**Alternative splicing regulatory networks: functions, mechanisms and evolution**

Jernej Ule1,2,5 and Benjamin J. Blencowe3,4,5

---

1 The Francis Crick Institute, London NW1 1AT, UK
2 Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, Queen Square, London, WC1N 3BG, UK
3 Donnelly Centre, University of Toronto, Toronto, Ontario M5S3E1, Canada
4 Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5S3E1, Canada

5 Correspondence: jernej.ule@crick.ac.uk; b.blencowe@utoronto.ca

**SUMMARY**

High-throughput sequencing-based methods and their applications in the study of transcriptomes have revolutionized our understanding of alternative splicing. Networks of functionally coordinated and biologically important alternative splicing events continue to be discovered in an ever increasing diversity of cell types in the context of physiologically normal and disease states. These studies have been complemented by efforts directed at defining sequence codes governing splicing and their cognate trans-acting factors, which have illuminated important combinatorial principles of regulation. Additional studies have revealed critical roles for position-dependent, multivalent protein-RNA interactions that direct splicing outcomes. Investigations of evolutionary changes in RNA binding proteins, splice variants, and associated cis-elements, have further shed light on the emergence, mechanisms and functions of splicing networks. Progress in these areas has emphasized the need for a coordinated, community-based effort to systematically address the functions of individual splice variants associated with normal and disease biology.
Introduction

Transcripts from nearly all human protein coding genes undergo one or more forms of alternative splicing, such as the inclusion or skipping of individual ‘cassette’ exons, switching between alternative 5´ and 3´ splice sites, differential retention of introns, mutually exclusive splicing of adjacent exons and other, more complex patterns of splice site selection (Pan et al., 2008; Wang et al., 2008). All of these forms of splicing require the spliceosome, a megadalton machine that catalyzes splicing reactions (Wahl et al., 2009). Spliceosome formation entails a complex interplay of trans-acting factors, including the small nuclear ribonucleoprotein particles (U1, U2, U4/U6 and U5 snRNPs), comprising snRNAs and associated proteins, together with ~150 additional proteins. The formation of spliceosomes and their mechanism of action has been illuminated in remarkable detail by a series of recent cryo-electron microscopy structures, work that has been reviewed elsewhere (Kastner et al., 2019; Plaschka et al., 2019; Yan et al., 2019).

The binding of snRNPs to pre-mRNA is typically stabilised by mutual “definition” interactions across introns and exons (De Conti et al., 2013) (Figure 1). Intron definition interactions predominate when introns are relatively short (e.g. in the range of up to a few hundred nucleotides), as is the case in yeast and most invertebrate species. In contrast, exon definition interactions predominate in vertebrates (Robberson et al., 1990) where introns have a median length of approximately 1kb (Hong et al., 2006). In either case, the principles governing splice site recognition and pairing are thought to be similar. For example, current models posit that U1 snRNP binds to the 5´ splice site and communicates via bridging interactions with splicing factor 1 (SF1) and the U2 snRNP auxiliary factor (a heterodimer of U2AF1 and U2AF2) bound to the 3´ splice site and its adjacent polypyrimidine tract (Abovich and Rosbash, 1997; De Conti et al., 2013). Additional interactions that contribute to exon and intron definition are mediated by members of the RNA Recognition Motif (RRM)-containing SR family of proteins (referred to below as ‘SR proteins’) and SR-related proteins, each of which contain one or more intrinsically disordered region (IDR) rich in alternating arginine and serine residues, referred to as the ‘RS domain’ (Figure 1). For example, it has been proposed that when SR proteins bind to exonic enhancer sequences, their RS domains interact with the RS domains of the U1 snRNP-specific 70KDa protein (SNRNP70), and U2AF1, to promote exon definition (Wu and Maniatis, 1993). Similarly, in S. pombe it has been shown that intron definition is promoted by interactions between the RS domains of Prp5 (DDX46 in human), which interacts with U2 snRNP, and Rsd1 (RBM38 in human), which interacts with U1 snRNP (Shao et al., 2012).

Numerous additional interactions come into play to forge intron- and exon-definition interactions. For example, the SR-related proteins SRRM1 and SRRM2 can bridge interactions between snRNPs bound at splice sites and SR proteins bound at exonic enhancers (Eldridge et al., 1999). Moreover, phosphorylated RS domains were reported to bind double-stranded RNA, which can promote base-pairing between snRNPs and pre-mRNA (Shen and Green, 2006). Collectively, these and additional early interactions, some of which are described later, facilitate the stable recruitment of U2 snRNP to the pre-mRNA branch site, followed by addition of U4/U6 and U5 snRNPs in the form of a tri-snRNP particle. The actions of many RNA helicases then promote rearrangements of snRNP interactions and the establishment of a catalytically competent spliceosome that carries out the two
trans-esterification reactions that lead to lariat formation, intron removal and exon ligation (Wahl et al., 2009).

Many types of RNA binding proteins (RBPs) can regulate alternative splicing. In addition to SR proteins, these include the heterogeneous ribonucleoprotein (hnRNP) family of proteins, as well as RBPs containing RRM, KH, zinc finger, or other domains (Lunde et al., 2007). The full set of proteins that control alternative splicing is not known, although recent large-scale screens employing systematic RNA interference or clustered regularly interspaced short palindromic repeats (CRISPR)-Cas-mediated ablation of genes have illuminated repertoires involving a few hundred proteins that act directly or indirectly to regulate specific alternative exons (Gonatopoulos-Pournatzis et al., 2018; Han et al., 2017; Papasaikas et al., 2015; Tejedor et al., 2015). Among other unexpected factors, these studies have highlighted previously annotated DNA binding proteins as having potential direct roles in RNA binding and splicing regulation. RBPs bind cis-elements in introns and exons and regulate splice site selection largely by promoting or repressing definition interactions (De Conti et al., 2013; Fu and Ares, 2014). They thus act mainly at the early stages of spliceosome formation, although regulation can also be imparted at later stages of assembly (Wahl et al., 2009).

In this review, we highlight recent advances in the identification and characterization of networks of splicing regulation, including significant strides that have been made in the systematic analysis of RBPs and associated regulatory mechanisms through the application of in vitro binding (Dominguez et al., 2018) and in vivo cross-linking and immunoprecipitation (CLIP) methods (Lee and Ule, 2018; Van Nostrand et al., 2018). Proteomics (Hentze et al., 2018) functional genomics (Gonatopoulos-Pournatzis et al., 2018) and increasingly powerful computational approaches (Baeza-Centurion et al., 2019; Jaganathan et al., 2019). We review how these and other complementary approaches are further providing unprecedented new insights into the evolution of mechanisms governing alternative splicing, as well as how the disruption of these mechanisms causes or contributes to human diseases and disorders. Finally, we conclude by discussing challenges for the field that lie ahead.

