes a splicing defect in ATM. *Nat. Genet.* **30:** 426–429.

Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A.L., Mohammad, N., Babak, T., Siu, H., Hughes, T.R., Morris, Q.D., et al. 2004. Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* **16:** 929–941.

Park, J.W., Parisky, K., Celotto, A.M., Reenan, R.A., and Graveley, B.R. 2004. Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc. Natl. Acad. Sci.* **101:** 15974–15979.

Pfarr, N., Prawitt, D., Kirschfink, M., Schroff, C., Knuf, M., Habermehl, P., Mannhardt, W., Zepp, F., Fairbrother, W., Loos, M., et al. 2005. Linking C5 deficiency to an exonic splicing enhancer mutation. *J. Immunol.* **174:** 4172–4177.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M., and Guthrie, C. 2007a. Rapid, transcript-specific changes in splicing in response to environmental stress. *Mol. Cell* **27:** 928–937.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M., and Guthrie, C. 2007b. Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biol.* **5:** e90. doi: 10.1371/journal.pbio.0050090.

Pollard, V.W. and Malim, M.H. 1998. The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **52:** 491–532.

Pongoski, J., Asai, K., and Cochrane, A. 2002. Positive and negative modulation of human immunodeficiency virus type 1 Rev function by *cis* and *trans* regulators of viral RNA splicing. *J. Virol.* **76:** 5108–5120.

Pozzoli, U. and Sironi, M. 2005. Silencers regulate both constitutive and alternative splicing events in mammals. *Cell. Mol. Life Sci.* **62:** 1579–1604.

Reed, R. and Cheng, H. 2005. TREX, SR proteins and export of mRNA. *Curr. Opin. Cell Biol.* **17:** 269–273.

Rio, D.C. 1991. Regulation of *Drosophila* P element transposition. *Trends Genet.* **7:** 282–287.

Robberson, B.L., Cote, G.J., and Berget, S.M. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10:** 84–94.

Robson-Dixon, N.D. and Garcia-Blanco, M.A. 2004. MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. *J. Biol. Chem.* **279:** 29075–29084.

Rothrock, C., Cannon, B., Hahm, B., and Lynch, K.W. 2003. A conserved signal-responsive sequence mediates activation-induced alternative splicing of CD45. *Mol. Cell* **12:** 1317–1324.

Rothrock, C.R., House, A.E., and Lynch, K.W. 2005. HnRNP L represses exon splicing via a regulated exonic splicing silencer. *EMBO J.* **24:** 2792–2802.

Sakabe, N.J. and de Souza, S.J. 2007. Sequence features responsible for intron retention in human. *BMC Genomics* **8:** 59. doi: 10.1186/1471-2164-8-59.

Schaal, T.D. and Maniatis, T. 1999. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol. Cell. Biol.* **19:** 261–273.

Schaub, M.C., Lopez, S.R., and Caputi, M. 2007. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. *J. Biol. Chem.* **282:** 13617–13626.

Sharma, S., Falick, A.M., and Black, D.L. 2005. Polypyrimidine tract binding protein blocks the 5′ splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol. Cell* **19:** 485–496.

Shaw, S.D., Chakrabarti, S., Ghosh, G., and Krainer, A.R. 2007. Deletion of the N-terminus of SF2/ASF permits RS-domain-independent pre-mRNA splicing. *PLoS ONE* **2:** e854. doi: 10.1371/journal.pone.0000854.

Shen, H. and Green, M.R. 2006. RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes & Dev.* **20:** 1755–1765.

Shen, H., Kan, J.L., and Green, M.R. 2004. Arginine-serine-rich domains bound at splicing enhancers contact the branch-point to promote prespliceosome assembly. *Mol. Cell* **13:** 367–376.

Shin, C. and Manley, J.L. 2002. The SR protein SRp38 represses splicing in M phase cells. *Cell* **111:** 407–417.

Shin, C. and Manley, J.L. 2004. Cell signalling and the control of pre-mRNA splicing. *Nat. Rev. Mol. Cell Biol.* **5:** 727–738.

Shin, C., Feng, Y., and Manley, J.L. 2004. Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature* **427:** 553–558.

Siebel, C.W., Fresco, L.D., and Rio, D.C. 1992. The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: Multiprotein complexes at an exon pseudo-5′ splice site control U1 snRNP binding. *Genes & Dev.* **6:** 1386–1401.

Siebel, C.W., Kanaar, R., and Rio, D.C. 1994. Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes & Dev.* **8:** 1713–1725.

Siebel, C.W., Admon, A., and Rio, D.C. 1995. Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes & Dev.* **9:** 269–283.

