UNDERSTANDING ALTERNATIVE SPLICING: TOWARDS A CELLULAR CODE

Arianne J. Matlin‡, Francis Clark and Christopher W. J. Smith‡*

Abstract | In violation of the 'one gene, one polypeptide' rule, alternative splicing allows individual genes to produce multiple protein isoforms — thereby playing a central part in generating complex proteomes. Alternative splicing also has a largely hidden function in quantitative gene control, by targeting RNAs for nonsense-mediated decay. Traditional gene-by-gene investigations of alternative splicing mechanisms are now being complemented by global approaches. These promise to reveal details of the nature and operation of cellular codes that are constituted by combinations of regulatory elements in pre-mRNA substrates and by cellular complements of splicing regulators, which together determine regulated splicing pathways.

ENHANCER
A *cis*-acting RNA sequence in an exon (ESE) or intron (ISE) on which a complex, often containing SR proteins, assembles to promote the use of a weak or regulated splice site.

SILENCER
A *cis*-acting RNA sequence in an exon (ESS) or intron (ISS) on which a complex, often containing heterogeneous nuclear ribonucleoproteins, assembles to repress the use of a splice site.

**Advanced Computational Modelling Centre, and ARC Centre for Bioinformatics, Priestly Building, University of Queensland, 4072, Australia.*
‡Department of Biochemistry, 80 Tennis Court Road, University of Cambridge, CB2 1GA, UK.
Correspondence to C.W.J.S.
e-mail: cwjs1@cam.ac.uk
doi:10.1038/nrm1645

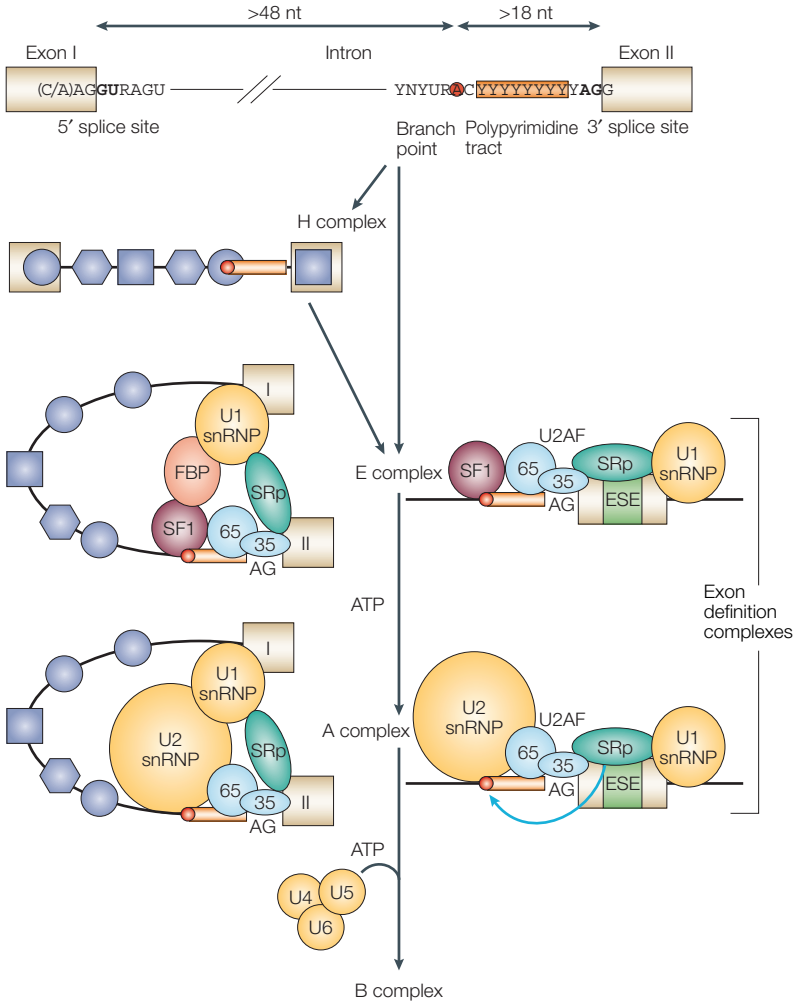
Pre-mRNA splicing is necessitated by the split nature of eukaryotic genes, in which the exons that will make up the mRNA product are interrupted by non-coding introns in the DNA and in the initial pre-mRNA transcript. Intron removal, and the concomitant joining of exons, is orchestrated by the spliceosome — a macromolecular ribonucleoprotein complex that assembles on the pre-mRNA in a series of complexes (E, A, B and C)[1,2] (BOX 1). Complex assembly is guided by consensus sequences at the ends of the introns. The importance of accurate splicing is illustrated by the fact that at least 15%, and perhaps as many as 50%, of human genetic diseases arise from mutations either in consensus splice site sequences or in the more variable auxiliary elements known as exon and intron splicing ENHANCERS (ESEs and ISEs, respectively) and SILENCERS (ESSs and ISSs, respectively)[3–6]. These auxiliary elements are involved in defining both constitutive and alternative exons (FIG. 1A).

Alternative splicing, in which different combinations of splice sites can be joined to each other (FIG. 1B), has assumed a high profile recently, owing to the dual realization that there are fewer human genes than originally anticipated, and that alternative splicing is more the rule than the exception. Analyses of expressed sequence tag (EST) and cDNA datasets conservatively estimated that ~40–60% of human genes are alternatively spliced (reviewed in REF. 7), and this number increased to 73% when alternative splicing microarray data were combined with ESTs[8]. This has been independently corroborated by 'GENOME TILING' MICROARRAYS across chromosomes 21 and 22, which indicated that alternative splicing occurs in >80% of genes[9]. Global comparisons of human and mouse alternative splicing reveal both conserved and species-specific events, with the balance between the two classes depending on the methodology used[10–13]. Conserved alternative exons are often flanked by more conserved intronic sequences than constitutive exons, and this has been used as an independent predictor of alternative splicing[14–16].

Most alternative splicing events affect the coding sequence, with half of these altering the reading frame[17] and a third apparently leading to NONSENSE-MEDIATED DECAY (NMD) of the RNA product[18]. Alternative splicing of untranslated regions can also have important regulatory consequences, including NMD, even though the open reading frame is unchanged. The *Drosophila melanogaster Dscam* gene exemplifies the extreme structural diversity that is achievable by alternative splicing. Dscam is a cell surface protein that is involved in axon

Box 1 | **Spliceosome assembly**

The positions and sequences of the consensus *cis*-acting elements help to define the splice sites of a typical metazoan intron. The removal of introns is catalysed by the spliceosome, an assembly of five small nuclear ribonucleoprotein (snRNP) particles (U1, U2, U4, U5 and U6) that are associated with a large number of additional proteins[2]. The spliceosome assembles onto the pre-mRNA through a series of complexes (see figure). Pre-mRNAs might become incorporated co-transcriptionally into an H complex together with heterogeneous nuclear ribonucleoproteins (hnRNPs); the precise complement of proteins might antagonize or promote the subsequent interaction of spliceosomal components with alternative splice sites. In the E complex, all consensus elements — and exon splicing enhancers (ESEs) if present — are recognized; U1 snRNP interacts with the 5′ splice site, SF1 (splicing factor 1) binds the branch point, U2AF (U2 auxiliary factor), a dimer of 65 and 35 kDa subunits, binds the polypyrimidine tract and the 3′ splice site, and SR proteins (SRp) bind to ESEs and contact U2AF, U1 snRNP and the branch point (light-blue arrow)[37]. The E complex is committed to the splicing pathway. However, kinetic trap assays indicate that pairing between splice sites is not committed before formation of the A complex[139]. Various lines of evidence indicate that many exons are first recognized as discrete units in a process known as EXON DEFINITION[140], so the E and A complexes are shown both as intron-based (left) and exon-definition complexes (right). Exon-definition complexes must be converted to intron-bridging interactions before spliceosome assembly can proceed further. In contrast to the kinetic trap assays, *trans* splicing studies indicate that coupling of 5′ and 3′ splice sites can accompany B complex formation[141]. So E, A and B complex formation are all potential points for alternative splicing regulation. Catalysis of the splicing reaction takes place in the C complex (not shown) following RNA conformational rearrangements. FBP, formin-binding protein; nt, nucleotide.



guidance in the developing brain, and exists as up to 38,016 alternatively spliced isoforms[19]. Pairwise interactions between Dscam variants show a preference for homophilic interactions only between identical isoforms. This ability to distinguish self from non-self could be crucial in correctly establishing neuronal connections[20]. At the other end of the complexity spectrum is the alternative splicing of the *Sex-lethal* (*Sxl*) and *transformer* (*tra*) genes in *D. melanogaster*. In each gene, sex-specific splicing gives rise to a functional protein product in females, but in males alternative splicing leads to the inclusion of stop codons, so that no functional protein is produced (reviewed in REF. 21).

How can we hope to understand alternative splicing regulation that is specific to a cell type, developmental stage or signalling pathway? The 'traditional' approach uses individual model systems — usually simple ones like *Sxl* and *tra* rather than the extreme cases such as *Dscam* — and characterizes the *cis*-acting regulatory elements and *trans*-acting factors that together impose regulation. Model systems from *D. melanogaster*, mammals and viruses have revealed various regulatory strategies. Early expectations of cell-specific regulatory proteins, the presence of which imposes cell-specific splicing patterns, have been met in some cases such as SXL and TRA in *D. melanogaster*[21] or the brain-specific Nova proteins in mammals[22]. More commonly, regulation involves the participation of multiple positively and negatively acting factors, with the precise combinations and relative ratios of factors dictating regulatory decisions. Tissue-specific splicing decisions might be determined by 'cellular codes' that are constituted by particular combinations of regulators specific to each cell type[23,24]. However, the details of these codes have remained elusive.

Traditional approaches are now being complemented by more global, 'systems' approaches for identifying regulatory elements and factors, for simultaneously detecting large numbers of alternative splicing events, and for detecting RNA targets of splicing regulators. These newer approaches will help in addressing the basis of the suggested cellular codes in a way that is not currently possible using model systems. Nevertheless, a detailed understanding of regulatory mechanisms and a full understanding of the operation of splicing codes will still require the analysis of model systems. We first discuss the contributions of model systems and then turn to the new perspective that is being afforded by the application of global approaches. Readers are referred to other reviews for more a in-depth discussion of the regulation of splicing by signalling[25,26], and the evolution of alternative splicing[27].

Basic principles

Like other levels of gene control, the regulation of splicing involves both *cis* and *trans* components, which are composed of sequences in the pre-mRNA and cellular factors (RNA or protein), respectively. It is often not possible to make a clear distinction between 'constitutive' and 'alternative' splicing elements and factors. The relative functional strengths of splice sites influence

'GENOME TILING' MICROARRAY
A type of microarray in which overlapping oligonucleotides are designed to cover a genomic region of interest.

the frequency with which an exon is selected. However, a pair of 'strong' splice sites is not sufficient to define an exon; many 'pseudo-exons' that are flanked by predicted splice sites are not spliced. Recent global analyses have indicated that the relative enrichment in ESEs or ESSs helps to distinguish between authentic exons and pseudo-exons[28,29]. The auxiliary splicing elements are highly variable in sequence, but they are important in defining constitutive and alternative exons.