Biological significance of alternative splicing regulatory networks

The development and application of custom microarrays, and later high-throughput RNA sequencing (RNA-Seq) methods, revealed the extraordinary complexity of regulated alternative splicing in metazoans, particularly in vertebrate species (Blencowe, 2015; Fu and Ares, 2014). Recent transcriptome sequencing efforts, involving both short- and long-read technologies, are increasingly focusing on specialized cell types and individual cells from different organs. To date, dynamic alternative splicing networks have been detected in embryonic stem and precursor cells, during the differentiation or reprogramming of various cell lineages as well as epithelial-mesenchymal transitions, and in adult organs such as the brain, heart, skeletal muscle, liver, kidney, adipose tissue, testes and the immune system (Baralle and Giudice, 2017; Bhat et al., 2015; Gabut et al., 2011; Han et al., 2013; Irimia et al., 2014; Kalsotra and Cooper, 2011; Licatalosi and Darnell, 2010; Mallory et al., 2015; Tapial et al., 2017; Wong et al., 2013; Zhang et al., 2016). Additional regulated alternative splicing networks have been detected in association with specific physiologic states of cells, such as thermal regulation and the stress response (Boutz et al., 2015; Gotic et al., 2016; Low et al., 2017).
Many regulatory RBPs function in a cell, tissue or condition-specific manner and are capable of coordinately regulating functionally coherent ‘networks’ of exons and introns (Braunschweig et al., 2013; Licatalosi and Darnell, 2010). Thus, our understanding of the repertoires of detected splice variants as well as other forms of transcript variation across cellular conditions in the context of normal and disease physiology continues to rise dramatically.

Notably, regulated alternative exons that overlap protein coding sequences are often located within predicted IDRs that are coincident with sites of post-translational modifications and protein-protein interactions, and the role of alternative splicing in diversifying protein interaction capabilities has been demonstrated experimentally (Buljan et al., 2012; Ellis et al., 2012; Yang et al., 2016). An important and likely general function of alternative splicing networks is therefore to control protein-protein interactions to impart important cellular, tissue- and condition-specific functions of widely expressed genes. In addition to remodelling the IDRs, a smaller number of conserved events in alternative splicing networks directly overlap critical modular protein domains and impact various important protein functions, such as those involved in control of transcription and chromatin, as reviewed elsewhere (Irimia and Blencowe, 2012; Kelemen et al., 2013; Porter et al., 2018). The identification and characterization of such events highlights the capacity of alternative splicing networks to have a broad impact on physiology through their ability to cross-talk with orthogonal gene regulatory layers.

A neuronal microexon network

Comparative vertebrate transcriptomics of a common set of organs from fish to human revealed that brain-specific alternative splicing events are among the most evolutionarily conserved (Barbosa-Morais et al., 2012; Merkin et al., 2012). Particularly striking in this regard is a network of a few hundred 3-27 nt neuronal microexons, which represent the most highly conserved class of alternative splicing events discovered to date (Irimia et al., 2014; Li et al., 2015). These microexons predominantly preserve open reading frames and insert one to several amino acids in proteins associated with neurogenesis, axon guidance and synaptic functions (Irimia et al., 2014; Johnson et al., 2019; Quesnel-Vallières et al., 2019; Ustianenko et al., 2017). Similar to longer exons, individual neuronal microexons can also impact orthogonal regulatory layers; for example, they alter the function of the lysine-specific histone demethylase 1A (KDM1A, also known as Lsd1), control activation domains of Mef2 family transcription factors, and regulate the activity of the cytoplasmic polyadenylation element binding protein 4 (CPEB4), which controls polyA tail length and translation (Ebert and Greenberg, 2013; Parras et al., 2018; Rusconi et al., 2017).

Notably, misregulation of activity-dependent splicing of microexons likely plays a causative role in autism, at least in part as a consequence of disrupted expression of the major transacting regulator of microexons, the neuronal SR-related protein nSR100/SRRM4 (Irimia et al., 2014; Quesnel-Vallières et al., 2015, 2016). SRRM4 activates splicing of microexons by binding specialized upstream intronic enhancer elements together with the SR proteins SRSF11 and RNPS1 (Gonatopoulos-Pournatzis et al., 2018; Raj et al., 2014) (Figure 1b). The C-terminal IDR of SRRM4 contains an ‘enhancer of microexons’ (eMIC) domain, which interacts with the branch point binding protein SF1, and U2AF, to promote the recruitment of U2 snRNP, and thus activates the earliest
Mechanisms underlying the regulation of alternative splicing networks

**Multivalency and RNP condensation in splicing regulation**

Key to a more general understanding of the function and regulation of alternative splicing networks is the systematic identification of corresponding splicing regulators and their cognate cis-acting binding sites. Most RNA binding domains of RBPs recognise short (i.e. 3-4nt) and degenerate motifs, and the capacity of such individual motifs to predict splicing outcomes is low. Instead, mutagenesis of minigene reporters as well as transcriptomic studies have demonstrated that many RBPs need to cooperatively interact with repeating, or ‘multivalent’ motifs around alternative exons in order to efficiently regulate splicing ([Cereda et al., 2014; Chou et al., 2000; Fu and Ares, 2014; Ule et al., 2003])

Biophysical studies with purified RBPs and RNA have demonstrated that multivalent interactions can drive condensation into dynamic complexes through the phenomenon of liquid-liquid phase separation (LLPS) ([Banani et al., 2017; Kato and McKnight, 2018; Li et al., 2012]) Such condensation can increase the concentration of molecules in RNP condensates by two orders of magnitude compared with the surrounding liquid ([Li et al., 2012]) and this likely enables RBPs to bind more stably to RNA transcripts with multivalent sites, compared to the interaction of a single RNA-binding domain of an RBP with an individual RNA motif. For example, incubation of purified PTBP1 with an RNA oligonucleotide containing multiple CU motifs leads to LLPS under *in vitro* conditions ([Li et al., 2012])

PTBP1 contains four RRM domains, each of which can bind a short CU-rich motif ([Chou et al., 2000] [Chou et al., 2000; Oberstrass et al., 2005]) and many other RBPs contain multiple domains with similar sequence specificity ([Lunde et al., 2007]) (Figure 2a), indicating that multivalent RNA sites could promote their condensation.

