



OPEN ACCESS

EDITED BY

Yadong Zheng,
Zhejiang Agriculture and Forestry
University, China

REVIEWED BY

Hari Krishna Yalamanchili,
Baylor College of Medicine, United States
Bruce McKay,
Carleton University, Canada

*CORRESPONDENCE

Rahul N. Kanadia,
✉ rahul.kanadia@uconn.edu

RECEIVED 23 January 2023

ACCEPTED 28 April 2023

PUBLISHED 16 May 2023

CITATION

Girardini KN, Olthof AM and Kanadia RN
(2023), Introns: the “dark matter” of the
eukaryotic genome.
Front. Genet. 14:1150212.
doi: 10.3389/fgene.2023.1150212

COPYRIGHT

© 2023 Girardini, Olthof and Kanadia. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Introns: the “dark matter” of the eukaryotic genome

Kaitlin N. Girardini¹, Anouk M. Olthof^{1,2} and Rahul N. Kanadia^{1,3*}

¹Physiology and Neurobiology Department, University of Connecticut, Storrs, CT, United States,

²Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark,

³Institute for Systems Genomics, University of Connecticut, Storrs, CT, United States

The emergence of introns was a significant evolutionary leap that is a major distinguishing feature between prokaryotic and eukaryotic genomes. While historically introns were regarded merely as the sequences that are removed to produce spliced transcripts encoding functional products, increasingly data suggests that introns play important roles in the regulation of gene expression. Here, we use an intron-centric lens to review the role of introns in eukaryotic gene expression. First, we focus on intron architecture and how it may influence mechanisms of splicing. Second, we focus on the implications of spliceosomal snRNAs and their variants on intron splicing. Finally, we discuss how the presence of introns and the need to splice them influences transcription regulation. Despite the abundance of introns in the eukaryotic genome and their emerging role regulating gene expression, a lot remains unexplored. Therefore, here we refer to introns as the “dark matter” of the eukaryotic genome and discuss some of the outstanding questions in the field.

KEYWORDS

intron, evolution, splicing, snRNA, spliceosome, eukaryotes, gene expression

Introduction

Historically, introns were considered the non-coding, non-functional sequence elements which disrupt those that are protein-coding, called exons (Gilbert, 1978). While this protein-centric definition of introns (Figure 1, left) has served its purpose, their presence in long non-coding RNA reveals that introns are not specific to protein-coding genes but instead serve a broader role in eukaryotic gene expression (Krchňáková et al., 2019; Abou Alezz et al., 2020). Moreover, introns have been found to host other lariat-derived RNAs, including microRNAs, long noncoding RNAs, small nucleolar RNAs, small nuclear RNAs, and circular RNAs that are crucial for gene regulation (Liu and Maxwell, 1990; Hesselberth, 2013; Seal et al., 2020; Kumari et al., 2022; Vakirlis et al., 2022). Introns can also house enhancer elements that drive tissue-specific expression kinetics during complex vertebrate development and embryogenesis (Emera et al., 2016; Blankvoort et al., 2018; Meng et al., 2021; Shiao et al., 2022). These intervening sequences necessitated co-evolution of splicing machinery to facilitate production of a contiguous transcript capable of encoding a functional unit (Grabowski et al., 1985; Nilsen, 2003). Inhibition of splicing results in retention of introns in the mature transcript, which often disrupts the open reading frame and ultimately dictates the fate of the final transcript (Kaida et al., 2007; Effenberger et al., 2017; Olthof et al., 2021). Since the discovery of splicing, introns have been extensively investigated and the significance of splicing in regulating gene expression is well documented (Singh and Padgett, 2009; Tellier et al., 2020; Zhu et al., 2020; Agirre et al., 2021; Reimer et al., 2021). Taken together, the presence of introns has a significant impact on eukaryotic gene

expression and underpins many of the complexities required to build higher eukaryotes. Therefore, here we present an intron-centric perspective (Figure 1, right) towards understanding regulation of eukaryotic gene expression.