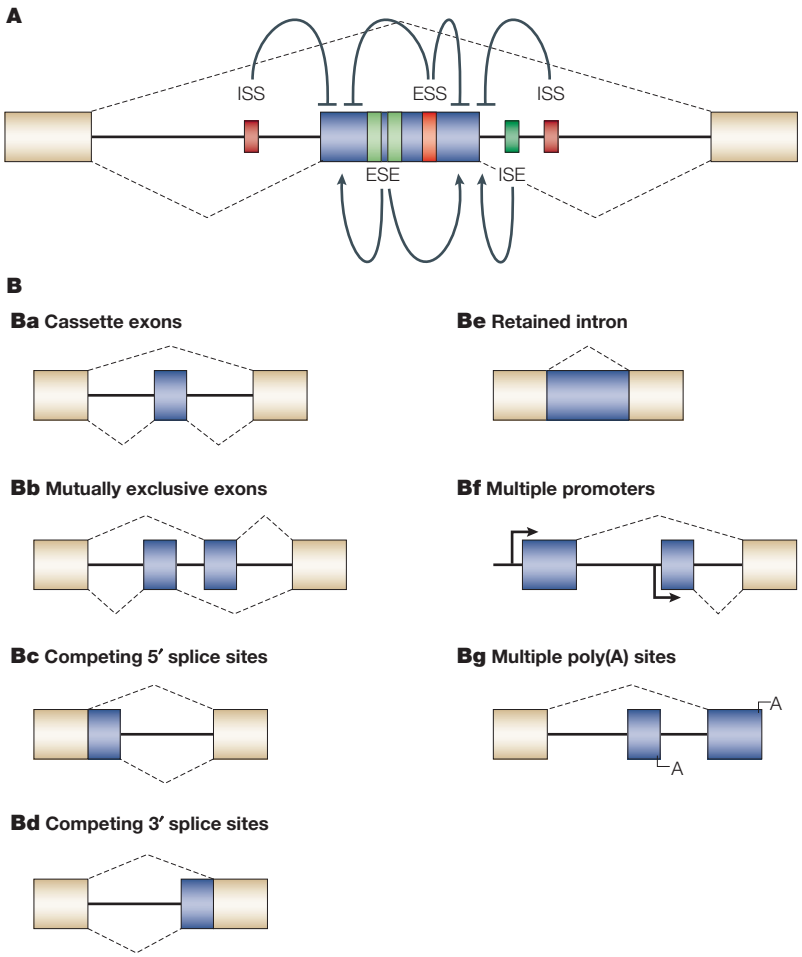


Figure 1 | **Elementary alternative splicing events and regulatory elements. A** | In addition to the splice-site consensus sequences, a number of auxiliary elements can influence alternative splicing. These are categorized by their location and activity as exon splicing enhancers and silencers (ESEs and ESSs) and intron splicing enhancers and silencers (ISEs and ISSs). Enhancers can activate adjacent splice sites or antagonize silencers, whereas silencers can repress splice sites or enhancers. Exon inclusion or skipping is determined by the balance of these competing influences, which in turn might be determined by relative concentrations of the cognate RNA-binding activator and repressor proteins. **B** | Elementary alternative splicing events represent binary choices. **Ba** | Cassette exons are discrete exons that can be independently included or excluded from the mRNA. They can be further subdivided into 'skipped' or 'cryptic' exons according to whether the main observed variant includes or excludes the exon, respectively. **Bb** | Mutually exclusive splicing involves the selection of only one from an array of two or more exon variants. **Bc,d** | Competing 5′ and 3′ splice sites represent 'exon modification' events, and in the case of 3′ splice sites can be as small as 3-nucleotide additions at NAGNAG 3′ splice sites[144]. Global computational analyses that consider only these first four categories find that 60% involve cassette and mutually exclusive exons, whereas 40% are exon modifications[17]. The remaining categories include **Be** | retained introns and **Bf** | alternative splicing in conjunction with the use of alternative promoters or **Bg** | 3′-end processing sites. Note that regulation of the last two categories need not be at the level of splicing.

Many ESEs contain binding sites for members of the SR FAMILY OF PROTEINS. SR proteins have roles in several steps of spliceosome assembly, and function as both essential splicing factors and regulatory factors[30] (BOX 1). Although they show redundancy in their essential splicing roles, they can show distinct alternative splicing activities, and the lethal effects of individual SR protein knockouts have established that they have non-overlapping roles *in vivo*[31–35]. They have a modular structure with one or two N-terminal RNA RECOGNITION MOTIF (RRM)-type domains that bind RNA, and C-terminal domains that are enriched in arginine and serine residues (RS domains). RS domains are also found in other core splicing factors such as U2AF65 (U2 auxiliary factor 65 kDa) and U2AF35, as well as alternative splicing regulators such as *D. melanogaster* TRA. They mediate both protein–protein and protein–RNA contacts[36,37].

Splicing silencers (ESSs and ISSs) are variable in sequence, but some of them bind members of the extended family of HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEINS (hnRNPs). These diverse proteins contain RRM-type and KH-type RNA-binding domains, as well as auxiliary domains that are often involved in protein–protein interactions, and they have multiple roles in pre-mRNA and mRNA metabolism[38,39]. A number of hnRNPs function as splicing repressors, including hnRNPA1, polypyrimidine tract binding protein (PTB/hnRNPI) and *D. melanogaster* SXL. Splicing activators and repressors commonly function by influencing the formation of the E and A complexes early in spliceosome assembly (BOX 1). In this regard, it is worth noting that the H ('heterogeneous') complex has a regulatory function due to its sequence-specific complements of hnRNPs[40], which can influence the ability of particular RNAs to assemble productive splicing complexes. Other *cis*-acting influences include the secondary structure of the RNA, which can either sequester sequence elements or bring them into closer apposition (reviewed in REF. 41), and the relative spacing of sequence elements. For example, the close proximity of sequence splice site elements can enforce mutually exclusive splicing events[42].

A final contextual point is that alternative splicing decisions are not made on a static, pre-synthesized template. As the pre-mRNA emerges from RNA polymerase (pol), it immediately becomes packaged with various processing and general RNA binding factors. The C-terminal domain of the RNA pol II large subunit helps to recruit RNA processing factors to the emerging transcript[43,44]. Moreover, the rate of polymerase elongation can influence splicing patterns by accelerating or delaying the synthesis of competing splice sites or regulatory elements (reviewed in REFS 4,43). It is in this dynamic setting that alternative splicing decisions are made.

Mechanisms of activation

Much work has been carried out with ESEs that activate 3′ or 5′ splice sites by binding SR proteins. A simple model for ESE-dependent 3′ splice site activation

by SR proteins involves the recruitment of U2AF65 to weak polypyrimidine tracts (FIG. 2a). An example of this activity is the female-specific ESE in exon 4 of *D. melanogaster dsx*. This complex ESE comprises a purine-rich element and six repeats of a 13-nucleotide *dsx* repeat element (dsxRE)[45]. Each element cooperatively binds a conventional SR protein along with two RS-domain-containing regulators, TRA and TRA2 (REF. 46). The separation of ~300 nucleotides between the ESE and the regulated 3′ splice site is crucial for sex-specific regulation, as the experimental reduction of this distance results in TRA-independent activation by TRA2 and SR proteins[47].

TRA functions as the crucial female-specific regulator by promoting the cooperative stabilization of the enhancer complex, which allows it to function from a distance. The enhancer has been shown *in vitro* to function through the interaction of the recruited RS domains with U2AF35, thereby promoting U2AF65 binding to the weak polypyrimidine tract[48]. Nucleotide substitutions that strengthen the 3′ splice site bypass the ESE requirement. Despite the evidence in its favour, this model is challenged by the fact that the RS domain of U2AF35 is not required for regulated splicing of *dsx* exon 4 *in vivo*[49]. In addition, other ESEs have been shown to activate splicing without affecting U2AF65 binding[50]. An alternative mechanism for enhancer activity involves the bridging interactions between ESEs and spliceosomal components, which are mediated by the SRm160 and SRm300 splicing co-activators[51–53] (FIG. 2a). These co-activators contain RS domains but lack RRM domains, and can form multiple interactions with SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES (snRNPs) and enhancer-bound SR proteins. ESEs can also activate 5′ splice sites. Female-specific selection of an alternative 5′ splice site in the *D. melanogaster fruitless* gene is dependent on an upstream ESE with 3 copies of the 13-nucleotide dsxRE motif that cooperatively binds TRA and TRA2 proteins[54]. ARTIFICIAL RECRUITMENT experiments *in vitro* indicate that the RS-domain interactions that are involved in 5′ and 3′ splice-site activation by TRA and TRA2 are similar[55].

The preceding models for ESE function are based on the ability of RS domains to mediate protein–protein interactions[56]. Recent evidence indicates a series of sequential RS-domain–RNA contacts at the BRANCH POINT and the 5′ splice site during splicing complex assembly[36,37]. In particular, RS domains that are recruited artificially or naturally to an ESE specifically contact the branch point RNA in the A complex[36,37]. The strong correlation between the ability to activate splicing and to contact the branch point, coupled with the detection of the interaction in the A complex, has led to the suggestion that the RS-domain–branch-point interaction might be responsible for splicing activation by ESEs. However, contacts of RS domains with protein and with RNA are not mutually exclusive, and their relative contribution to ESE function remains to be resolved (discussed in REF. 57).

Although more commonly viewed as repressors, individual hnRNPs can also stimulate the use of alternative splice sites, mainly through ISEs. Activation of a weak 5′ splice site in the *eNOS* (endothelial nitric oxide synthase) transcript is effected by the binding of hnRNPL to a CA-repeat ISE[58]. Similarly, the closely related TIA1 and TIAR proteins function at uracil-rich ISEs to promote the interaction of U1 snRNP with weak 5′ splice sites, apparently by direct contacts with the U1 snRNP-specific C protein[59–63] (FIG. 2b).

Mechanisms of repression

Repression of splicing often involves hnRNPs such as SXL, PTB and hnRNPA1. In the simplest scenario, silencer-bound proteins directly antagonize splicing factor binding. However, repression frequently requires cooperative repressor binding to multiple silencer elements of a common type. Two models for cooperative repression involve the creation of a 'zone of silencing', either by the propagation of repressor binding from high-affinity silencers or by looping out of RNA between bound repressors. In each case, exons, splice sites or enhancers are made inaccessible to their cognate splicing factors[3,64].