Even RBPs that contain only one or two RNA-binding domains can undergo RNP condensation on RNAs with multivalent motifs if their IDRs promote homomultimerisation. This is often achieved through short, multivalent protein-protein interacting motifs in IDRs, such as repeating amino acids, short linear motifs and/or dynamic secondary structures that are capable of relatively promiscuous and weak homomeric interactions ([Banani et al., 2017; Kato and McKnight, 2018]). Multiple recent studies have demonstrated that RNP condensation mediated by IDRs, and the multivalent binding properties of RBPs, can function in transcriptome-wide assembly and function of splicing regulators ([Attig et al., 2018; Gueroussov et al., 2017; Ying et al., 2017; Zhou et al., 2019]) (Figure 2b). For instance, the tyrosine-rich IDR of RBFOX1 can mediate LLPS *in vitro*, while *in vivo* it is crucial for
the interaction of RBFOX1 with a large RNP complex referred to as LASR, which contains eight RBPs, including hnRNPC, hnRNPH, hnRNPM and MATR3 (Damianov et al., 2016; Ying et al., 2017). Interestingly, the IDR is not required for sequence-specific binding of RBFOX1 to RNA, but is required for its assembly into higher-order complexes and localization in nuclear speckles, which affects the capacity of RBFOX1 to regulate a subset of its target alternative exons. Notably, the IDR of RBFOX1 also contributes to the formation of RNA granules upon stress as part of a cellular adaptive response (Kucherenko and Shcherbata, 2018) indicating that the multivalent interactions formed by the IDR can contribute to the assembly of multiple types of RNP condensates.

Another recent study has revealed that IDRs of many RBPs, including those of most members of the hnRNP A and D families, contain glycine/tyrosine repeats that can promote LLPS in vitro, and promote binding to multivalent RNA sites to globally regulate alternative splicing (Gueroussov et al., 2017). Interestingly, the C-terminal IDRs of hnRNP A and D family proteins were found to contain conserved exons that are alternatively spliced in mammals but constitutively spliced in other vertebrate species. The differential inclusion of these exons controls the formation of tyrosine-dependent multivalent hnRNP assemblies that, in turn, function to globally regulate splicing. Thus, IDRs in RBPs contribute to the regulation of alternative splicing, and are themselves often regulated through alternative splicing. Notably, as has been proposed for transcriptional control (Hnisz et al., 2017) RNP condensation could increase the responsiveness of regulatory networks to local concentrations and modifications of IDR-containing proteins, which in turn may play an important role in mediating inputs from external and internal signals.

RNA position dependence of splicing regulation

Analysis of RBP binding profiles around co-regulated alternative splicing events has been an effective means of unravelling regulatory mechanisms, which can be summarized by three main insights. First, SR proteins typically enhance splicing when binding between other spliceosomal components, because their RS domains generally mediate heteromeric interactions with other RS domain-containing splicing factors in a manner that promotes combinatorial assembly of the spliceosome on RNA elements, including the branchpoint, splice sites and enhancer motifs. SR proteins typically bind to exonic enhancer sequences to activate splicing (Fu and Ares, 2014), but in the case of microexons, nSR100/SRRM4 binds to intronic enhancer sequences upstream of the 3’ splice site (Raj et al., 2014) (Figure 1b). When binding at these positions, SR proteins generally oppose the repressive effects of hnRNPs (Chen and Manley, 2009) for example, SRRM4 opposes the repressive effects of PTBP1 on the inclusion of microexons (Gonatopoulos-Pournatzis et al., 2018; Li et al., 2015; Raj et al., 2014) In contrast to their general enhancing activities, some characterized SR proteins were reported to repress splicing when binding to intronic sequences downstream of target exons (Erkelenz et al., 2013; Ibrahim et al., 2005; Kanopka et al., 1996). It remains to be determined whether such reciprocal patterns of SR proteins commonly contribute to splicing regulation across the transcriptome.

Second, computational studies have been used to construct “RNA maps” of RBP activity in splicing regulation by evaluating the position of their binding around exons. Initially, studies of Nova proteins
revealed RNA maps accounting for their dual effects on splicing in the brain that depend on their binding position in pre-mRNA, defined either by analysis of multivalent motifs or with CLIP data (Licatalosi et al., 2008; Ule et al., 2006). The multivalency of RNA motifs has been exploited to further develop computational approaches to identify regulatory motifs and RNA maps at co-regulated exons, recently with the additional context of the predicted RNA structure (Cereda et al., 2014; Dominguez et al., 2018; Feng et al., 2019). Through the efforts of many laboratories, RNA maps of numerous additional RBPs have been determined around regulated exons, thus revealing their position-dependent splicing regulation (Van Nostrand et al., 2018; Witten and Ule, 2011). For instance, Nova, PTBP1, or TDP-43 tend to repress exon inclusion when their binding overlaps an exon or splice site, but activate splicing when bound to intronic sequences downstream of the 5´ splice site (Llorian et al., 2010; Tollervey et al., 2011; Ule et al., 2006; Xue et al., 2009) (Figure 3a, 3b). These RBPs generally rely on IDR(s) that connect multiple domains with similar RNA specificity (as in the case of Nova or PTBP1), or promote formation of homomeric RBP complexes (as in the case of TDP-43), which gives these RBPs the flexibility to recognise variably-spaced multivalent motifs (Figure 2). As such, these RBPs can bind to multivalent motifs that are located between splice sites and SR protein binding sites, which can lead to efficient competition for spliceosome assembly on pre-mRNA (Figure 3a).

The mechanisms of enhancing effects of non-SR proteins have been explored in depth for TIA1, which interacts with the U1 snRNP-specific C protein to recruit U1 snRNP to the 5´ splice site (Förch et al., 2002) and PTBP1, which was proposed to interact with the stem loop 4 of U1 snRNA to stimulate docking of U1 snRNP in a productive or nonproductive conformation depending on the position of PTBP1 binding on pre-mRNA (Hamid and Makeyev, 2017; Sharma et al., 2011) (Figure 3b). Interestingly, stem loop 4 of U1 snRNA also interacts with the SF3a component of the U2 snRNP, which can promote intron or exon definition (Sharma et al., 2014). The position-dependent rules governing alternative splicing are similar across hnRNPs and many other RBPs in various types of cells and tissues in species such as C. elegans, Drosophila, mouse and human (Ajith et al., 2016; Brooks et al., 2011; Kuroyanagi et al., 2013; Wang et al., 2012; Witten and Ule, 2011). These diverse RBPs recognise a range of different RNA motifs and have limited sequence homology with each other, and thus it remains to be determined if they operate through common mechanisms, such as employing interactions with different components of U1 snRNP.