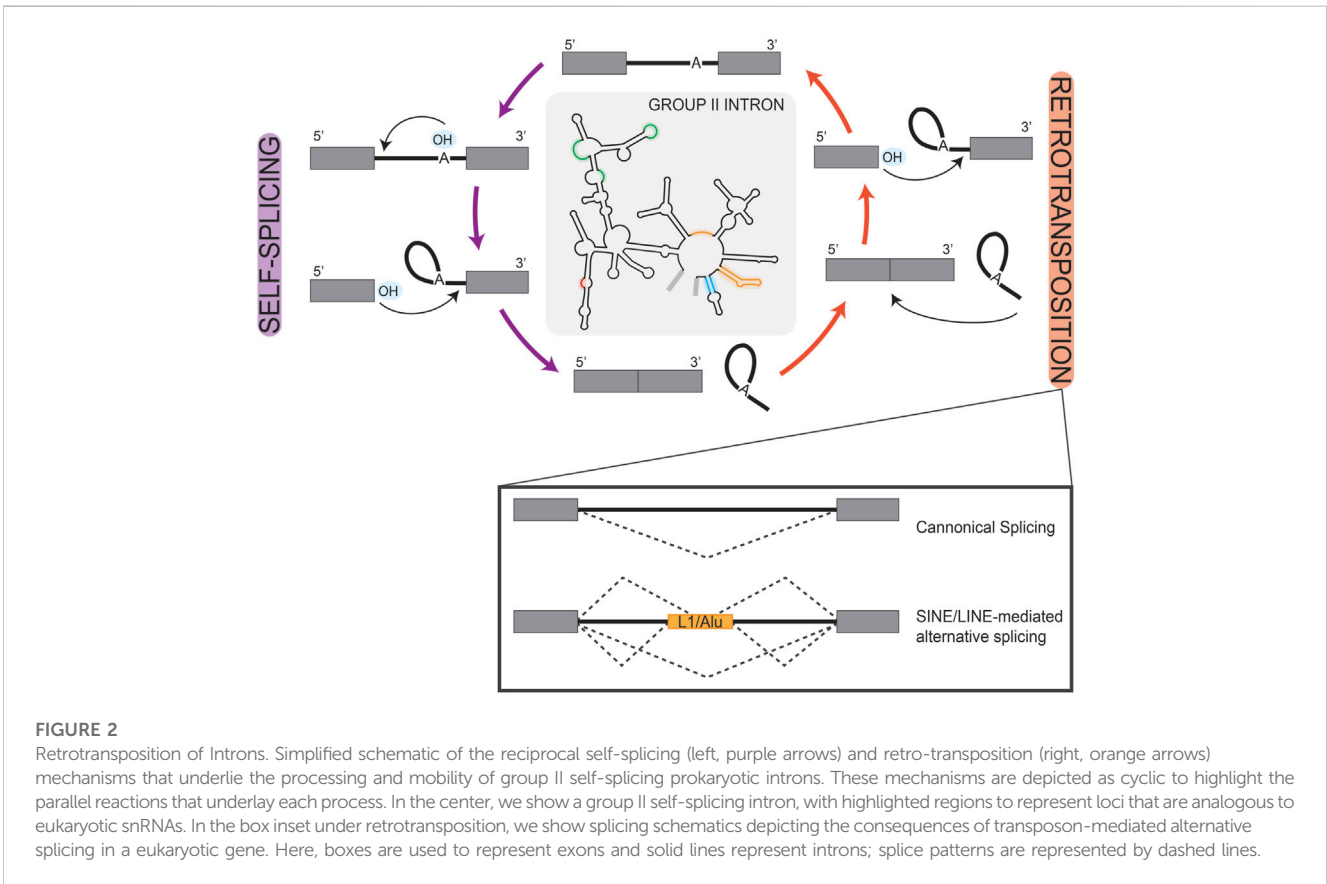
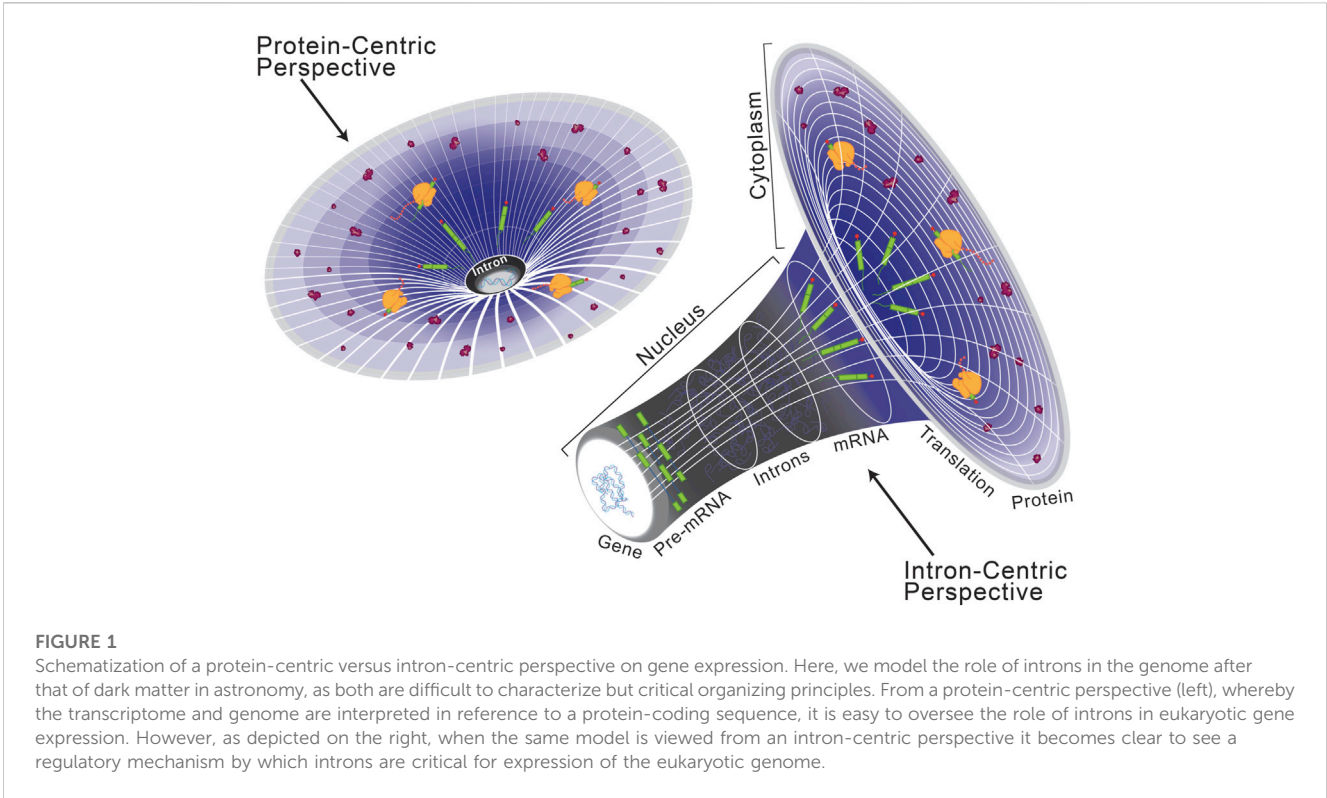
Function and evolution of intronic elements