Alternative splicing of *tra* pre-mRNA in *D. melanogaster* is an example of simple steric inhibition of early spliceosome assembly. In female flies, SXL protein binds to the pyrimidine tract of an upstream 3′ splice site, thereby blocking access to U2AF. A downstream 3′ splice site, which binds U2AF but not SXL, is then selected[65] (FIG. 2c). Splicing repression by mammalian PTB can also occur by competing with U2AF65 binding[66,67]. However, repression usually requires cooperative PTB binding to multiple sites, which are often not associated with the polypyrimidine tract, so repression can often be more complex than simple binding competition with U2AF65 (REFS 64,68–71). Autoregulation of *Sxl* splicing occurs by the cooperative binding of SXL protein to multiple sites around exon 3, leading to exon skipping[72]. Remarkably, *in vitro* experiments indicate that rather than inhibiting early spliceosome assembly, repression by SXL occurs after the first catalytic step of splicing[73].

hnRNPA1 functions as a general sequence-independent regulator of alternative splicing by antagonizing SR proteins such as SF2/ASF[74,75], and variations in the ratios of SF2/ASF and hnRNPA1 can influence many alternative splicing events[76]. Variations in the relative activities can result from a change in the subcellular localization of these proteins. Phosphorylation of hnRNPA1 as a result of the stress-mediated $MKK_{3/6}$–p38 signalling cascade results in its export to the cytoplasm, with consequent changes in alternative splicing regulation in the nucleus[77]. High-affinity hnRNPA1 binding sites function as specific silencers[78–81]. The artificial recruitment of just the C-terminal glycine-rich domain of hnRNPA1, which mediates hnRNPA1 oligomerization, can restore ESS function[80]. The mechanism of repression depends on the relative location of binding sites and can include antagonism of binding of U1 and U2 snRNPs and

NONSENSE-MEDIATED DECAY (NMD). A surveillance pathway leading to the targeted degradation of mRNAs that contain premature termination codons. In mammals, termination codons that lie more than 50–55 nucleotides upstream of an exon–exon junction are recognized as premature.

EXON DEFINITION
The initial interaction and stabilization across an exon of factors bound to the 3′ and 5′ splice sites prior to the formation of cross-intron bridging interactions.

SR FAMILY OF PROTEINS
Required for both constitutive and regulated splicing, with a modular structure comprising one or two N-terminal RRMs and a C-terminal arginine/serine-rich (RS) domain.

RNA RECOGNITION MOTIF (RMM). A protein domain that is frequently involved in sequence-specific single-stranded RNA binding, and consists of a βαββαβ fold with the β strands forming a surface that displays two highly conserved RNP motifs and forms contacts with the RNA.

HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN (hnRNP). A group of >20 proteins that associate with high-molecular-weight nuclear RNA. Some hnRNPs, such as members of the hnRNPA/B family, shuttle in and out of the nucleus, whereas others are strictly nuclear.

KH
An hnRNPK homology domain that binds to single-stranded RNA, and contains α and β secondary structure elements.

SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLE (snRNP). snRNPs U1, U2, U4, U5 and U6, which contain proteins and U-rich RNA, are core components of the spliceosome.

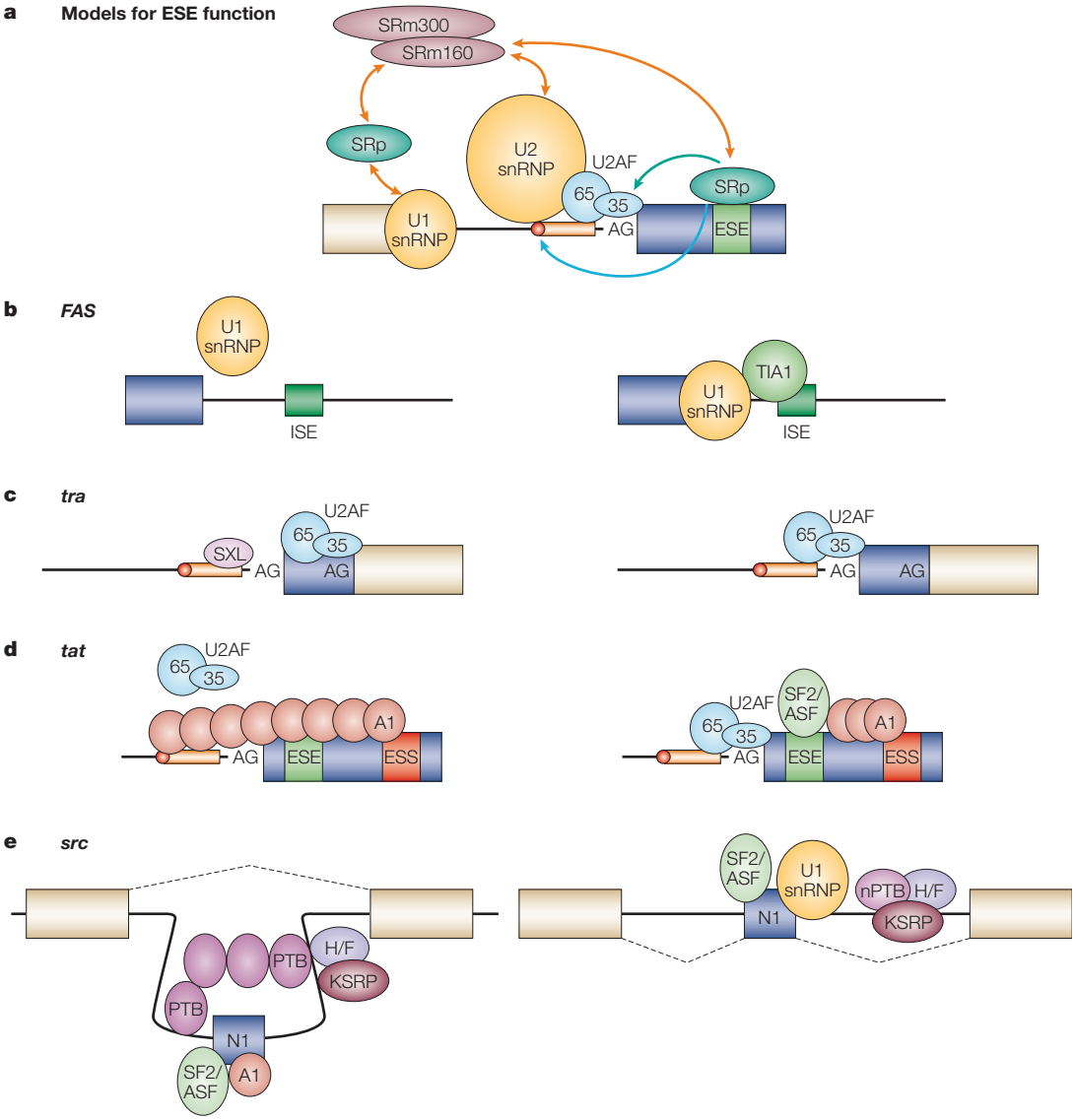


Figure 2 | **Mechanisms of enhancement and silencing of alternative splicing. a** | Models for exon splicing enhancer (ESE) function. SR proteins (SRp) bound to an ESE could activate splicing by recruiting U2AF65 to weak polypyrimidine tracts through interaction of their RS domains with U2AF35 (green arrow), by interacting with the co-activator SRm160 (orange arrow) or by contacting the RNA branch point (light-blue arrow). **b** | A weak 5′ splice site in the *FAS* transcript is enhanced by TIA1 binding to a downstream intron splicing enhancer (ISE). TIA1 cooperatively promotes the interaction of U1 small nuclear ribonucleoprotein particles (snRNPs) with the pre-mRNA. **c** | Repression of the non-sex-specific *tra* 3′ splice site involves the interaction of SXL with an intron splicing silencer (ISS) embedded in the polypyrimidine tract and the prevention of U2AF binding. This leads to selection of the downstream female-specific 3′ splice site. **d** | Inclusion of exon 3 of HIV1 *tat* pre-mRNA is determined by the nuclear ratio of specific heterogeneous nuclear ribonucleoprotein (hnRNP) and SR proteins. Propagative multimerization of hnRNPA1 from a high-affinity exon splicing silencer (ESS) is sterically blocked by the interaction of SF2/ASF with the upstream ESE. In this case, ESE function requires the RRM domains but not the RS domain of SF2/ASF. **e** | The regulation of N1 exon splicing in the *src* transcript provides an example of combinatorial control by cooperation and antagonism between numerous positively and negatively acting factors. In non-neuronal cells (left), N1 is excluded, whereas in neurons (right), it is included in the mature mRNA. Constitutive exons are shown as beige boxes, whereas alternative exons are shown as blue boxes. KSRP, KH-type splicing regulatory protein; nPTB, neural polypyrimidine tract binding protein.

ARTIFICIAL RECRUITMENT
An experimental approach that allows the analysis of 'effector' domains (repressors or activators), independent of the ability to bind RNA, by fusing an RNA-binding domain (typically, the bacteriophage MS2 coat protein) to the protein or domain of interest, and by introducing cognate RNA-binding sites into the reporter RNA.

SR proteins. In the case of hnRNPA1-induced alternative splicing events in the *hnRNPA1* gene, the skipped exon is looped out between high-affinity sites, but the important effect might be to bring the external exons into proximity rather than to directly inhibit the skipped exon[78,82]. This is reminiscent of the effects caused by the secondary structure of RNA, which can similarly influence splice site selection by bringing widely separated *cis*-acting elements into juxtaposition[83].

Although mainly noted for their roles as activators, RS-domain-containing factors can also function as repressors. The adenovirus L1 transcription unit provides an example of splicing repression mediated by the binding of SR proteins to an ESE-like sequence that, owing to its position upstream of the branch point, functions as a silencer[84]. Artificial recruitment shows that the second RRM domain of SF2/ASF mediates repressor activity[85]. Binding of SF2/ASF to this ISS inhibits A-complex assembly, but SF2/ASF dephosphorylation by protein phosphatase 2A, which is recruited by viral protein E4 open reading frame 4 (ORF4) during late phases of infection, destabilizes the repressor complex and allows splicing to the regulated splice site[86]. This example also illustrates the general principle that reversible phosphorylation of SR proteins on serine residues controls their activity and localization (reviewed in REFS 25,26).