Third, RBPs can follow a reciprocal position dependence also when regulating an alternative exon indirectly via the adjacent exon. For instance, while Nova and TIA proteins enhance inclusion when binding downstream of an alternative exon, they can also repress inclusion of an alternative exon when binding downstream of a preceding constitutive exon (Ule et al., 2006; Wang et al., 2010) (Figure 3c). Similar reciprocity at adjacent exons can be seen in the RNA maps of several other RBPs studied by the ENCODE project (Van Nostrand et al., 2018). Reciprocal effects of regulation are commonly seen at adjacent splice sites as a result of splice site competition, since silencer sequences are most effective on a splice site when a competing splice site is present nearby (Yu et al., 2008). Competition between splice sites of adjacent exons can also control splicing outcomes if both introns that flank the alternative exon are available when the splicing choice is made. Indeed, splicing of alternative exons is often somewhat delayed after transcription, which might enable such competition (Drexler et al., 2019; Kim et al., 2017). Moreover, studies of Nova-regulated exons
showed that the excision of introns is often asymmetric, such that a specific intron is spliced first in the pathway to exon inclusion, and Nova tends to bind this intron when directly regulating exon inclusion (Ule et al., 2006). However, Nova can also indirectly repress inclusion of an alternative exon by enhancing the 5′ splice site at the preceding exon, perhaps through competition with the exon inclusion pathway that requires initial splicing of the intron downstream of the alternative exon (Figure 3c). Interestingly, splice site competition has also recently been shown to cause a non-linearity in the genotype-phenotype map of exonic mutations (Baeza-Centurion et al., 2019). The broader importance of the order of intron excision and splice site competition in splicing regulation, and its interplay with multivalency and positioning of cis-acting motifs, remains to be fully established.

Regulatory roles for cryptic splice sites and exons

Cryptic splice sites and intronic decoys

One of the most remarkable feats of the cell is its ability to accurately recognize and differentially regulate correct splice sites from the myriad sequences that match splice site consensus sequences but are not selected to form mature mRNA. These cryptic sites can be distinguished from canonical splice sites in part because their positioning is not compatible with efficient definition interactions. Moreover, they are often flanked by a high density of splicing silencer motifs, which bind to various hnRNPs that repress their recognition (Sibley et al., 2016; Wang et al., 2004; Zhang et al., 2008). For instance, cryptic exons can be repressed by hhnNPC when binding to U-tracts in antisense Alu elements (Zarnack et al., 2013) by Nova when binding to long clusters of YCAY motifs (Eom et al., 2013), and by TDP-43 or hnRNPL when binding to microsatellite-derived UG or CA repeats, respectively (Ling et al., 2015; McClory et al., 2018).

Derepression of cryptic splice sites in human diseases can be caused either by mutations in cis-acting silencing elements or deregulation of trans-acting repressors (Kahles et al., 2018; Ling et al., 2015; Sibley et al., 2016; Zarnack et al., 2013). Even mutations in the core splicing machinery can also lead to derepression of cryptic splice sites and exons, as has been shown by the analysis of cryptic 3′ splice sites induced by cancer-associated SF3B1 mutations (Darman et al., 2015; DeBoever et al., 2015). Interestingly, a Caenorhabditis elegans genetic screen recently identified alleles of the core spliceosome component Prp8 that specifically alter cryptic splicing frequency (Mayerle et al., 2019). Notably, alterations in the accuracy of splice site recognition can lead to substantial transcriptome heterogeneity in cancer, with consequences for the generation of neoantigens of possible relevance in immunotherapy (Kahles et al., 2018). Thus, it is becoming clear that aside from roles in ensuring transcriptome integrity, control of cryptic alternative splicing events can also contribute to the regulation of gene expression, while their perturbation can lead to disease, or can serve as a biomarker or therapeutic target.

Cryptic splice sites and binding sites for splicing factors can also act as intronic decoys that compete with canonical splice sites for spliceosomal recruitment and thereby affect splicing of adjacent introns or exons (Côté et al., 2001; Parra et al., 2018; Sutandy et al., 2018). The activity of such intronic decoys is tightly regulated, as they tend to be flanked by splicing silencer motifs (Wang et al., 2004,
Splicing factors such as U2AF2 can bind to numerous intronic decoys, where it can promote the use of a cryptic splice site, and interfere with splicing at the immediate downstream 3′ splice site (Côté et al., 2001; Shao et al., 2014; Zarnack et al., 2013). Many RBPs can bind to intronic decoys to negatively or positively regulate U2AF2 binding, thereby indirectly regulating splicing decisions (Howard et al., 2018; Sutandy et al., 2018; Zarnack et al., 2013). The potential for splicing regulation through intronic decoys is likely to be greatest in long introns, which are enriched in genes expressed in the brain (Sibley et al., 2015) and for which cell-type-specific CLIP studies are particularly valuable for elucidating competitive and cooperative binding mechanisms and functions of RBPs (Saito et al., 2019; Ule et al., 2018).

Recursive splicing and its regulation

Recursive splicing (ReS) is a process by which introns are excised in multiple steps (Hatton et al., 1998). It is present most often in long introns of Drosophila and human genes, and has been associated with increased splicing accuracy (Burnette et al., 2005; Duff et al., 2015; Pai et al., 2018; Sibley et al., 2015). It is also required for the skipping of cryptic exons that form during ReS, hereafter referred to as ‘ReS-exons’ (Joseph et al., 2018; Sibley et al., 2015) (Figure 4). ReS-exons are initially defined by the spliceosome and spliced to a preceding exon, which generates a ‘reconstituted’ 5′ splice site that is then used in the second splicing event to ‘discard’ the ReS-exon (Sibley et al., 2015). In human genes, most cryptic ReS-exons are present in long first introns of neuronal genes. Interestingly, alternative promoters commonly arise from long introns, leading to production of a new first exon, which is not capable of reconstituting a strong 5′ site after the first step of recursive splicing, an event that leads to inclusion of the ReS-exon. Cryptic ReS-exons contain stop codons in all three frames, and their inclusion thus elicits nonsense-mediated decay (NMD) and terminates translation. In this way, recursive splicing serves as a mechanism for quality control of transcripts produced from alternative promoters by preventing production of aberrant proteins with altered N-termini (Sibley et al., 2015).