Introns date back to the last eukaryotic common ancestor, after invasion into the early eukaryotic genome (Russell et al., 2006; Carmel et al., 2007; Csuros et al., 2011). While an endogenous model has been proposed to explain the emergence of introns (Catania et al., 2009), there is a general consensus that prokaryotic group II self-splicing introns underwent invasion and mutational degeneration during early eukaryogenesis, resulting in inert introns and trans-acting splicing machinery (Michel et al., 1989; Sharp, 1991; Sontheimer et al., 1999; Shukla and Padgett, 2002). As the origin of eukaryotic introns has been extensively described (Koonin, 2006; Rogozin et al., 2012; Vosseberg and Snel, 2017; Baumgartner et al., 2019; Smathers and Robart, 2019), here, we focus on the continued maintenance and diversification of introns in eukaryotic genomes.

Prokaryotic group II self-splicing introns behaved largely as transposable elements, which may have facilitated their invasion of the eukaryotic genome (Figure 2) (Lambowitz and Zimmerly, 2011). Initially characterized in the maize genome, transposable elements are repetitive sequences found across eukaryotes and are critically known for their ability to relocate in the genome and alter gene expression (McClintock, 1950; SanMiguel et al., 1996; Elliott et al., 2005; Wells and Feschotte, 2020). Short and long interspersed retro-transposable elements (SINEs/LINEs) belong to the non-long terminal repeat class of elements which have retained transposable activity and are highly represented in the human genome as Alu and L1 elements, respectively (Kazazian and Moran, 1998; Lander et al., 2001; Balachandran et al., 2022). When carrying splice sites, these transposable elements can create novel exon/intron boundaries, which hold the potential to alter expression of that gene (Figure 2); a detailed description of exon/intron boundaries and their recognition by splicing machinery is discussed in the following sections. For example, a recent study queried pathogenic mutations that were associated with novel intron-exon boundaries in humans and identified those which aligned with transposable elements. They found that clusters of transposable elements are more liable to exonization, likely due to the combined effort of LTR and Alu elements in potentiating all necessary splice sites (Alvarez et al., 2021). In another computational investigation of the human genome, mutagenesis of Alu elements into weak splice sites was found to be well-tolerated if not retained long-term and was often associated with exon skipping events (Sorek et al., 2002). Exon skipping is a frequently observed form of alternative splicing, which more broadly serves as an important regulatory node for gene expression in developing systems (Baralle and Giudice, 2017). One can then speculate that Alu elements in this manner allow for transient sampling of novel functions of proteins encoded by these alternatively spliced transcripts. This idea is an extension of the already known role of Alu

elements in tissue-specific transcription regulation (Franchini et al., 2011). Notably, weak splice sites in Alu elements can eventually become constitutively spliced exons, losing their capacity for transposition and become exons used in regulating tissue-specific gene expression, as is observed in the human *NARF* gene (Lev-Maor et al., 2007).