Combinatorial control

In contrast to the examples mentioned above, which mostly involve the promotion or inhibition of early spliceosome assembly by individual regulators, many alternative splicing events involve a more complex interplay between positive and negative regulators that function through cognate enhancers and silencers. For example, an ESS in HIV1 *tat* exon 3 binds hnRNPA1 with high affinity and inhibits splicing by propagating further hnRNPA1 binding towards the 3′ splice site (FIG. 2d). This propagative binding can be inhibited, and splicing activated, by the binding of SF2/ASF to an upstream ESE. In contrast to other ESEs, in which artificial recruitment of just an RS domain can restore function, at this ESE it is only the RRM domains of the SR protein that are required[79]. A purine-rich ESE in the M2 exon of the *IgM* gene functions in a similar way, by antagonizing an adjacent PTB-binding ESS[87]. The ESS itself promotes the formation of a 'dead-end' A-like complex, rather than blocking complex formation completely[50]. The formation of dead-end, splicing-related complexes as a mechanism of repression has also been observed in other systems. Retention of intron 3 of the *D. melanogaster* P-element transposase is somatic cell specific. The KH-domain protein P-element somatic inhibitor (PSI) contacts U1 70K protein and diverts U1 snRNP binding away from the authentic 5′ splice site to a pair of pseudo-5′ splice sites[88]. Nonproductive interaction of the 5′ splice site of caspase-2 exon 9 with a 'decoy' 3′ splice site leads to exon skipping[89]. In these cases, sequences that resemble splice sites function as silencer elements.

Antagonism between members of the SR protein family can also mediate alternative splicing decisions. For example, the regulation by SRp20 of its own splicing is antagonized by SF2/ASF[90]. Several SR-related factors have also been identified recently that can modulate the splicing regulatory activity of canonical SR proteins. The *D. melanogaster* RSF1 repressor antagonizes the ability of SF2/ASF to promote the binding of U1 to pre-mRNA. RSF1 comprises an N-terminal RRM, which binds purine-rich ESE-like sequences and competes directly with SR proteins, and a C-terminal Gly/Arg/Ser-rich region that interacts with the RS domain of SF2/ASF[91,92]. The mammalian SR-related protein SRrp86 exerts differential control over the splicing regulatory effects of specific canonical SR proteins. It enhances SRp20 activity but represses other family members including SC35, SF2/ASF and SRp55 (REF. 93), and is itself antagonized by the SR protein 9G8, hnRNPG and SAFB (scaffold attachment factor B)[94]. SRrp86 shows tissue-restricted expression at the translational level, whereas its targets are present ubiquitously; this results in a unique ratio of factors in each tissue type[94].

Functional antagonism between different hnRNPs is another emerging theme in the regulation of alternative splicing. The opposing effects of repressive PTB and activating CELF (CUG-BP and ETR3-like factors) proteins[95] modulate the splicing of several target pre-mRNAs[96–99]. In the splicing of *N*-methyl-D-aspartate (NMDA) receptor subunit NR1 and α-actinin pre-mRNAs, direct competition between binding of PTB and CELFs to adjacent sites is involved[97,99]. In the case of cardiac troponin-T, functional antagonism occurs without binding competition[96]. In addition to their anti-repressor activity, CELF proteins also function as repressors of other splicing events, sometimes elsewhere in the same transcript, by binding to CUG- or UG-repeat elements[97–99].

Neuronal-specific alternative splicing of the mammalian *src* N1 exon is a well-studied model system with many regulatory factors, none of which alone seems to be sufficient to explain the tissue specificity of regulation (reviewed in REF. 100) (FIG. 2e). Inhibition of N1 splicing in non-neuronal cells depends on four flanking CUCUCU elements. These bind PTB cooperatively, as indicated by reduced PTB binding at the upstream sites after mutation of the downstream elements and vice versa[70]. U2 snRNP binding to the upstream intron correlates inversely with PTB binding. By contrast, U1 snRNP binding to the N1 exon seems to be unaffected by PTB-mediated repression, which argues for cooperative binding with looping of intervening RNA, rather than propagative PTB binding, which would be expected to displace U1 snRNP. N1 inclusion in neuronal cells depends on a downstream ISE, which overlaps with one of the CUCUCU motifs and interacts with hnRNPF and hnRNPH, and KH-type splicing regulatory protein (KSRP). Assembly of the ISE complex is promoted by nPTB[101], a neuronally restricted PARALOGUE of PTB, which binds the CUCUCU elements but is less repressive than PTB. nPTB seems to have a dual role in antagonizing a repressor as well as promoting assembly of an ISE complex. Despite its restricted expression, nPTB is not sufficient to activate N1 splicing. An ESE in the N1 exon binds SF2/ASF and hnRNPF and hnRNPH, whereas hnRNPA1 binding mediates repression by a mechanism that is distinct from that of PTB[102]. The mechanistic details of N1 splicing activation by the ISE and ESE are not yet well understood; however, this system illustrates the complexity of regulation of a seemingly simple system.

BRANCH POINT
An adenosine within a variably conserved 'branch point sequence' upstream of the 3′ splice site that is involved in the first chemical step of splicing. A new 2′–5′ phosphodiester is formed between the branch point and the 5′ end of the intron, concomitant with breakage of the link between the 5′ exon and the intron.

PARALOGUE
A sequence, or gene, that originates from a common ancestral sequence, or gene, by a duplication event.

In some cases, the presence or absence of an individual mammalian tissue-specific regulator is sufficient to determine splicing patterns. The brain-specific KH-domain-containing Nova proteins are involved in determining a number of tissue-specific alternative splicing events through binding to ISEs, ISSs and ESSs that contain UCAY motifs[22,103]. Nova1 promotes alternative exon inclusion in glycine receptor α2, and functions antagonistically with nPTB[104]. However, most tissue-specific alternative splicing events studied so far seem to be regulated by a more complex network of synergistic and antagonistic influences between groups of regulators. In general, the individual *cis* elements that have been identified are not sufficient to direct tissue-specific splicing pathways (extensive lists of experimentally determined *cis* elements are provided in REFS 105,106).

Global identification of regulatory motifs

Global approaches that permit genome-wide identification of regulators and target sequences are beginning to contribute to our understanding of alternative splicing. An initial aim of experimental and computational screens is to provide a catalogue of regulatory sequence motifs, and substantial progress has been made on this front.

In vitro *approaches.* *In vitro* binding SELEX (systematic evolution of ligands by exponential enrichment) experiments have been carried out with a number of SR proteins and hnRNPs. Optimal binding sites vary between individual SR proteins, but they often function as splicing enhancers (reviewed in REFS 3,107) and frequently resemble 'classical' purine-rich ESEs. Similarly, optimal binding sites for repressor hnRNPs, such as hnRNPA1 (REF. 108), PTB/hnRNPI (REF. 109) and SXL[67], correspond to known splicing silencers. The SELEX-selected sequence for FOX1 protein[110] (GCAUG) corresponds to a regulatory sequence motif that had previously been identified in a number of pre-mRNAs and independently by computational analysis of brain-specific exons[111].

In vitro functional SELEX experiments have revealed various classes of ESE, some of which could be associated with particular SR proteins[112–115]. Splicing that is dependent on the addition of individual SR proteins to cytoplasmic extract allowed the selection of distinct motifs for SF2/ASF, SC35, SRp40 and SRp55 (REFS 112,113). These functionally defined motifs, which were more redundant than *in vitro* binding, SELEX selected sequences for the same proteins (possibly reflecting the fact that ESE function does not require the tightest binding), have led to the highly successful ESEfinder web resource (see online links box)[116].

In vivo *approaches.* *In vivo* functional selection approaches have used cell-transfection assays. The first example of this approach was based on the selection for ESEs in a mini-gene construct, and involved the replacement of a necessary ESE in an internal exon by a random 13-mer sequence. Following transfection, active ESEs were identified by PCR after reverse transcription of RNA (RT-PCR) and sequencing of the included exons[117]. This revealed both the expected purine-rich ESEs and a class of AC-rich ESEs, which were subsequently shown to be binding sites for the RNA-binding protein YB1 (REF. 118). An elaboration of this approach, known as 'fluorescence-activated screen for exonic splicing silencers', was recently reported by the Burge laboratory[28]. A three-exon mini-gene was engineered to produce enhanced green fluorescent protein (eGFP) only when exon 2 is skipped. A library of random decamers was cloned into exon 2, stably transformed cells were selected, and fluorescence-activated cell sorting (FACS) was used to select cells that produced eGFP. PCR of genomic DNA then allowed ESS sequences to be derived. Using this approach, 141 decamers with ESS activity were identified. These could be grouped into seven putative ESS hexamer motifs, some of which resembled optimal binding sites for hnRNPA1 and hnRNPH. The cell-based screens for ESEs and ESSs[28,117] could be carried out in various cell types, potentially revealing motifs that mediate cell-type-specific regulation. Importantly, the experimental design of Wang *et al.*[28] could readily be extended to ISEs and ISSs.

Computational approaches. Purely computational approaches, followed by experimental validation, have also contributed to the global definition of splicing regulatory motifs[29,119–121]. The usual approach is to consider a sequence dataset that is expected to be enriched in binding sites of interest, and to compare it with one or more control sets (for example, a search for ESE motifs in coding exons from multi-exon genes versus coding sequences from intronless genes as a control[119]). Hexamers (or other N-mers, depending on the size of the datasets) that are significantly enriched are considered as putative binding sites, and generally group into families of highly similar sequences that, together, define a binding motif. Computational approaches are necessarily focused on common binding sites for which hexamer overrepresentation can be observed.

In the RESCUE-ESE (relative enhancer and silencer classification by unanimous enrichment; see online links box) approach, hexamers were identified in constitutive human exons by enrichment in exons versus introns, and in exons with weak splice sites versus exons with strong splice sites[120]. Using stringent cut-offs, 238 distinct hexamers were identified as possible ESEs. These clustered into ten motifs, of which two were specific to 5′ splice sites, five were specific to 3′ splice sites, and three associated with both splice sites. Representatives of all ten classes were validated as ESEs in transfection assays. In addition, an average of 5.2 ESE 'clumps' (multiple overlapping ESE hexamers) were found per exon, with most exons having between 3 and 7 ESE clumps. Lower ESE densities were observed in a dataset of 2,000 alternative (skipped) exons. In a comparative analysis between human, mouse and *Fugu* species, potential ESEs and ISEs were identified[121]. In the RESCUE-ISE approach, hexamers were selected by

SELEX
(Systematic evolution of ligands by exponential enrichment). An iterative method for the selection of optimal sites for RNA-binding proteins. Starting from a pool of degenerate cDNAs, sequences are subjected to several rounds of transcription, protein binding and RT-PCR until a consensus sequence emerges.

enrichment in introns versus exons and in introns with weak splice sites versus introns with strong splice sites. Two primary ISE motifs, the known ISE triplet GGG motif and a C-rich motif, were found in human and mouse, but not in *Fugu* species.

To avoid the complicating effects of protein-coding biases, Zhang and Chasin[29] compared octamers (allowing a single mis-match) from non-coding 5′ UTR constitutive exons with both pseudo-exons and intronless 5′ UTR sequences. They identified 2,069 putative ESE octamers and 974 putative ESS octamers (~5% of all octamers), grouping into 80 putative ESEs and 69 putative ESSs. The ESE octamers had substantial overlap with the RESCUE-ESE identified motifs[120]. Of 20 potential ESEs and ESSs tested, 18 were experimentally validated. Furthermore, the motifs identified in this study and in the work by Wang *et al.*[28] could be used to distinguish between authentic exons and pseudo-exons.