Not all ReS-exons are cryptic, as hundreds of annotated ReS-exons are included in human mRNAs because their recursive splicing is repressed by the exon junction complex (EJC) (Blazquez et al., 2018) (Figure 4). The EJC comprises several proteins that associate with sequences ~24 nt upstream of spliced junctions and previously has been implicated in mRNA turnover, export, translation, as well as splicing (Le Hir et al., 2016). It is not entirely clear why the EJC can repress recursive splicing of canonical but not cryptic ReS-exons, although the relative strengths of the canonical and reconstituted 5′ splice sites clearly plays a role (Blazquez et al., 2018; Sibley et al., 2015). In addition to repressing recursive splicing, the EJC can also repress cryptic 5′ and 3′ splice sites located near exon-exon junctions (Blazquez et al., 2018; Boehm et al., 2018) if such a cryptic 5′ splice site is located close to the start of an exon, its recognition leads to inclusion of cryptic ‘microexon’ sequences. It is intriguing that RNPS1, an auxiliary EJC component, can repress inclusion of such cryptic microexons, while on the other hand it promotes splicing of canonical neural microexons (Gonatopoulos-Pournatzis et al., 2018).
It is also interesting to consider that the repressive efficiency of core and auxiliary EJC components likely varies between tissues, which could lead to regulated recursive splicing of canonical exons. Notably, deficient EJC activity in mice has the strongest impact on the brain, where it leads to skipping of ReS-exons in genes with neurodevelopmental functions, likely contributing to a microcephaly phenotype [Blazquez et al., 2018]. Moreover, several human neurodevelopmental disorders are associated with mutations in EJC components [McMahon et al., 2016] and it is plausible that deregulated recursive splicing might contribute to these disorders.

**Emergence and evolution of alternative splicing**

The degenerate nature of splice sites provides opportunities for alternative splicing regulation, which as described above is often achieved through the action of cis-acting splicing enhancer and silencer sequences [Fu and Ares, 2014]. Analysis of these cis-elements by machine learning approaches can infer context-dependent and combinatorial mechanisms to derive “splicing codes” that predict splice site selection in genomic sequence, and in some cases also the regulation of alternative splicing across cell types and tissues [Baeza-Centurion et al., 2019; Barash et al., 2010; Bretschneider et al., 2018; Jaganathan et al., 2019; Louadi et al., 2019; Xiong et al., 2015]. Notably, a splicing code inferred from mouse data can predict splicing patterns in other vertebrate species with reasonable accuracy [Barbosa-Morais et al., 2012], in agreement with the conserved nature of the binding specificities of orthologous RBPs and their cognate cis-acting elements [Irimia et al., 2011; Jelen et al., 2007; Oddo et al., 2016; Solana et al., 2016].

During the early stages of genome sequencing projects, a prediction was made that differences in transcriptome diversity, rather than the absolute numbers of protein coding genes, might more strongly account for the vast range of biological complexity and phenotypic attributes across metazoan species [Ewing and Green, 2000]. Analyses of expressed sequence tags and microarray profiling studies subsequently provided initial evidence that alternative splicing patterns have diverged rapidly among species [Modrek and Lee, 2003; Pan et al., 2004]. Moreover, comparative RNA-Seq analyses of transcriptomes of diverse organ types in vertebrate species spanning ~350 million years of evolution have revealed that whereas organ-dependent gene expression programs have remained relatively conserved, patterns of alternative splicing have diverged remarkably rapidly such that, overall, alternative splicing patterns are more similar between organs within a species than they are between the same organs of different species [Barbosa-Morais et al., 2012; Deveson et al., 2018; Merkin et al., 2012].

The evolutionary origin of alternative splicing can be traced to the last eukaryotic common ancestor, which contained introns and the spliceosome [Collins and Penny, 2005; Csuros et al., 2011]. The role of alternative splicing in proteome expansion appears to be particularly important in animals, where subsets of regulated alternative exons are strongly enriched for frame preservation, which is not observed in plants or other eukaryotes [Grau-Bové et al., 2018]. Moreover, the frequency of alternative splicing through exon skipping has increased during animal evolution, with the highest skipping frequencies detected in the primate nervous system [Barbosa-Morais et al., 2012; Kim et al., 2007; McGuire et al., 2008]. These increases are likely in part a consequence of changes in the
gene architecture of animal genomes, particularly the increased density of introns, shorter exons and heterogeneous strengths of splice sites \cite{Csuros et al., 2011; Grau-Bové et al., 2018}

Related to these observations, biochemical and cryoEM studies have demonstrated that, compared to the yeast spliceosome, the human spliceosome contains more regulatory components and is also considerably more dynamic \cite{Kastner et al., 2019; Plaschka et al., 2019; Shi, 2017; Yan et al., 2019}

Moreover, the structure of the human post-catalytic spliceosome revealed that it has co-opted additional metazoan-specific RBPs, which are implicated in alternative splicing, to regulate 3’ splice site selection and exon ligation \cite{Fica et al., 2019}. For example, Cactin, SDE2, and NKAP stabilize the exon-ligation conformation of the spliceosome \cite{Fica et al., 2019} and their orthologs are known to regulate alternative splicing in S. pombe \cite{Lorenzi et al., 2015; Thakran et al., 2018} Thus, the human spliceosome likely allows more plasticity and RNA-specific regulation at multiple steps of the splicing reaction.

In 1977 Francois Jacob termed the phrase ‘evolutionary tinkering’ to propose that mutations accumulate gradually in genes while minimizing disruption of existing functions \cite{Jacob, 1977}

Alternative splicing provides a remarkably fertile ground for such tinkering, particularly through evolutionary reassortment of cis-acting elements \cite{Barbosa-Morais et al., 2012; Deveson et al., 2018; Merkin et al., 2012} but also through changes in trans-acting regulators, especially in the IDRs of RBPs. Intronic sequences derived from transposable elements (TEs) are also a particularly common source of cis-acting elements that vary across evolution and can regulate splicing \cite{Gal-Mark et al., 2009; Keren et al., 2010}(Figure 5a). For example, hnRNPC binds to antisense Alu sequences that are prevalent in introns, and thereby represses splicing of thousands of Alu-derived exons \cite{Zarnack et al., 2013}. hnRNPC forms a tetramer that can further assemble into higher-order hnRNP particles \cite{Skoglund et al., 1983; Whitson et al., 2005}. Each tetramer contains four RRM domains in a conformation that is ideally suited for binding to two U-tracts that are present within the antisense Alu sequence (Figure 2c). Binding of hnRNPC to these U-tracts is crucial for it to achieve an affinity required to displace U2AF2, which is essential to repress the use of adjacent cryptic 3’ splice sites \cite{Zarnack et al., 2013}