Inherent to the jumping nature of transposable elements is the impartiality of transposon landing. Transposon insertion would likely be deleterious in the protein-coding region of a gene, leading to evolutionary selection against that gene configuration. However, in a heterozygote, transposon-induced activation of a novel splice site within an intron could allow for a low-cost trial of differentially spliced isoforms, while still maintaining a functionally expressed copy. A susceptibility of spliceosomal introns to genomic recombination was demonstrated in two *Saccharomyces cerevisiae* genes, *RPL8B* and *ADH2*. Truncated versions of these genes were used in a splicing reporter construct, such that the second exon was expressed in frame with a fused EGFP cassette. Additionally, each construct carried an embedded *S. pombe his5⁺* gene within the first intron, encoded for in the opposite direction as EGFP. Here, the *his5⁺* gene contains an artificial intron lacking a catalytic branch point, and containing splice sites in such an orientation that they are only capable of splicing from the EGFP transcript. Thus, splicing of the artificial intron followed by transposition of the EGFP intron into the genomic loci was required to confer a positive result (Lee and Stevens, 2016). Meanwhile, Gozashti et al. (2022), has attributed rapid, lineage-specific intron gains to Introner elements derived from transposable elements. Through analysis of 1,700 species, these “intron-generating transposable elements families” were identified in approximately 5% of genomes and significantly overrepresented in aquatic lineages. Based on statistical association models and a consideration of likely propagation mechanisms, they concluded that Introner elements may facilitate recent intron gain, particularly through horizontal gene transfer in aquatic lineages. The activity of Introner elements is particularly interesting, as mechanisms of Introner elements in *Micromonas pusilla* and *Aureococcus anophagefferens* exhibit seemingly preferential insertion between pre-existing nucleosomes (Huff et al., 2016). The rationale here is such that the linker sequence between nucleosomes is often open and available for insertion events. Further support for this idea is seen in the unequal distribution and position of nucleosomes observed between protein-coding exons, pseudo exons, and introns in human and *Caenorhabditis elegans* (Andersson et al., 2009). Using transcriptomic and genomic sequencing data, Huff et al. (2016), reported that Introners are largely capable of co-opting splice sites and inserting by DNA transposition in both orientations, though with biases consistent with species-specific patterns in genome organization. Outside of splice site generation, transposons have also been implicated in regulation of splicing-competent snRNAs, such as those L1 transposons which are associated with formation of U6 pseudogenic snRNA (Doucet et al., 2015). Pseudogenes can encode variations of spliceosomal snRNAs, the implications of which are discussed further below. In all, transposable elements further expand gene structure by modifying intronic elements, thus revealing a critical role of non-coding intronic elements in eukaryotic genome evolution.



Classification and splicing of introns

After the discovery of splicing, identified introns appeared to show a pattern of conserved terminal di-nucleotides at the exon-intron and intron-exon boundaries, and this feature became a defining characteristic of spliced introns (Breathnach et al., 1978; Crick, 1979; Breathnach and Chambon, 1981). As sequencing techniques have progressed and data now includes more diverse eukaryotic genomes, it is increasingly clear that introns are defined by several extended consensus sequences. These include the 5' splice site (5'SS), the branch point sequence (BPS), and the 3' splice site (3'SS) (Dietrich et al., 1997; Mercer et al., 2015). Not long after their discovery, it was determined that most introns are processed by five Uridyl-rich snRNAs—U1, U2, U4, U5, and U6—that are highly conserved between eukaryotes and assemble into a ribonucleoprotein complex, the spliceosome (Bringmann et al., 1983; 1984; Bringmann and Lührmann, 1986; Nilsen, 2003; Wahl et al., 2009). Specifically, U1 snRNA has complementarity at the 5' splice site, marking the exon-intron boundary, while U2 snRNA base pairs around a conserved adenosine toward the 3' end, at what has become known as the branch point sequence (Yan and Ares, 1996; Malca et al., 2003). The direct base pairing of these snRNAs with splice site consensus sequences helps to recognize and remodel the intron during splicing, conferring the core function of the spliceosome.

As this mechanism was coming into focus, Jackson (1991), discovered spliced transcripts, that when mapped to the genome, showed intronic splice site sequences that were incompatible with the identified snRNAs. The fact that these introns were nonetheless spliced suggested the existence of a separate mechanism for their removal. This discordant finding led to sequence-based investigations for U snRNAs with complementary to non-consensus splice sites. This included an exploratory genomics investigation by Hall and Padgett (1994), and ultimately led to the hypothesis that newly identified U11 and U12 snRNAs serve in roles analogous to U1 and U2 during splicing (Montzka and Steitz, 1988). A role for U11 and U12 was confirmed *in vitro* (Tarn and Steitz, 1996a) and *in vivo* (Hall and Padgett, 1996; Kolosova and Padgett, 1997), and bolstered by the additional identification of snRNAs analogous to U4/U6, U4atac and U6atac (Tarn and Steitz, 1996b; Incorvaia and Padgett, 1998). Based on their relative abundance in analyzed genomes, the intron types and their respective spliceosomes were henceforward labeled major (U2-type) and minor (U12-type) in those eukaryotes that maintain them in parallel (Burge et al., 1998; Lynch and Richardson, 2002; Lin et al., 2010). Of note, major introns and the major spliceosome are ubiquitous in the eukaryotic lineage, while minor introns and the minor spliceosome are reportedly absent in some lineages, such as *Caenorhabditis elegans* (Burge et al., 1998).