In addition to these searches for general motifs that regulate splicing, some computational analyses have searched for motifs that are associated specifically with alternative exons. Analysis of a set of 25 brain-specific exons reported in the literature allowed Brudno *et al.*[111] to show that the hexanucleotide UGCAUG, a binding site for FOX1 protein[110], was overrepresented in the downstream intron, and this enrichment was conserved for the orthologous exons in other species[122]. Miriami *et al.* identified longer C- and G-rich elements in the introns that flank skipped exons[123], prompting the suggestion that these might be involved in looping out skipped exons. Finally, a set of motifs associated with exons (or splice sites) that are skipped in brain or liver was found to be enriched in known silencers[124]. The current limitation of these focused approaches for regulatory elements is the starting datasets of co-regulated alternative splicing events. Alternative splicing microarrays (BOX 2) should provide the necessary high-quality datasets for more revealing analyses in the near future.

Global identification of regulator targets

The computational and experimental approaches discussed above aim to identify a 'parts list' of sequence motifs that can influence splicing. A complementary global approach is to start with individual splicing regulators and define the sets of cellular pre-mRNAs to which the regulators bind, and of splicing events that are influenced by the regulators. So far, alterations in splicing have been used to validate targets identified in binding assays. But with the advent of alternative splicing microarrays (BOX 2), it might become possible to directly identify alterations in alternative splicing events.

Identifying RNA-binding targets. Global assays for cellular RNAs that are bound to specific proteins mostly involve the immunoprecipitation of RNAs associated with that protein, followed by microarray analysis to identify the mRNAs[125]. In its simplest form, this approach is not ideal for splicing regulators. Although cDNA microarrays can detect complexes involving intron regions of pre-mRNA[126], genomic arrays would be more suitable. PRE-mRNP complexes are also likely to be present at lower levels than stable mRNPs. Moreover, RNP immunoprecipitation can suffer from artefactual re-assortment of RNA–protein complexes after cell lysis[127]. The last problem can be overcome by crosslinking complexes before cell lysis. Darnell and co-workers developed UV crosslink immunoprecipitation (CLIP)[103] (BOX 3) to identify RNA targets of the brain-specific Nova splicing regulators in mouse. The approach was remarkably successful and identified 340 RNA sequence 'tags' of ~60–100 nucleotides from mouse brain. CLIP tags associated with a functionally related set of genes that encoded components of inhibitory synapses. Most tags were intronic and were enriched in optimal Nova-binding motifs (YCAYC), as identified by SELEX[128]. In functional validation experiments using Nova$^{-/-}$ mice, 7 of 18 newly identified target genes showed large alterations in the splicing of exons that are associated with CLIP tags.

Genomic SELEX (BOX 3) is an intermediate approach between CLIP, which 'freezes' RNP complexes *in vivo* before analysis *in vitro*, and conventional SELEX, which identifies optimal binding sites from random sequence pools *in vitro*. Genomic SELEX with *D. melanogaster* SRp55 (known as B52) successfully identified candidate targets[129]. Of 15 genes that were detectably expressed in B52$^{-/-}$ larvae, 4 were validated as B52 targets by the obvious alterations in splicing of exons that were associated with putative B52 binding sites. Although it does not sample cell-specific RNA–protein interactions, genomic SELEX has the advantage of normalizing the representation of intronic and exonic binding sites.

Chromatin immunoprecipitation (ChIP) is widely used for analysis of DNA-associated proteins, but can be used to monitor the association of proteins with nascent RNA *in vivo* (reviewed in REF. 130) (BOX 3). U1 snRNP was shown to associate specifically with intron-containing genes in *Saccharomyces cerevisiae* and to associate with the expected intronic regions, but not 5′ exons or distant 3′ locations[131]. ChIP therefore promises to contribute to global analysis of splicing regulators. Although CLIP, genomic SELEX and ChIP have yet to prove their general applicability, it is likely that between them, they will be widely harnessed to catalogue pre-mRNA targets of splicing regulators in different cell types.

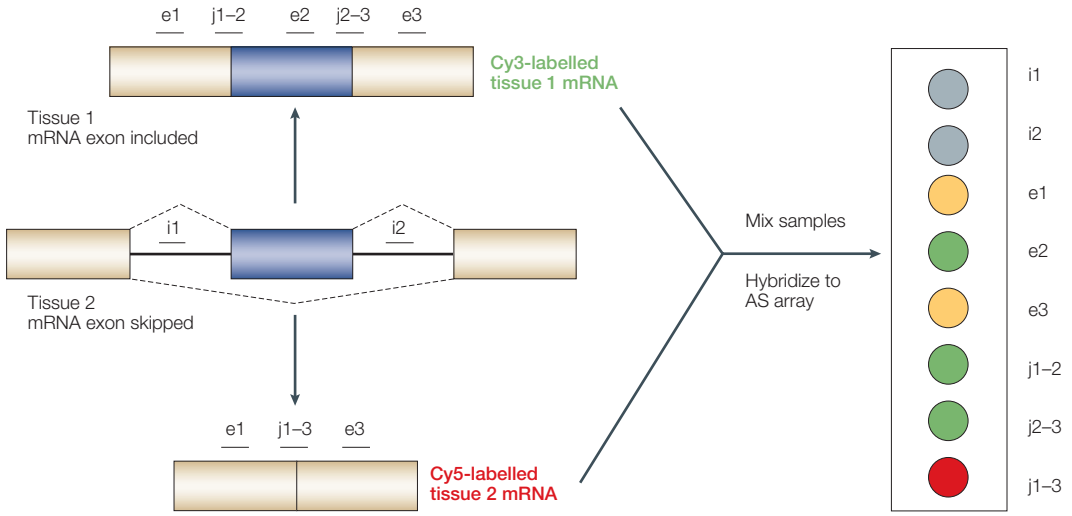
Identifying functional targets. The general approach to define functional targets of splicing regulators is to perturb the levels of a splicing regulator *in vivo* and then search for changes in splicing. Methods that have been used successfully to reduce splicing factor levels *in vivo* include RNA interference (RNAi) of various factors in cell lines, RNAi of SR proteins and hnRNPs in *C. elegans*[132], temperature-sensitive mutants of *D. melanogaster* U2AF50 (REF. 126), replacement of the endogenous SF2/ASF by an inducible version in chicken DT40 cells[33], and Nova knockout in mouse[22]. These studies predated the recent development of

PRE-mRNP (or mRNP). A ribonucleoprotein complex, of variable protein composition, that is associated with pre-mRNAs or mRNAs, respectively.

Box 2 | **Alternative splicing microarrays**

Microarrays contain multiple immobilized cDNA or oligonucleotide probes that are hybridized with labelled 'target' RNA, cDNA or cRNA. Usually, target samples from two different cell or tissue sources are labelled with Cy5 (red) and Cy3 (green) fluorescent dyes, mixed and then hybridized to the array, which allows the comparison of expression levels between the two samples. Immobilized cDNA probes cannot distinguish between alternatively spliced isoforms, but oligonucleotide probes, typically 25–60 nucleotides in length, can be designed to hybridize to isoform-specific mRNA regions. Recently, alternative splicing microarrays have been designed with probes that are specific to both exons and exon–exon junctions (see figure). Probes e1, e2 and e3 are exon specific, whereas j1–2, j2–3 and j1–3 are isoform-specific junction probes. Some arrays also contain intron probes (i1 and i2) to indicate signals from pre-mRNA.

Various array design and data processing strategies facilitate the quantitative analysis of alternative splicing patterns, some of which have been subsequently confirmed by PCR after reverse transcription of RNA (RT-PCR)[8,135,136,143]. Johnson *et al.*[8] used arrays with probes for all adjacent exon–exon junctions in 10,000 human genes and hybridized these with samples from 52 human tissues and cell lines. This revealed cell-type-specific clustering of alternative splicing events, and allowed the discovery of new alternative splicing events. Pan *et al.*[136] analysed 3,126 known cassette-type alternative splicing events in mouse using exon-specific and exon–exon junction probes. Analysis of RNAs in ten tissues showed clustering of alternative splicing events by tissue type, and further revealed that tissue-specific programmes of transcription and alternative splicing operate on different subsets of genes. A direct comparison also showed that computational prediction of tissue-specific alternative splicing based on ESTs and cDNAs performed poorly compared with the alternative splicing microarray and RT-PCR.



alternative splicing microarrays, which can facilitate the identification of targets by analysing large numbers of alternative splicing events in parallel (BOX 2). Nevertheless, global expression analysis of the effects of *D. melanogaster* U2AF50 depletion provided the surprising finding that nearly 40% of downregulated genes lacked introns, possibly indicating a novel role for U2AF50 in mRNA export[126].

Many splicing regulators have essential roles in early development, so the investigation of their roles in tissue-specific splicing requires methods to selectively affect their activity in adult tissues. Conditional Cre-mediated knockout mice have been used to address the roles of the SR proteins SC35 and SF2/ASF in adult heart[32,35], and of SC35 in thymus[133]. Strikingly, adult cardiac muscle cells were viable without SC35 or SF2/ASF, although in both cases distinct cardiac phenotypes developed during adult life. Conventional microarrays identified a large group of upregulated cardiac hypertrophy-associated genes that are present in both SC35$^{-/-}$ and SF2/ASF$^{-/-}$ mice, which clearly represented secondary effects, and a smaller number of downregulated genes. RT-PCR and western blot surveys of the main contractile proteins and proteins involved in Ca^{2+} signalling revealed three pre-mRNAs with specific alterations in alternative splicing in SF2/ASF$^{-/-}$ hearts[32]. These data give an unexpected insight into the role of SR proteins *in vivo*, namely that their functions in constitutive splicing might be redundant in many cases, with specific alternative splicing events being regulated by individual SR proteins.

These initial studies highlight future challenges. SR proteins are splicing regulators, essential splicing factors and, similar to hnRNPs, have roles at other post-transcriptional stages of gene expression in the nucleus and cytoplasm[134]. Ablation of such factors could therefore have multiple effects. The use of alternative splicing microarrays should help to identify effects at the level of splicing. It will be more complicated to distinguish between genuine patterns of alternative splicing, and mis-splicing that results from the complete lack of an essential splicing factor. Alternative

Box 3 | **Methods for identifying RNA-binding targets of regulatory proteins**

CLIP (crosslinking and immunoprecipitation)
This method can identify proteins that directly contact RNA in live cells. It involves the following steps:
- Irradiate cells or tissue with ultraviolet B to induce covalent links between RNA and bound protein
- Lyse the cells and carry out a partial RNase digestion
- Immunoprecipitate the protein of interest with crosslinked RNA
- Label the 5′ ends of the RNA fragments
- Purify the RNA–protein complexes by electrophoresis
- Remove the proteins by treatment with protease, leaving the RNA fragments
- Ligate linkers, PCR after reverse transcription of RNA (RT-PCR) and sequence the CLIP tags

CLIP allows the stringent purification of RNA–protein complexes, without re-assortment of complexes, and allows the identification of regions of pre-mRNA that bind protein. However, the method is non-quantitative, the efficiency of RNA–protein crosslinking is variable, and only proteins in direct contact with RNA are detected.