Similar to Alu elements, the vast majority of long interspersed nuclear elements (LINEs) are evolutionary relics that, with few exceptions, are transcribed as part of host genes, most often within the longest introns. Most LINEs have acquired a vast number of mutations, some of which have created new functional elements, such as enhancers, promoters, or exons \cite{Chuong et al., 2017}. Interestingly, comparisons of mammalian genomes revealed that LINEs bind differently to RBPs depending on how evolutionarily old they are \cite{Attig et al., 2018}. Repressive proteins such as PTBP1 and MATR3 favor evolutionarily young LINE elements, because these are longer and more multivalent. For example, an antisense LINE1 sequence contains many dozens of UC-rich motifs dispersed over hundreds of nucleotides. PTBP1 and MATR3 both recognise UC-rich motifs, and they directly interact, mediated by a region of the IDR in MATR3, which is required to ensure that they cooperate, rather than compete, when binding multivalent RNA binding sites \cite{Attig et al., 2018; Coelho et al., 2015}. MATR3 is required for efficient recruitment of PTBP1 to LINEs, indicating that cooperative binding of the two proteins is particularly important when condensation on long multivalent RNA is required. Notably, the evolutionary young LINEs are depleted from a ±1kb vicinity of constitutive exons, both in mouse and human genes, indicating that the repression of multivalent
RNPs formed on LINEs can spread to the surrounding sequences in RNA. Thus, multivalent RNPs that assemble on intronic RNA derived from TEs can make large portions of the intron inaccessible to other proteins, thereby maintaining the intron in a repressed state (Figure 5a). The flexibility of multivalent assembly provided by IDRs likely enables the repressive RNPs to tolerate changes in the positioning of binding motifs introduced by genetic drift, such that they can maintain introns in a repressed state in spite of genetic variation.

With evolutionary age, the number of repressive multivalent binding motifs in Alus and LINEs decreases, and these changes correlate with decreased binding of repressive RBPs and increased binding of splicing factors such as U2AF2 ([Attig et al., 2016, 2018]) (Figure 5b). Comparison between TEs within and across genomes has demonstrated that exon emergence proceeds through a gradual loss in the multivalency of repressive motifs. In this way, new exons initially emerge as minor transcript isoforms, thus minimizing any potential disruptive effects on gene functions and decreasing the negative selection against such new exons. Thus, multivalent RNPs presumably enable the newly emerging cryptic exons to persist in populations, such that cryptic exons can undergo further evolutionary tinkering through mutations that gradually weaken their repression, thereby enabling the evolution of new transcript isoforms ([Attig and Ule, 2019]). However, due to the large numbers of cryptic exons in our transcriptome, occasionally a mutation strongly induces inclusion of a cryptic exon, which can cause a variety of human diseases ([Sibley et al., 2016]).

Finally, variations in IDRs also commonly diversify splicing regulatory capacities between evolutionary lineages. For example, a new exon that extended an IDR in the SRRM family of splicing factors led to the birth of microexons in bilaterians (Figure 5c). This can be traced to a differential alternative splicing event within the IDR that created the ‘enhancer of microexons’ (eMIC) domain in an ancestral invertebrate Srm ortholog, which originated the neuronal microexon program ([Torres-Méndez et al., 2019]). Gene duplication and specialization of emerging SRRM2-4 orthologs subsequently resulted in further evolution of vertebrate microexon regulation. Moreover, IDRs of other proteins can also be a force for innovation: an alternative exon that encodes a linker region between RRMs in PTBP1 is specifically skipped in mammals, where it reduces the repressive activity of PTBP1 to facilitate activation of a mammalian brain-specific alternative splicing program ([Gueroussov et al., 2015]) (Figure 5d). As mentioned earlier, lineage-dependent alternative splicing events often change the IDRs of hnRNPs to shape their global regulatory properties ([Gueroussov et al., 2017]). Taken together, it is becoming clear that alternative splicing has been particularly important for enriching proteomic complexity in animals in ways that have provided an expanded tool-kit for evolution.

**Function versus noise or evolutionary fodder?**

As the number of alternative splicing events detected in large-scale sequencing studies continues to rise, it has been argued that only a minor fraction of splice variants are regulated, translated or are of functional importance ([Tress et al., 2017]). These arguments stem in large part from computational analyses suggesting that the majority of annotated alternative exons are evolving at neutral rates, and also the poor detection frequency of peptides corresponding to alternative splicing variants in
analyses of shot-gun mass spectrometry data. However, it is well established that such data are severely underrepresented for multiple technical reasons that necessitate a careful modelling of false-negative rates ([Blencowe, 2017; Nilsen and Graveley, 2010]) For example, trypsin-only proteomics is predicted (i.e. based on an in silico analysis) to miss detection of the majority of splice junction-spanning peptides ([Wang et al., 2018]) and MS analysis methods, such as those employed by Tress and colleagues, which required the detection of the same peptide at least twice in two or more independent datasets, are heavily biased against detection of differentially regulated alternative splicing events, even though their functionality is supported by enrichment for frame-preservation and evolutionary conservation ([Brown et al., 2014; Fagnani et al., 2007; Gerstein et al., 2014; Sugnet et al., 2006; Wang et al., 2008]).

Moreover, approximately 75% of human exon-skipping events detected by RNA-seq data in transcripts with medium-to-high abundance were detected in ribosome profiling data ([Weatheritt et al., 2016]) and another recent study has provided evidence for widespread translation of neutrally evolving peptide sequences ([Ruiz-Orera et al., 2018]) As mentioned above, even TE-derived exons that are at the earliest stages of emergence and are poorly conserved might have unforeseen roles under specific physiological states. For example, increased inclusion of TE-derived exons was observed upon UV irradiation in key cell cycle checkpoint regulators, and as most of these exons are out of frame or contain premature stop codons, and thus can prevent translation of full-length proteins or cause nonsense mediated decay of mRNA, they may contribute to cell cycle arrest until the DNA damage process is complete ([Avgan et al., 2019]). Thus, alternative splicing events lie on an evolving spectrum of regulation and functionality, and as such it is very challenging to draw a line between those that are functional or non-functional.

Despite the significant challenges associated with assigning functional roles to individual alternative splicing events, more rapid progress in this area will ultimately come from the development of new functional genomic approaches. For example, CRISPR-based systems have the potential to afford the systematic ablation of alternative splicing events in screens coupled to phenotypic readouts and are expected to provide assessments as to the extent of functionally important splicing variants ([Yuan et al., 2018]). It is envisioned that such screens will initially focus on alternative splicing events in pathway-specific genes where an associated phenotypic readout is known or can be predicted. Such screens will provide a foundation for further in-depth studies and will also facilitate the predictions of phenotypic effects of genetic variation and disease mutations that frequently impact the splicing process ([Climente-González et al., 2017; Jaganathan et al., 2019; Ohno et al., 2018; Park et al., 2018; Sterne-Weiler and Sanford, 2014]).

Conclusions and Future Perspectives.