Both the major and minor spliceosomes employ U5 snRNA, and each snRNA further associates with specific proteins in their splicing-competent forms (Tarn and Steitz, 1996a; Tarn and Steitz, 1997). Though the individual snRNAs have specific proteins associated with their regulation and maturation, many of the remaining proteins that comprise the spliceosome are shared between both the major and minor molecular machineries (Will et al., 1999); for a more comprehensive presentation of individual spliceosome components, see Olthof et al. (2022).

Worth noting, the same protein can carry out different roles in each spliceosome, as is observed by URP (also called ZRSR2) (Tronchère et al., 1997; Shen et al., 2010). While the size and dynamic composition of the spliceosome can make it difficult to fully resolve, identifying the proteins involved in splicing regulation remains an area of active investigation. Recent biochemical and cryogenic electron microscopy investigations to this end have significantly enhanced our understanding of minor spliceosome-specific proteins. For example, the protein compositions of U4.U6/U5 and U4atac.U6atac/U5 tri-snRNP complexes were previously thought to be identical. However, co-immunoprecipitation and co-migration analyses have suggested that CENATAC may aid in 5'SS recognition for a subclass of minor introns characterized by AT-AN terminal di-nucleotides. Previously known as *CCDC84*, CENATAC was renamed following its mutagenic link to intron retention in human genes that contribute to chromosome stability and segregation (de Wolf et al., 2021). Interestingly, phylogenetic profiling of CENATAC across 90 eukaryotic species showed that it co-enriched with other components of the minor spliceosome, including the newly characterized SCNM1 protein (de Wolf et al., 2021). The U12 snRNA is flanked by the N-terminal C₂H₂ zinc fingers of SCNM1, which interacts with the U12/BPS duplex and the U12 Sm ring (Bai et al., 2021). The N-terminus of SCNM1 also functions to stabilize U6atac and RNF113A at the 5'SS, maintenance of which is required for spliceosome activation *in vivo* (Incorvaia and Padgett, 1998; Bai et al., 2021). Structural insights were also important in identifying the novel minor spliceosome protein, RBM48, which is now known to bind ARM7 and interact with terminal ends of U6atac snRNA via conserved RNA binding residues (Bai et al., 2021; Siebert et al., 2022). Structural analyses of the minor spliceosome are a recent advancement and do not yet cover all phases of splicing, notably excluding the U11/U12 di-snRNP. As such, there remains the possibility for other unidentified components regulating the nuances of minor intron splicing.

A delineation between major versus minor intron splicing is often based on the quantitative analysis of splice site conservation, and thus relative splice site strength. Intron splice sites are generally scored based on the degree of similarity to the major versus minor intron consensus sequences found in Figure 3, using position weight matrices (Sheth et al., 2006; Alioto, 2007; Olthof et al., 2019; Moyer et al., 2020). The resultant major or minor intron classification inherently dictates how we interpret its processing, such that bioinformatically classified minor introns are predicted to be spliced by the minor spliceosome, and *vice versa*. However, RNA sequencing data has shown that, upon inhibition of the minor spliceosome, not all bioinformatically classified minor introns show a splicing defect (Olthof et al., 2019). Thus, parallel existence of major and minor spliceosomes, combined with diverging intron consensus sequences, reveal an added complexity in the relationship between a given intron and its recruited spliceosome. Akin to how the concept of a single intron type was disrupted by the discovery of minor introns; it seems increasingly likely that the binary classification of major versus minor itself is insufficient to fully resolve all introns. Rather, evidence has begun to suggest that the stringency of the classification schema fails to consider the fluidity of exons and introns. For example, use of novel splice sites within exonic regions in the unicellular *Paramecium* is evidence of intronization activity in