Singh, N.K., Singh, N.N., Androphy, E.J., and Singh, R.N. 2006. Splicing of a critical exon of human Survival Motor Neuron is regulated by a unique silencer element located in the last intron. *Mol. Cell. Biol.* **26:** 1333–1346.

Sorek, R. and Ast, G. 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13:** 1631–1637.

Sorek, R., Shemesh, R., Cohen, Y., Basechess, O., Ast, G., and Shamir, R. 2004. A non-EST-based method for exon-skipping prediction. *Genome Res.* **14:** 1617–1623.

Srinivasan, K., Shiue, L., Hayes, J.D., Centers, R., Fitzwater, S., Loewen, R., Edmondson, L.R., Bryant, J., Smith, M., Rommelfanger, C., et al. 2005. Detection and measurement of alternative splicing using splicing-sensitive microarrays. *Methods* **37:** 345–359.

Stadler, M.B., Shomron, N., Yeo, G.W., Schneider, A., Xiao, X., and Burge, C.B. 2006. Inference of splicing regulatory activities by sequence neighborhood analysis. *PLoS Genet.* **2:** e191. doi: 10.1371/journal.pgen.0020191.

Sterner, D.A., Carlo, T., and Berget, S.M. 1996. Architectural limits on split genes. *Proc. Natl. Acad. Sci.* **93:** 15081–15085.

Stoltzfus, C.M. and Madsen, J.M. 2006. Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. *Curr. HIV Res.* **4:** 43–55.

Sugnet, C.W., Srinivasan, K., Clark, T.A., O'Brien, G., Cline, M.S., Wang, H., Williams, A., Kulp, D., Blume, J.E., Haussler, D., et al. 2006. Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* **2:** e4. doi: 10.1371/journal.pcbi.0020004.

Sun, H. and Chasin, L.A. 2000. Multiple splicing defects in an intronic false exon. *Mol. Cell. Biol.* **20:** 6414–6425.

Tian, H. and Kole, R. 1995. Selection of novel exon recognition elements from a pool of random sequences. *Mol. Cell. Biol.* **15:** 6291–6298.

Townley-Tilson, W.H., Pendergrass, S.A., Marzluff, W.F., and Whitfield, M.L. 2006. Genome-wide analysis of mRNAs bound to the histone stem–loop binding protein. *RNA* **12:** 1853–1867.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302:** 1212–1215.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J., and Darnell, R.B. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444:** 580–586.

Underwood, J.G., Boutz, P.L., Dougherty, J.D., Stoilov, P., and Black, D.L. 2005. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol. Cell. Biol.* **25:** 10005–10016.

Voelker, R.B. and Berglund, J.A. 2007. A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Res.* **17:** 1023–1033.

Wagner, E.J., Baraniak, A.P., Sessions, O.M., Mauger, D., Moskowitz, E., and Garcia-Blanco, M.A. 2005. Characterization of the intronic splicing silencers flanking FGFR2 exon IIIb. *J. Biol. Chem.* **280:** 14017–14027.

Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell* **119:** 831–845.

Wang, Z., Xiao, X., Van Nostrand, E., and Burge, C.B. 2006. General and specific functions of exonic splicing silencers in splicing control. *Mol. Cell* **23:** 61–70.

Warf, M.B. and Berglund, J.A. 2007. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA* **13:** 2238–2251.

Will, C. and Lührmann, R. 2007. Spliceosome structure and function. In *The RNA world* (eds. R.F. Gesteland et al.), pp. 369–400. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Xiao, X., Wang, Z., Jang, M., and Burge, C.B. 2007. Coevolutionary networks of splicing *cis*-regulatory elements. *Proc. Natl. Acad. Sci.* **104:** 18583–18588.

Xie, J. and Black, D.L. 2001. A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels. *Nature* **410:** 936–939.

Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S., and Fu, X.D. 2002. Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.* **20:** 353–358.

Yeo, G. and Burge, C.B. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11:** 377–394.

Yeo, G., Hoon, S., Venkatesh, B., and Burge, C.B. 2004. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl. Acad. Sci.* **101:** 15700–15705.

Yeo, G.W., Van Nostrand, E., Holste, D., Poggio, T., and Burge, C.B. 2005. Identification and analysis of alternative splicing events

conserved in human and mouse. *Proc. Natl. Acad. Sci.* **102:** 2850–2855.

Yeo, G.W., Nostrand, E.L., and Liang, T.Y. 2007. Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet.* **3:** e85. doi: 10.1371/journal.pgen.0030085.

Zhang, X.H. and Chasin, L.A. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes & Dev.* **18:** 1241–1250.

Zheng, Z.M. 2004. Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *J. Biomed. Sci.* **11:** 278–294.

Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419:** 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.