This review provides a focused account of recent developments in our understanding of the mechanisms, functions and evolution of alternative splicing regulatory networks. As the number of new ‘omics’ technologies continues to grow, generating new types of transcriptomic datasets with which to derive the principles of gene regulation, it is becoming increasingly apparent that underexplored alternative splicing mechanisms await further investigation. For example, an
increasing arsenal of ‘omic’ methods is being used to map different types of RNA modifications comprising the ‘epitranscriptome’ and an important area for future study will be to determine how these may impact alternative splicing, as well as other steps in post-transcriptional gene regulation ([Haussmann et al., 2016; Lence et al., 2016; Pacini and Koziol, 2018; Shi et al., 2019; Zhou et al., 2019]) Similarly, methods for the transcriptome-wide mapping of spliceosomal assembly and RNA-RNA contacts, including both intra- and inter-molecular contacts, will be important for understanding how RNA structures regulate the formation of functional RNP complexes and the availability of splice sites or the contacts between splice sites, which can facilitate either regular splicing or back-splicing during circular RNA (circRNA) biogenesis ([Aw et al., 2016; Briese et al., 2018; Burke et al., 2018; Chen et al., 2018; Kristensen et al., 2019; Lu et al., 2016; Nguyen et al., 2016; Sharma et al., 2016]). Such mapping studies and complementary methods are beginning to uncover new roles for non-coding RNA (ncRNA) products of splicing, such as excised introns and circRNA isoforms ([Kristensen et al., 2019; Morgan et al., 2019; Parenteau et al., 2019]) New roles of trans-acting ncRNAs in splicing mechanisms continue to be unraveled, such as in masking functional sites on RNAs or sequestering splicing factors ([Modic et al., 2019; Romero-Barrios et al., 2018]) These studies are expected to illuminate new roles that autoregulation, crossregulation and feedback networks play in controlling splicing regulatory networks in the context of development and disease ([Jangi and Sharp, 2014]).

For a deeper mechanistic understanding, the application of Cryo-EM and complementary methods will be critically important to further elucidate the structures of splicing regulatory complexes, their roles in the initial definition reactions, as well as in the subsequent early stages of spliceosome formation. Moreover, the increased application of single cell profiling, long-read sequencing of full-length transcripts, as well as methods to study RNP condensates and measure coupled steps in post-transcriptional gene regulation, are expected to further unravel the complexity of alternative splicing and its coordination with transcription and other regulatory processes ([Anvar et al., 2018; Ding and Elowitz, 2019; Fiszbein et al., 2019; Guo et al., 2019; Wan and Larson, 2018]) An important collective goal of these studies will be to derive more sophisticated predictive codes for alternative splicing regulation that will enable the modelling of alternative splicing decisions in a broader range of cell types, and to more accurately predict the impact of normal and disease variation on such decisions ([Siegfried and Karni, 2018]) Such advances will complement emerging functional genomic and machine learning approaches for interrogating the regulation and roles of individual splice variants ([Gonatopoulos-Pournatzis et al., 2018; Narykov et al., 2018]) The resulting information will facilitate further development of strategies for the therapeutic targeting of the splicing process, for example, through the use of small molecules, antisense oligonucleotides, and CRISPR-based approaches for editing cis elements or modulating splicing regulation ([Fong et al., 2019; Konermann et al., 2018; Lee and Abdel-Wahab, 2016; Wang et al., 2019; Yuan et al., 2018])

Finally, we emphasize that the time has come to establish general guidelines for a coordinated, community based effort to systematically annotate and integrate important information on alternative splicing regulation and function into public data resources ([Chakrabarti et al., 2018; Tapial et al., 2017; Van Nostrand et al., 2018]) Such a step, which must involve cooperation with journals to ensure that new data are curated according to rigorous standards as they appear, will help achieve
the ultimate goal of generating an exon- and intron-resolution understanding of gene function and regulation.

Acknowledgements

We thank our lab members, in particular Rasim Barutcu, Ulrich Braunschweig, Hannes Bretschneider, Thomas Gonatopoulos-Pournatzis, Martina Hallegger, Flora Lee, Lorea Blazquez, Anob Chakrabarti, Rupert Faraway, Paulo Gameiro, Hong Han and Jonathan Roth, as well as Juan Valcarcel, Brent Graveley, Manuel Irimia, Kiyoshi Nagai, Chris Smith and Julian König for helpful discussions and critical review of the manuscript. We are grateful to Joe Brock for assistance with the figures. Jernej also thanks Agnieszka and Gustav for inspiration, as the review was drafted in the days after Gustav’s birth. Research in BJB’s group is supported by grants from the Canadian Institutes of Health Research, Simons Foundation, and Canada First Research Excellence Medicine by Design Program; Research in JU’s group is supported by grants from the European Research Council (617837-Translate) and the Wellcome Trust (103760/Z/14/Z). BJB holds the Banbury Chair in Medical Research at the University of Toronto. The Francis Crick Institute receives its core funding from Cancer Research UK (FC001002), the UK Medical Research Council (FC001002), and the Wellcome Trust (FC001002).

Figures: https://www.dropbox.com/s/e0fu3gja8iu1oll/Figures.pdf?dl=0

Figure 1. Exon and intron definition interactions.

a) Schematic of spliceosomal components and regulatory proteins that participate in exon and intron definition, and interactions between them. Trans-acting splicing factors are shown as blue shapes, and their names are shown next to the shape. RS domain is marked by ‘RS’. Blue arrows denote intron or exon definition interactions, many of which are mediated by the RS domain. Exons are represented as grey boxes, intronic RNA and snRNAs are represented as grey lines, cis-acting motifs as colored lines, with the consensus sequences of these motifs shown underneath. The pairing of U1 snRNA with the 5’ splice site is indicated by black lines.

b) Schematic of microexon definition, shown in the same manner as described in Figure 1a.

Figure 2. Multivalent RNP assembly
Intrinsically disordered regions (IDRs) enable multiple RNA-binding domains with similar RNA specificity to recognise multivalent RNA motifs, which can be achieved in three primary ways:

a) Domains with similar sequence specificity are connected by IDRs within a multi-domain RBP, such as PTBP1.

b) Multiple copies of the same RBP can be brought together by multivalent homomeric interactions formed by the IDR, which can promote RNP condensation. Such condensates might be better capable of recognising multivalent RNA sequences.

c) IDRs can connect multiple types of domains to enable homomeric assembly of an RBP into a regular structural unit. For example, hnRNPC contains a single RRM domain, but it is capable of multivalent assembly on sequences containing multiple U-tracts because it forms a tetramer through the action of distinct domains: RRM is connected via IDR to bZLM, a basic region zipper-like motif was proposed to interact in a sequence-independent manner with the RNA that wraps around the tetramer, followed by CLZ, a leucine zipper-like oligomerization domain that forms a very stable a coiled-coil tetramer, and the CTD, a C-terminal domain that was proposed to regulate protein-RNA interactions. (The image is adapted from (Whitson et al., 2005)).