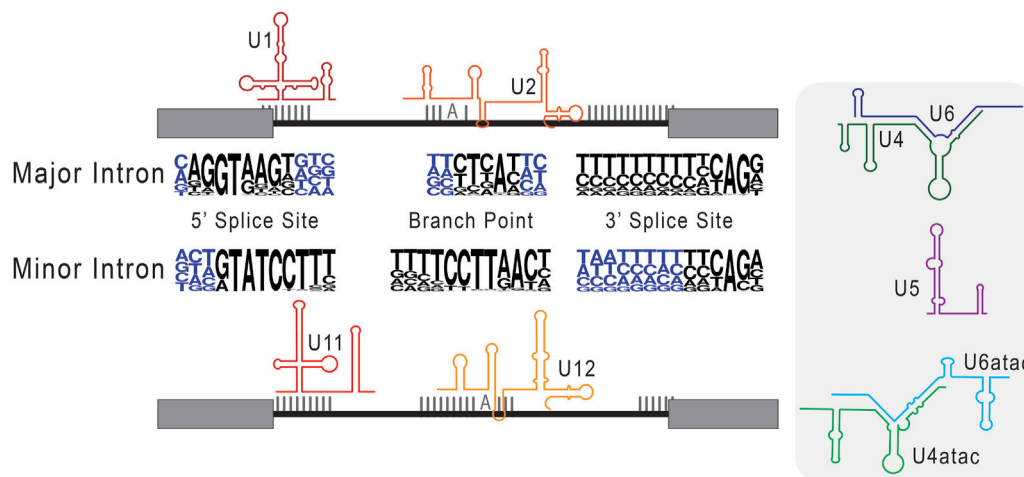


FIGURE 3

Consensus sequences used in the classification of major versus minor introns. Here, we schematize splice site selection by the respective components of the major and minor spliceosomes. The snRNAs of the major (U1 and U2) and minor (U11 and U12) spliceosome are shown base pairing to their cognate consensus sequences. In the center, next to the respective major intron and minor intron labels, we depict consensus sequences as nucleotide frequency plots. Here, the relative size of the nucleotide represents how frequently it is observed in that genomic position. Right of this schematic, we include the remaining core snRNAs that are unique to major (U4 and U6) and minor (U4atac and U6atac) intron splicing, as well as the shared U5 snRNA.

eukaryotes (Ryll et al., 2022). In essence, these findings increasingly suggest that the current approach to intron classification is too reductive to fully capture the complexities and dynamic regulation of eukaryotic introns. Towards this end, an examination of minor-type splice sites in *Physarum polycephalum* has suggested that minor introns may exist in divergent, if not degenerative, types (Larue et al., 2021) and this idea is currently being refined in other studies that combine principles of speciation and comparative genomics.

How gene architecture informs splice site selection

Spliceosomal introns are known to range from tens of base pairs in length to hundreds of kilobases in length, with a mean length that is smaller in lower eukaryotes and larger in higher eukaryotes (Sakharkar et al., 2004; Piovesan et al., 2015; Abebrese et al., 2017; Li et al., 2017; Jakt et al., 2022). The size of an intron has an inherent impact on gene expression, as it will take longer for transcription machinery to create nascent transcripts. In turn, this will impact the kinetics of co-transcriptional intron splicing; these ideas have been explored in depth (Herzel et al., 2017; Wallace and Beggs, 2017; Neugebauer, 2019). It is long since established that relative intron and exon lengths can differentially affect splicing efficiency due to a presence or absence of regulatory elements and differing requirements for catalysis (Fox-Walsh et al., 2005; Kandul and Noor, 2009; Pai et al., 2017). Splicing efficiency refers to the proportion of spliced versus un-spliced transcripts relative to the number of total transcripts. This is commonly assessed using computational strategies that characterize splice events in the transcriptome (de Melo Costa et al., 2021; Jiang et al., 2023), followed by a validation of observed changes in expression using

techniques such as RT-PCR. In one assessment of how splicing efficiency and gene expression patterns may be coupled, intron length was found to contribute to the temporal coordination that is required for co-expression of genes with interdependent biochemical functions (Keane and Seoighe, 2016). This idea is further reflected by distinct differences in splice site strength relative to intron length, and by differences in splicing efficiency and mRNA abundance relative to gene length (Gelfman et al., 2011; Sánchez-Escabias et al., 2022). Vertebrates are known to increase splicing efficiency around longer introns via cell-specific recursive splicing and transposable elements that form stems with intronic RNA loops to juxtapose splice sites (Shepard et al., 2009; Zhang et al., 2018). For details on recursive splicing, please see published reviews (Georgomanolis et al., 2016; Gehring and Roignant, 2021; Joseph et al., 2022; Pitolli et al., 2022).