Genomic SELEX (systematic evolution of ligands by exponential enrichment)
This involves *in vitro* binding selection using RNA transcribed from genomic DNA fragments. Compared with conventional SELEX, which uses random sequence pools, the identification of real genomic sequences is an advantage, but the starting pool also contains intergenic sequences. Only sequences to which purified protein binds are identified, whereas physiological binding might involve cooperative interactions with other proteins.

ChIP (chromatin immunoprecipitation)
This method monitors the association of proteins with specific genomic regions. It involves the following steps:
- Treat live cells with formaldehyde to induce protein–nucleic-acid crosslinks
- Lyse the cells and shear the DNA
- Immunoprecipitate the protein of interest along with crosslinked DNA
- Reverse crosslinks by heat treatment and phenol extract to remove proteins
- Carry out PCR to detect gene regions that are associated with the protein

ChIP allows the quantitative detection of protein association along a gene, and can detect proteins that are not in direct contact with RNA or DNA. It can be applied globally by amplifying all immunoprecipitated DNA and hybridizing with a genomic array. However, because it monitors the association of protein with DNA, it only detects the co-transcriptional association of splicing factors. A variant of ChIP known as RNP immunoprecipitation (RIP) has been developed, in which RT-PCR is used to monitor RNA sequences to which a protein binds[142].

splicing microarrays that are designed on the basis of known alternative splicing events[135,136] would not detect mis-splicing (or previously undetected genuine events), whereas 'unbiased' exon junction alternative splicing arrays[8], or even genomic tiling arrays[9], have the potential to detect novel splicing events.

The development of transgenic models that allow quantitative alterations in splicing regulator levels, either by inducible knockdown or overexpression, might give an improved picture over simple knockouts. However, the widespread autoregulation by RNA-binding proteins (for example, see REF. 137 and references therein) might present a further challenge to this approach. 'Functional target' assays also have the associated problem of disentangling direct from indirect effects, as illustrated by the cardiac-specific SC35 and SF2/ASF knockout mice[32,35]. RNA-binding approaches can establish a direct connection, but not for regulators such as protein kinases that act on RNA-binding regulators, but do not themselves bind RNA. This highlights a further important challenge, as many alternative splicing regulators are themselves regulated by post-translational modifications[25,26] and subcellular localization.

Associating regulators with splicing patterns

Attempts are currently being made to associate tissue-specific patterns of splicing with complements of splicing regulators. Computational analyses have gone some way to providing a global view of tissue-specific splicing, but these studies are limited by inadequate transcript coverage and by difficulties in assigning ESTs to tissue types[11]. In a direct comparison with alternative splicing microarray and RT-PCR, computational predictions of tissue-specific alternative splicing performed relatively poorly[136]. Nevertheless, computational predictions of tissue-specific alternative splicing combined with published microarray data on splicing factor expression produced some interesting findings[124]. Liver was found to have the most distinctive global patterns of alternative splicing, combined with the most divergent pattern of SR and hnRNP expression.

Future progress in this area will probably come from experimental approaches that make use of the procedure outlined by Rélogio *et al.*[135] who developed a microarray with exon-junction and exon-specific probes for 100 cassette-type alternative splicing events, combined with probes for the

expression of 86 known splicing factors and regulators. Samples from Hodgkin's lymphoma cell lines showed clustering of alternative splicing patterns according to cell transformation and tumour grade. Moreover, the ectopic expression of the usually neuron-restricted Nova2 protein was seen to accompany a usually neuron-specific splicing pattern of JNK2 kinase in grade IV tumour cells. Future elaborations on this experimental design will probably include probes for many more alternative splicing events (as described in REF. 136) and, crucially, to analyse the expression of all isoforms of the splicing regulators. Large datasets of co-regulated splicing events will facilitate the characterization of the associated complements of splicing regulators and of the *cis* elements that mediate these alternative splicing programmes. But even these datasets will not be perfect; the fact that many alternative splicing events lead to products that are degraded by NMD[18,137] means that steady-state ratios of mRNA isoforms do not accurately reflect the relative activities of the pathways that lead to their formation.

An implicit assumption of the preceding discussion is that we have a full parts list of splicing regulators. Although it is possible to catalogue large numbers of known splicing factors and regulators, there are certain to be many other alternative splicing regulators that await identification. A recent RNAi screen in *D. melanogaster* cells that targeted 250 mRNAs encoding predicted RNA-binding proteins and known spliceosomal components shows how we might arrive at a fuller parts list of regulators[138]. The effects of targeting potential regulators were monitored using five alternative splicing events. Of the 47 identified alternative splicing regulators, 26 were new, whereas the remainder included known core spliceosomal proteins (for example, U1A and C proteins, U2AF50 and U2AF38) and proteins involved in other nuclear steps in gene expression.

Concluding remarks

The experimental approaches described above hold the promise of revealing details of the 'cellular codes' of splicing regulators. However, even a quantitative definition of a full set of regulator-encoding mRNAs, including all appropriate alternative isoforms of the regulators themselves, would not fully define a code. The many examples of the regulation of splicing regulators by phosphorylation and nuclear–cytoplasmic redistribution[25,26] indicate that a comprehensive definition would require proteomic analysis of nucleoplasmic RNA-binding proteins. Moreover, given the potential for alterations in transcription kinetics to regulate alternative splicing decisions[4,43], information on the state of the transcriptional machinery might ultimately need to be included to fully define the splicing code of a cell.

Despite these difficulties, there are many opportunities to advance our understanding of alternative splicing with the large body of data on the *cis* and *trans* elements that determine splicing pathways. In the future, data-mining and computer-modelling approaches can be expected to provide important insights. Alternative splicing pathways involve the formation and re-modelling of the pre-mRNP along competing pathways. With cellular catalogues of splicing factors, their binding motifs and some understanding of how these factors might cooperatively or antagonistically interact with each other, it becomes possible to start building computer models of pre-mRNP formation and consequent splicing. This could be modelled as a dynamic network of protein–RNA and protein–protein interactions in a highly simplified but biologically appropriate way. All this might require a community effort to develop the necessary datasets, to make them publicly available, and to engage and encourage modellers in this work. Data from global approaches combined with the means to examine that data and test hypotheses will provide further systematic focus on crucial issues in our mechanistic understanding of splicing regulation and its codes.

1. Burge, C., Tuschl, T. & Sharp, P. in *The RNA World* 2nd edn (eds Gestetland, R., Cech, T. & Atkins, J.) 525–560 (Cold Spring Harbor Laboratory Press, 1999).
2. Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**, 5–14 (2003).
3. Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Rev. Genet.* **3**, 285–298 (2002).
4. Caceres, J. F. & Kornblihtt, A. R. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* **18**, 186–193 (2002).
5. Faustino, N. A. & Cooper, T. A. Pre-mRNA splicing and human disease. *Genes Dev.* **17**, 419–437 (2003).
6. Pagani, F. & Baralle, F. E. Genomic variants in exons and introns: identifying the splicing spoilers. *Nature Rev. Genet.* **5**, 389–396 (2004).
7. Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nature Genet.* **30**, 13–19 (2002).
8. Johnson, J. M. *et al*. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).
Used a splice-sensitive microarray with probes for every known exon–exon junction in >10,000 human genes to reveal several global features of alternative splicing, as well as the discovery and validation of previously uncharacterized splicing events.
9. Kampa, D. *et al*. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* **14**, 331–342 (2004).
10. Thanaraj, T. A., Clark, F. & Muilu, J. Conservation of human alternative splice events in mouse. *Nucleic Acids Res.* **31**, 2544–2552 (2003).
11. Pan, Q. *et al*. Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet.* **277**, 73–77 (2005).
12. Nurtdinov, R. N., Artamonova, I. I., Mironov, A. A. & Gelfand, M. S. Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum. Mol. Genet.* **12**, 1313–1320 (2003).
13. Modrek, B. & Lee, C. J. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nature Genet.* **34**, 177–180 (2003).
14. Philipps, D. L., Park, J. W. & Graveley, B. R. A computational and experimental approach toward a priori identification of alternatively spliced exons. *RNA* **10**, 1838–1844 (2004).
15. Sorek, R. & Ast, G. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13**, 1631–1637 (2003).
16. Sorek, R. *et al*. A non-EST-based method for exon-skipping prediction. *Genome Res.* **14**, 1617–1623 (2004).
17. Clark, F. & Thanaraj, T. A. Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Hum. Mol. Genet.* **11**, 451–464 (2002).
18. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100**, 189–192 (2003).
A bioinformatics survey indicating that a large proportion of human alternative splicing leads to products that are degraded by nonsense-mediated decay.
19. Schmucker, D. & Flanagan, J. G. Generation of recognition diversity in the nervous system. *Neuron* **44**, 219–222 (2004).
20. Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L. & Clemens, J. C. Alternative splicing of *Drosophila Dscam* generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* **118**, 619–633 (2004).
21. Lopez, A. J. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* **32**, 279–305 (1998).
22. Jensen, K. B. *et al*. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**, 359–371 (2000).
The mouse knockout revealed Nova1 to be a mammalian tissue-specific splicing regulator.
23. Black, D. L. Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* **103**, 367–370 (2000).
24. Smith, C. W. & Valcarcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).
25. Shin, C. & Manley, J. L. Cell signalling and the control of pre-mRNA splicing. *Nature Rev. Mol. Cell Biol.* **5**, 727–738 (2004).