**Figure 3 Alternative splicing regulation mechanisms**

The multivalent RBP schematized here corresponds to Nova proteins, which contain three KH domains and act according to the principles shown in the figure. Many other RBPs have been shown to follow similar principles.

a) Multivalent motifs (marked with red lines) that are intercalated between splice sites and enhancing motifs will recruit repressive RBPs that can compete for pre-mRNA binding with other splicing factors. The RNA and cis-acting motifs are shown as in Figure 1A.

b) If multivalent motifs are located downstream of the exon, many RBPs can bind to support exon definition, most likely by interacting with U1 snRNP. The RNA, RBPs and cis-acting motifs are as in Figure 1A.

c) RBPs can indirectly regulate an alternative exon by enhancing the 5' splice site of a preceding exon, which promotes exon skipping (pathway 1). This likely requires an asymmetric pathway to exon inclusion that relies on initial splicing event being the removal of intron downstream of the alternative exon (pathway 2). In such a case, RBP likely promotes skipping by changing the competition between pathways 1 and 2.

**Figure 4. Recursive splicing and its regulation**
Schematic of the mechanism underlying recursive splicing and its regulation by the exon junction complex (EJC). For recursive splicing to be possible, an ReS-exon needs to be defined to enable splicing of the intron upstream of the ReS-exon (labelled here as ‘intron 1’). This splicing event reconstitutes a new 5’ splice site at the exon-exon junction, and also leads to deposition of the EJC on the preceding exon. If spliceosome assembles on the reconstituted 5’ splice-site, it can lead to ReS-exon removal through recursive splicing (as shown by arrow 1). However, in the case of canonical ReS-exons, EJC tends to efficiently repress recognition of the reconstituted 5’ splice-site, thus promoting ReS-exon inclusion (as shown by arrow 2).

**Figure 5. Emergence and evolution of alternative splicing networks**

a) IDR-mediated interactions can promote assembly of multivalent RNPs that keep introns in a repressed state, while allowing the intron and exon definition interactions (marked in blue lines). Multivalent motifs are indicated as red lines, and blue/green lines represent cryptic splice sites that are efficiently repressed by the multivalent RNPs. Many multivalent RNA sequences are derived from transposable elements (TEs), such as Alus, that often assemble the hnRNPC tetramer, or LINEs, which assemble the PTBP1/MATR3 complex. IDRs often enable interactions that promote RNP condensation, such as homomeric contacts of hnRNPA, or the heteromeric contact between PTBP1 and MATR3. Multi-domain RBPs might be able to further contribute to condensation of intronic RNA by binding to motifs that are present on distal RNA regions, as exemplified here for PTBP1. The RNA and cis-acting motifs are shown as in Figure 1A.

b-d) Evolutionary variation can decrease the repressive environment, thus enabling splicing factors to assemble on the newly-emerging exon. The decrease is usually gradual, leading to a change in the proportion of transcript isoforms, which minimises the disruption of the existing gene functions. This can be achieved through multiple means.

b) Mutations that gradually decrease the multivalency of repressive motifs and/or introduce binding sites for SR proteins.

c) Variation in IDRs, such as the addition of eMIC domain to the IDR of the invertebrate SRRM orthologue rendered it capable of binding upstream of microexons to promote their exon definition [(Torres-Méndez et al., 2019)] This led to further evolutionary variation in cis-acting motifs that shaped the species-specific splicing patterns of microexons.

d) The repressive capacity of PTBP1 can be decreased by a splicing change that shortens the IDR that links the RRM domains, which in turn facilitates activation of a mammalian brain-specific alternative splicing program [(Gueroussov et al., 2015)]
References


Bhate, A., Parker, D.J., Bebee, T.W., Ahn, J., Arif, W., Rashan, E.H., Chorghade, S., Chau, A.,


De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing...
<table>
<thead>
<tr>
<th>Title</th>
<th>Journal/Volume/Publication Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>4, 245–255.e5.</td>
<td></td>
</tr>
<tr>
<td>economies of scale in gene expression.</td>
<td></td>
</tr>
</tbody>
</table>


Förch, P., Puig, O., Martínez, C., Séraphin, B., and Valcárcel, J. (2002). The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5’ splice sites. EMBO J. 21, 6882–6892.


Han, H., Irimia, M., Ross, P.J., Sung, H.-K., Alipanahi, B., David, L., Golipour, A., Gabut, M.,


Ohno, K., Takeda, J.-I., and Masuda, A. (2018). Rules and tools to predict the splicing effects of
exonic and intronic mutations. Wiley Interdiscip. Rev. RNA


Parra, M., Booth, B.W., Weiszmann, R., Yee, B., Yeo, G.W., Brown, J.B., Celniker, S.E., and Conboy, J.G. (2018). An important class of intron retention events in human erythroblasts is regulated by cryptic exons proposed to function as splicing decoys. RNA 24, 1255–1265.


| 521 | 371–375. |


Figure 1

A

trans-acting factors:
SR proteins (eg. SRSF1)  
U1 snRNP  
RS
SF3
U2 snRNP  
RS
eMIC
RS
RS
RS
SF1

cis-acting motifs:
(eg. GAA)  
GGURAG

intron definition interactions
exon definition interactions
exon interactions

B

microexon definition interactions

cis-acting motifs:
YURAY  
Y..YNAG  
GGURAG

U2 snRNP  
U1 snRNP  
SRRM4  
RNPS1  
SF1  
U2AF2  
U2AF1

SF1  
RS
RS
RS
RS
RS
SF1  
RS
RS
RS
Figure 2

A

B

IDR-mediated RNP condensation

C

RNA

bZLM

IDR

CLZ

RRM

CTD
Figure 4

ReS-exon inclusion

ReS-exon

recursive splice site

recursive splicing

RS

intron 1 intron 2

RNPS1 PNN EJC

recursive splice site

recursive splicing

ReS-exon inclusion

ReS-exon
Figure 5

A

gradual derepression of a cryptic exon creates a new transcript isoform

variation in cis-acting motifs

variation in IDR of SRRM2-4 & in cis-acting motifs

variation in IDR of PTBP1

D

PTBP1+E9

IDR encoded by E9

PTBP1ΔE9

SF1

SF1

SF1

SF1

B

C

D

PTBP1+E9

IDR encoded by E9

PTBP1ΔE9

SF1

SF1

SF1

SF1

0% 100%

eMIC

hnRNPA

hnRNPC tetramer

antisense Alu-derived sequence

LINE-derived sequence

RNP condensation interactions

U1 snRNP

SR proteins

U2AF

PTBP1

SF1

SF1

SF1

SF1

variation in IDR of PTBP1