Separate from this, longer introns may also have a propensity to contain multiple splice sites within one intronic feature, leading to alternative splicing from competing splice site use (Sun and Chasin, 2000; Roca et al., 2003; Kapustin et al., 2011). Meaning it becomes increasingly likely that multiple splice sites be present, in addition to exonic splicing enhancers and silencer elements, which themselves can act as determinants of splice site usage (Black, 2003; Wang et al., 2006). It thus follows that the sequence content of the intron to be excised can drive splicing progression. Splice site selection is thought to occur by competing intron- and exon-definition models, which describe how the spliceosome assembles either through cross-bridging interactions across the intron itself or across the flanking exon. Specifically, the intron-definition model refers to the mechanism whereby 3' SS selection is informed by recognition of the upstream 5' SS, such that the spliceosome assembles across the intron. For exon-definition interactions, 3' SS recognition depends instead on recognition of the downstream 5' SS (Robberson et al., 1990; Berget, 1995; Romfo et al., 2000; De Conti et al., 2013; Olthof

et al., 2021). For example, most genes in *Saccharomyces cerevisiae*, contain only one, short intron. With this gene architecture, it is not surprising that intron-definition interactions predominate. Surprisingly, cryo-electron microscopy structures of the pre-catalytic spliceosome demonstrated that the same splicing machinery can perform exon-definition interactions in multi-intronic genes (Li et al., 2019). This finding brings to bear uncertainty as to how and when an intron- versus exon-centric model is utilized. This becomes especially important in higher vertebrates which have a larger intronic burden.

Reconciliation between the intron- and exon-definition models is coupled with new insight into how proximity rules inform splice site selection. Based on the length of an intron, the intron-centric proximity rule dictates a preference for the spliceosome to assemble over a splice site pair that minimizes the distance between 5' and 3' end selection (Reed and Maniatis, 1986). More recently, computational analyses by Carranza et al. (2022) refined the exon-centric proximity rule, by which splice sites are selected to minimize the exon-spanning distance. Meaning if one were to imagine an intron with two adjacent sets of 5' and 3' splice sites, the intron-centric proximity rule would employ the innermost splice site pair, maximizing the resultant exons. Meanwhile, the exon-centric proximity rule would, in contrast, use the exon-proximal splice sites to maximize the size of the intron being excised. In either case, commitment to the intron-centric or exon-centric proximity rule has commensurate intronization/exonization consequences as molecular machinery decides whether to select for the smaller or larger exonic sequences. In addition to intron size, studies suggest that GC content of the intron may also be a determinative factor in the mechanism employed for splice site selection. In one study, (Tammer et al. (2022), examined the nucleotide composition of exons versus introns and subsequently identified genes they refer to as “differential” or “leveled”. In “leveled” genes, GC content is found to be similarly high in exons and introns, while “differential” genes are ones wherein GC content is low in exons, and even lower in introns. Notably, Tammer et al. (2022), describe a partiality for intron-definition interactions across “leveled” genes, while exon-definition interactions predominate over “differential” genes. This finding is in line with previously reported links between differential GC content and splice site selection (Amit et al., 2012).

Spliceosomal snRNAs

As described above, snRNAs confer the primary function of the spliceosome through formation of specific base pair interactions with consensus sequences in the intron. The presence and function of snRNAs is essential for recognition and restructuring of the nascent mRNA transcript in the sequential, exothermic transesterification reactions that constitute splicing.

In general, most snRNAs (U1, U2, U4, U5, U11, and U12) are transcribed by RNA polymerase II, while U6 and U6atac expression are largely dependent on RNA polymerase III (Reddy et al., 1987; Jawdekar and Henry, 2008; Younis et al., 2013). Initiation of transcription of these snRNAs is highly reliant on the proximal and distal sequence elements located upstream of the snRNA-encoding region. Specifically, because they serve as promoter and enhancer elements for recruitment of transcription machinery

through interactions with the SNAPc transcription factor complex and stabilizing co-activators (Sadowski et al., 1993; Henry et al., 1998; Mittal et al., 1999; Dergai et al., 2018). Structural insights by cryogenic electron microscopy of SNAPc during the transcription of U6 has revealed the importance of conserved subunits which recognize and bind the proximal sequence element (Sun et al., 2022). One unique exception to this rule is for the expression of human U4atac snRNA, which is embedded into an intron of *CLASPI* (Ederly et al., 2011). Therefore, U4atac expression relies on RNA polymerase II mediated transcription of this gene, as well as successful splicing of this intron.

Within the genome, spliceosomal snRNAs often exist both as gene copies and gene families, whereby divergent genes can encode for variant snRNAs with nucleotide polymorphisms (Denison et al., 1981; Abel et al., 1989). There are both productive and unproductive variants of the snRNAs annotated; productive snRNAs are capable of splicing, while those that are not are termed pseudogenic (Mabin et al., 2021). For example, the U6 snRNA has many pseudogenes and fewer productive copies that are dispersed throughout the genome, whereas U1 and U2 snRNAs are encoded by many functional copies that are organized in homogenous repeats (Van Arsdell and Weiner, 1984; Theissen et al., 1985; Tichelaar et al., 1998; Domitrovich and Kunkel, 2003; O'Reilly et al., 2013; Anjos et al., 2015). The presence of multiple gene copies may in part explain the splicing-independent roles of U1 and U2 in regulating transcription termination and 3' end processing (Friend et al., 2007; Di et al., 2019; So et al., 2019). Moreover, the idea that multiple gene copies exist for minor spliceosomal snRNAs, including U4atac and U6atac, warrants further investigation. Even if multiple gene copies do exist, it must be noted that U6atac expression is maintained at a lower level through rapid post-transcriptional turnover (Younis et al., 2013).