26. Stamm, S. Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum. Mol. Genet.* **11**, 2409–2416 (2002).
27. Ast, G. How did alternative splicing evolve? *Nature Rev. Genet.* **5**, 773–782 (2004).
28. Wang, Z. *et al*. Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831–845 (2004).
A transfection-based screen for ESSs, which can potentially be used to identify different classes of regulatory element that are active in different cell types.
29. Zhang, X. H. & Chasin, L. A. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* **18**, 1241–1250 (2004).
A computational screen for ESSs and ESEs; the datasets produced by reference 28 and 29 allow pseudo-exons to be discriminated from authentic exons.
30. Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA* **6**, 1197–1211 (2000).
31. Ring, H. Z. & Lis, J. T. The SR protein B52/SRp55 is essential for *Drosophila* development. *Mol. Cell. Biol.* **14**, 7499–7506 (1994).
32. Xu, X. *et al*. ASF/SF2-regulated CaMKIId alternative splicing temporally reprograms excitation–contraction coupling in cardiac muscle. *Cell* **120**, 59–72 (2005).
Conditional Cre-mediated knockout of SF2/ASF in cardiac muscle produces a surprisingly specific phenotype with disruption of a few specific alternative splicing events.
33. Wang, J., Takagaki, Y. & Manley, J. L. Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes Dev.* **10**, 2588–2599 (1996).
34. Jumaa, H., Wei, G. & Nielsen, P. J. Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. *Curr. Biol.* **9**, 899–902 (1999).
35. Ding, J. H. *et al*. Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J.* **23**, 885–896 (2004).
36. Shen, H. & Green, M. R. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol. Cell* **16**, 363–373 (2004).
37. Shen, H., Kan, J. L. & Green, M. R. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol. Cell* **13**, 367–376 (2004).
References 36 and 37 show specific interactions of RS domains with the branch point and 5′ splice site sequences within splicing complexes, which indicates that the key functional role of RS domains might be to bind RNA.
38. Krecic, A. M. & Swanson, M. S. hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* **11**, 363–371 (1999).
39. Dreyfuss, G., Kim, V. K. & Kataoka, N. Messenger RNA-binding proteins and the messages they carry. *Nature Rev. Mol. Cell Biol.* **3**, 195–205 (2002).
40. Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G. & Reed, R. Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly *in vitro*. *Mol. Cell. Biol.* **12**, 3165–3175 (1992).
41. Buratti, E. & Baralle, F. E. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.* **24**, 10505–10514 (2004).
42. Smith, C. W. J. & Nadal Ginard, B. Mutually exclusive splicing of α-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. *Cell* **56**, 749–758 (1989).
43. Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J. & Nogues, G. Multiple links between transcription and splicing. *RNA* **10**, 1489–1498 (2004).
44. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506 (2002).
45. Lynch, K. W. & Maniatis, T. Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev.* **9**, 284–293 (1995).
46. Lynch, K. W. & Maniatis, T. Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila* doublesex splicing enhancer. *Genes Dev.* **10**, 2089–2101 (1996).
47. Tian, M. & Maniatis, T. A splicing enhancer exhibits both constitutive and regulated activities. *Genes Dev.* **8**, 1703–1712 (1994).
48. Zuo, P. & Maniatis, T. The splicing factor U2AF35 mediates critical protein–protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* **10**, 1356–1368 (1996).
49. Rudner, D. Z., Breger, K. S. & Rio, D. C. Molecular genetic analysis of the heterodimeric splicing factor U2AF: the RS domain on either the large or small *Drosophila* subunit is dispensable *in vivo*. *Genes Dev.* **12**, 1010–1021 (1998).
50. Kan, J. L. & Green, M. R. Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. *Genes Dev.* **13**, 462–471 (1999).
51. Li, Y. & Blencowe, B. J. Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. *J. Biol. Chem.* **274**, 35074–35079 (1999).
52. Eldridge, A. G., Li, Y., Sharp, P. A. & Blencowe, B. J. The SRm160/300 splicing co-activator is required for exon-enhancer function. *Proc. Natl Acad. Sci. USA* **96**, 6125–6130 (1999).
53. Blencowe, B. J. *et al*. The SRm160/300 splicing co-activator subunits. *RNA* **6**, 111–120 (2000).
54. Heinrichs, V., Ryner, L. C. & Baker, B. S. Regulation of sex-specific selection of fruitless 5′ splice sites by transformer and transformer-2. *Mol. Cell. Biol.* **18**, 450–458 (1998).
55. Lam, B. J. *et al*. Enhancer-dependent 5′-splice site control of fruitless pre-mRNA splicing. *J. Biol. Chem.* **278**, 22740–22747 (2003).
56. Wu, J. Y. & Maniatis, T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061–1070 (1993).
The original demonstration of specific interaction between RS domains, on which many subsequent models of SR protein action have been based.
57. Graveley, B. R. A protein interaction domain contacts RNA in the prespliceosome. *Mol. Cell* **13**, 302–304 (2004).
58. Hui, J., Stangl, K., Lane, W. S. & Bindereif, A. HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nature Struct. Biol.* **10**, 33–37 (2003).
59. Del Gatto-Konczak, F. *et al*. The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5′ splice site. *Mol. Cell. Biol.* **20**, 6287–6299 (2000).
60. Forch, P. *et al*. The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Mol. Cell*. **6**, 1089–1098 (2000).
61. Forch, P., Puig, O., Martinez, C., Seraphin, B. & Valcarcel, J. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites. *EMBO J.* **21**, 6882–6892 (2002).
62. Le Guiner, C. *et al*. TIA-1 and TIAR activate splicing of alternative exons with weak 5′ splice sites followed by a U-rich stretch on their own pre-mRNAs. *J. Biol. Chem.* **276**, 40638–40646 (2001).
63. Zhu, H., Hasman, R. A., Young, K. M., Kedersha, N. L. & Lou, H. U1 snRNP-dependent function of TIAR in the regulation of alternative RNA processing of the human calcitonin/CGRP pre-mRNA. *Mol. Cell. Biol.* **23**, 5959–5971 (2003).
64. Wagner, E. J. & Garcia-Blanco, M. A. Polypyrimidine tract binding protein antagonizes exon definition. *Mol. Cell. Biol.* **21**, 3281–3288 (2001).
65. Valcarcel, J., Singh, R., Zamore, P. D. & Green, M. R. The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* **362**, 171–175 (1993).
66. Lin, C. H. & Patton, J. G. Regulation of alternative 3′ splice site selection by constitutive splicing factors. *RNA* **1**, 234–245 (1995).
67. Singh, R., Valcarcel, J. & Green, M. R. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173–1176 (1995).
68. Zhang, L., Liu, W. & Grabowski, P. J. Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. *RNA* **5**, 117–130 (1999).
69. Southby, J., Gooding, C. & Smith, C. W. Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of α-actinin mutally exclusive exons. *Mol. Cell. Biol.* **19**, 2699–2711 (1999).
70. Chou, M. Y., Underwood, J. G., Nikolic, J., Luu, M. H. T. & Black, D. L. Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of *c-src* neural-specific splicing. *Mol. Cell* **5**, 949–957 (2000).
Using site-specifically labelled RNAs, the authors show cooperative binding of PTB to sites flanking the *src* N1 exon as the cause of exon skipping.
71. Wagner, E. J. & Garcia-Blanco, M. A. RNAi-mediated PTB depletion leads to enhanced exon definition. *Mol. Cell* **10**, 943–949 (2002).
Demonstrates, by RNAi and artificial recruitment, that PTB functions as a splicing repressor *in vivo*.
72. Wang, J. & Bell, L. R. The Sex-lethal amino terminus mediates cooperative interactions in RNA binding and is essential for splicing regulation. *Genes Dev.* **8**, 2072–2085 (1994).
73. Lallena, M. J., Chalmers, K. J., Llamazares, S., Lamond, A. I. & Valcarcel, J. Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45. *Cell* **109**, 285–296 (2002).
Shows that SXL represses splicing of its own exon 3, apparently by inhibiting splicing after the first catalytic step.
74. Mayeda, A. & Krainer, A. R. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**, 365–375 (1992).
The first demonstration of antagonistic effects of hnRNPA1 and SF2/ASF upon alternative splicing.
75. Eperon, I. C. *et al*. Selection of alternative 5′ splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol. Cell. Biol.* **20**, 8303–8318 (2000).
76. Caceres, J. F., Stamm, S., Helfman, D. M. & Krainer, A. R. Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science* **265**, 1706–1709 (1994).
77. van der Houven van Oordt, W. *et al*. The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J. Cell Biol.* **149**, 307–316 (2000).
An important demonstration of a signalling pathway that alters alternative splicing patterns by relocalizing hnRNP A1 to the cytoplasm, thereby altering the nuclear ratio of hnRNP A1 and SF2/ASF.
78. Blanchette, M. & Chabot, B. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* **18**, 1939–1952 (1999).
79. Zhu, J., Mayeda, A. & Krainer, A. R. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol. Cell* **8**, 1351–1361 (2001).
Characterizes an ESE that functions by blocking the propagation of hnRNP A1 binding from a downstream ESS; in contrast to other ESEs, this one relies only on the RRMs and not the RS domain of bound SF2/ASF.
80. Del Gatto Konczak, F., Olive, M., Gesnel, M. C. & Breathnach, R. hnRNP A1 recruited to an exon *in vivo* can function as an exon splicing silencer. *Mol. Cell. Biol.* **19**, 251–260 (1999).
Showed that an ESS functions by binding hnRNP A1, and that recruitment of just the glycine-rich domain of hnRNP A1 is sufficient for ESS function.
81. Tange, T. O., Damgaard, C. K., Guth, S., Valcarcel, J. & Kjems, J. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J.* **20**, 5748–5758 (2001).
82. Villemaire, J., Dion, I., Elela, S. A. & Chabot, B. Reprogramming alternative pre-messenger RNA splicing through the use of protein-binding antisense oligonucleotides. *J. Biol. Chem.* **278**, 50031–50039 (2003).
83. Baraniak, A. P., Lasda, E. L., Wagner, E. J. & Garcia-Blanco, M. A. A stem structure in fibroblast growth factor receptor 2 transcripts mediates cell-type-specific splicing by approximating intronic control elements. *Mol. Cell. Biol.* **23**, 9327–9337 (2003).
84. Kanopka, A. *et al*. Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* **393**, 185–187 (1998).
85. Dauksaite, V. & Akusjarvi, G. Human splicing factor ASF/SF2 encodes for a repressor domain required for its inhibitory activity on pre-mRNA splicing. *J. Biol. Chem.* **277**, 12579–12586 (2002).
Using artificial recruitment, the authors show that the second RRM domain of SF2/ASF functions as a splicing repressor when recruited upstream of a regulated 3′ splice site.
86. Estmer Nilsson, C. *et al*. The adenovirus E4-ORF4 splicing enhancer protein interacts with a subset of phosphorylated SR proteins. *EMBO J.* **20**, 864–871 (2001).
87. Shen, H., Kan, J. L., Ghigna, C., Biamonti, G. & Green, M. R. A single polypyrimidine tract binding protein (PTB) binding site mediates splicing inhibition at mouse *IgM* exons M1 and M2. *RNA* **10**, 787–794 (2004).
88. Labourier, E., Adams, M. D. & Rio, D. C. Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol. Cell* **8**, 363–373 (2001).
Shows a direct interaction between P-element somatic inhibitor and U1 snRNP 70K protein, which leads to splicing repression by diverting U1 binding to pseudo-5′ splice sites in the upstream exon.
89. Cote, J., Dupuis, S. & Wu, J. Y. Polypyrimidine track-binding protein binding downstream of caspase-2 alternative exon 9 represses its inclusion. *J. Biol. Chem.* **276**, 8535–8543 (2001).

90. Jumaa, H. & Nielsen, P. J. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.* **16**, 5077–5085 (1997).
91. Labourier, E. *et al*. Antagonism between RSF1 and SR proteins for both splice-site recognition *in vitro* and *Drosophila* development. *Genes Dev.* **13**, 740–753 (1999).
92. Labourier, E. *et al*. Recognition of exonic splicing enhancer sequences by the *Drosophila* splicing repressor RSF1. *Nucleic Acids Res.* **27**, 2377–2386 (1999).
93. Barnard, D. C. & Patton, J. G. Identification and characterization of a novel serine-arginine-rich splicing regulatory protein. *Mol. Cell. Biol.* **20**, 3049–3057 (2000).
94. Li, J. *et al*. Regulation of alternative splicing by SRrp86 and its interacting proteins. *Mol. Cell. Biol.* **23**, 7437–7447 (2003).
95. Ladd, A. N., Charlet-B., N. & Cooper, T. A. The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol. Cell. Biol.* **21**, 1285–1296 (2001).
96. Charlet-B., N., Logan, P., Singh, G. & Cooper, T. A. Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol. Cell* **9**, 649–658 (2002).
97. Zhang, W., Liu, H., Han, K. & Grabowski, P. J. Region-specific alternative splicing in the nervous system: implications for regulation by the RNA-binding protein NAPOR. *RNA* **8**, 671–685 (2002).
98. Suzuki, H., Jin, Y., Otani, H., Yasuda, K. & Inoue, K. Regulation of alternative splicing of α-actinin transcript by Bruno-like proteins. *Genes Cells* **7**, 133–141 (2002).
99. Gromak, N., Matlin, A. J., Cooper, T. A. & Smith, C. W. Antagonistic regulation of α-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein. *RNA* **9**, 443–456 (2003).
100. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
101. Markovtsov, V. *et al*. Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol. Cell. Biol.* **20**, 7463–7479 (2000).
102. Rooke, N., Markovtsov, V., Cagavi, E. & Black, D. L. Roles for SR proteins and hnRNP A1 in the regulation of *c-src* exon N1. *Mol. Cell. Biol.* **23**, 1874–1884 (2003).
103. Ule, J. *et al*. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
Reports a novel technique that allowed the identification of cellular targets of Nova proteins in mouse brain, many of which were subsequently validated by analysing their splicing patterns in Nova-knockout mice.
104. Polydorides, A. D., Okano, H. J., Yang, Y. Y., Stefani, G. & Darnell, R. B. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc. Natl Acad. Sci. USA* **97**, 6350–6355 (2000).
105. Ladd, A. N. & Cooper, T. A. Finding signals that regulate alternative splicing in the post–genomic era. *Genome Biol.* **3**, reviews 0008 (2002).
106. Zheng, Z. M. Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *J. Biomed. Sci.* **11**, 278–294 (2004).
107. Tacke, R. & Manley, J. L. Determinants of SR protein specificity. *Curr. Opin. Cell Biol.* **11**, 358–362 (1999).
108. Burd, C. G. & Dreyfuss, G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**, 1197–1204 (1994).
109. Perez, I., Lin, C. H., McAfee, J. G. & Patton, J. G. Mutation of PTB binding sites causes misregulation of alternative 3′ splice site selection *in vivo*. *RNA* **3**, 764–778 (1997).
110. Jin, Y. *et al*. A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.* **22**, 905–912 (2003).
111. Brudno, M. *et al*. Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Res.* **29**, 2338–2348 (2001).
112. Liu, H. X., Chew, S. L., Cartegni, L., Zhang, M. Q. & Krainer, A. R. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol. Cell. Biol.* **20**, 1063–1071 (2000).
113. Liu, H. X., Zhang, M. & Krainer, A. R. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* **12**, 1998–2012 (1998).
References 112 and 113 describe a functional SELEX approach in which ESEs were selected in cytoplasmic S100 extract in the presence of individual SR proteins, allowing the identification of ESEs specific for each protein. This allowed the development of the ESEfinder web resource.
114. Schaal, T. D. & Maniatis, T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.* **19**, 1705–1719 (1999).
115. Tian, H. & Kole, R. Selection of novel exon recognition elements from a pool of random sequences. *Mol. Cell. Biol.* **15**, 6291–6298 (1995).
116. Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31**, 3568–3571 (2003).
117. Coulter, L. R., Landree, M. A. & Cooper, T. A. Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.* **17**, 2143–2150 (1997).
A functional SELEX approach, based on the recovery of sequences that led to exon inclusion, allowed the identification of various ESEs, including a novel AC-rich class.
118. Stickeler, E. *et al*. The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *EMBO J.* **20**, 3821–3830 (2001).
119. Fedorov, A., Saxonov, S., Fedorova, L. & Daizadeh, I. Comparison of intron-containing and intron-lacking human genes elucidates putative exonic splicing enhancers. *Nucleic Acids Res.* **29**, 1464–1469 (2001).
120. Fairbrother, W. G., Yeh, R. F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
A computational survey for ESEs, based on the enrichment in exons versus introns, and in exons with weak splice sites versus exons with strong splice sites; led to the development of the RESCUE-ESE web resource.
121. Yeo, G., Hoon, S., Venkatesh, B. & Burge, C. B. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl Acad. Sci. USA* **101**, 15700–15705 (2004).
122. Minovitsky, S., Gee, S. L., Schokrpur, S., Dubchak, I. & Conboy, J. G. The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *Nucleic Acids Res.* **33**, 714–724 (2005).
123. Miriami, E., Margalit, H. & Sperling, R. Conserved sequence elements associated with exon skipping. *Nucleic Acids Res.* **31**, 1974–1983 (2003).
124. Yeo, G., Holste, D., Kreiman, G. & Burge, C. B. Variation in alternative splicing across human tissues. *Genome Biol.* **5**, R74 (2004).
Analysis of publicly available tissue-specific EST and microarray expression data showed that liver had the most distinctive patterns of alternative splicing and expression of SR and hnRNP proteins.
125. Tenenbaum, S. A., Carson, C. C., Lager, P. J. & Keene, J. D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl Acad. Sci. USA* **97**, 14085–14090 (2000).
126. Blanchette, M., Labourier, E., Green, R. E., Brenner, S. E. & Rio, D. C. Genome-wide analysis reveals an unexpected function for the *Drosophila* splicing factor U2AF(50) in the nuclear export of intronless mRNAs. *Mol. Cell* **14**, 775–786 (2004).
Conditional mutants for the *D. melanogaster* large subunit of U2AF and RNAi knockdown indicate that the U2AF large subunit has an unexpected role in nuclear export of intronless RNAs.
127. Mili, S. & Steitz, J. A. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694 (2004).
128. Buckanovich, R. J. & Darnell, R. B. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets *in vitro* and *in vivo*. *Mol. Cell. Biol.* **17**, 3194–3201 (1997).
129. Kim, S., Shi, H., Lee, D. K. & Lis, J. T. Specific SR protein-dependent splicing substrates identified through genomic SELEX. *Nucleic Acids Res.* **31**, 1955–1961 (2003).
Successful application of genomic SELEX to identify cellular RNA targets of *D. melanogaster* B52 (SRp55), a number of which were validated in functional assays.
130. Hieronymous, H. & Silver, P. A. A systems view of mRNP biology. *Genes Dev.* **18**, 2845–2860 (2004).
131. Kotovic, K. M., Lockshon, D., Boric, L. & Neugebauer, K. M. Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol. Cell. Biol.* **23**, 5768–5779 (2003).
132. Longman, D., Johnstone, I. L. & Caceres, J. F. Functional characterization of SR and SR-related genes in *Caenorhabditis elegans*. *EMBO J.* **19**, 1625–1637 (2000).
133. Wang, H. Y., Xu, X., Ding, J. H., Bermingham, J. R. Jr & Fu, X. D. SC35 plays a role in T cell development and alternative splicing of CD45. *Mol. Cell* **7**, 331–342 (2001).
134. Huang, Y. & Steitz, J. A. SRprises along a messenger's journey. *Mol. Cell* **17**, 613–615 (2005).
135. Rélogio, A. *et al*. Alternative splicing microarrays reveal functional expression of neuron specific regulators in Hodgkin lymphoma cells. *J. Biol. Chem.* **280**, 4779–4784 (2005).
Used a microarray that contained probes specific for alternative splicing events as well as the expression of splicing regulators. This design potentially allows the deciphering of splicing factor codes that regulate particular splicing events.
136. Pan, Q. *et al*. Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* **16**, 929–941 (2004).
A quantitative alternative splicing microarray analysed 3,126 known alternative splicing events, allowing many new insights into the global aspects of alternative splicing programmes.
137. Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A. & Smith, C. W. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **13**, 91–100 (2004).
Demonstrates a negative feedback loop in which PTB promotes skipping of its own exon 11, producing an RNA that is degraded by NMD.
138. Park, J. W., Parisky, K., Celotto, A. M., Reenan, R. A. & Graveley, B. R. Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc. Natl Acad. Sci. USA* **101**, 15974–15979 (2004).
An RNAi screen allowed the identification of a large number of splicing regulators, including a number of constitutive splicing factors.
139. Lim, S. R. & Hertel, K. J. Commitment to splice site pairing coincides with A complex formation. *Mol. Cell* **15**, 477–483 (2004).
Shows that although pre-mRNAs are committed to splicing in the E complex, the decision as to which splice sites will be paired is subsequently made in the A complex.
140. Berget, S. M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414 (1995).
141. Chiara, M. D. & Reed, R. A two-step mechanism for 5′ and 3′ splice-site pairing. *Nature* **375**, 510–513 (1995).
142. Niranjanakumari, S., Lasda, E., Brazas, R. & Garcia-Blanco, M. A. Reversible crosslinking combined with immunoprecipitation to study RNA–protein interactions *in vivo*. *Methods* **26**, 182–190 (2002).
143. Wang, H. *et al*. Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics* **19** (Suppl. 1), i315–i322 (2003).
144. Hiller, M. *et al*. Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nature Genet.* **36**, 1255–1257 (2004).

Acknowledgements

We thank Juan Valcárcel and Sushma Nagaraja Grellscheid for critically reading the manuscript. Work in the laboratory of C.W.J.S. is supported by a programme grant from the Wellcome Trust. We apologize to those colleagues whose work, for reasons of space constraints, we have been unable to cite directly.

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez
protein phosphatase 2A
Swiss-Prot: http://us.expasy.org/sprot
hnRNPA1 | hnRNPL | KSRP | PTB | SC35 | SF2/ASF | SRm160/SRm300 | SRp20 | SRrp86 | U2AF35 | U2AF65
FlyBase: http://flybase.bio.indiana.edu
Dscam | *dsx* | *fruitless* | PSI | RSF1 | *Sxl* | *tra* | TRA2 | U1 70K

FURTHER INFORMATION

ESEfinder: http://rulai.cshl.edu/tools/ESE
RESCUE-ESE: http://genes.mit.edu/burgelab/rescue-ese
Alternative Splicing Database (ASD): http://www.ebi.ac.uk/asd/index.html
Access to this interactive links box is free online.