Perhaps counterintuitively, U5 snRNA has the smallest gene family, yet it is the only shared snRNA between the major and minor spliceosomes. Investigations by Mabin et al. (2021) into the relevance of snRNA variants in splicing led to the discovery of high sequence identity between U5 variants. In fact, they report several U5 variants with a conserved stem consensus sequence (CUUUU) that can be incorporated into catalytic spliceosomes. Based on these observations, it has been suggested that U5 may not have a canonical snRNA; rather, specific variants may be optimal for use in one spliceosome type over the other (Mabin et al., 2021). While mechanistically unvalidated, this logic is consistent with the analogous nature of the other major versus minor snRNAs. Yet, it also remains possible that these U5 variants are regulated in a context-dependent way, as is observed for U1 snRNA variants during human stem cell programming (Vazquez-Arango et al., 2016). Additionally, U5 snRNA variants have been identified in regulating development in humans, *Drosophila*, and *Lytechinus variegatus* (Sontheimer and Steitz, 1992; Morales et al., 1997; Chen et al., 2005). The expression of snRNA variants to specify a differentiating transcriptome is not unique to U5 snRNA, but more broadly detected for other snRNAs and across species (Lo and Mount, 1990; Cáceres et al., 1992; Sierra-Montes et al., 2005; O'Reilly et al., 2013; Lu and Matera, 2014).

Functional sequence variants of the snRNAs have the potential to contact cryptic or degenerating splice sites, make novel protein

interactions, and adopt secondary structures that alter spliceosome conformation. It is thus possible, given our evolving understanding of consensus sequences, that these variant snRNAs do confer complementarity to specific intron splice sites. Accordingly, from an intron-centric perspective, we must allow for the possibility that seemingly unproductive snRNAs are leveraged to splice a specific subset of introns. A role for non-consensus intron classes was voiced by Hudson et al. (2019), whose bioinformatics analyses of diplomonad and parabasalid lineage eukaryotes revealed splice site sequences that diverged from both the major and minor consensus sequences. They similarly identify divergent snRNAs, though they maintained key functional structures including stem loops and putative Sm binding sites. Perhaps more compelling, the discovered snRNAs showed aggregate features of both the major- and minor-type snRNAs, suggesting a propensity for the spliceosome to adopt complementarity to *trans*-spliced introns.

It remains to be established if variant snRNAs are evolutionarily selected for use in differential splicing or if they arise stochastically. Though, one could imagine that selective use of a variant splicing component would provide an opportunity to splice novel or divergent splice site sequences. It is known that mutations in the snRNAs can have pathogenic effects, as demonstrated by *RNU12* which is causal to early onset cerebellar ataxia (Elsaid et al., 2017). Additionally, snRNA secondary structure is important for splicing as it dictates the RNA-protein interactions necessary for spliceosome assembly. For example, U11/U12-65K binds the 3' stem loop II (SLII) of U12 snRNA based on the integrity of this structure and its RNA binding motif. Further, 3' truncation mutants that disrupt the U12 SLIII are targeted for degradation by the nuclear exosome targeting complex upon reimport to the nucleus (Norppa and Frilander, 2021). In another example, the U2/U6 and U12/U6atac complexes are remodeled and stabilized prior to the first catalytic step in splicing by intramolecular base pairing with RBM22 (Ciavarella et al., 2020). Regardless, developmentally regulated snRNA variants demonstrate that mutations outside of critical structures may maintain, albeit differential, functionality. Thus, it stands to reason that variant snRNAs without disease-causing consequences to splicing may have a context-dependent role in the regulation of introns with divergent consensus sequences.

The evolutionary advantages of introns

Introns have served a valuable evolutionary role for eukaryotes in that they are more prone to genetic drift compared to exons. Introns appear to be under weaker selection than exons in somatic cells, which may be due to a mismatch repair system employed for exons that is notably lacking for introns (Hoffman and Birney, 2006; Resch et al., 2007; Frigola et al., 2017; Rodriguez-Galindo et al., 2020). Using a combinatorial multi-omics approach, Huang et al. (2018), has attributed the selective protection and mismatch repair of actively transcribed genes to an enrichment of H3K36me markers, which help regulate molecular responses to DNA damages induced by prolonged euchromatic conformation. More broad analyses of the differentiating human methylome reveal distinct differences in methylation pattern between genomics features, such that methylation is generally more

common to exons than introns (Laurent et al., 2010). This unequal distribution may explain the higher frequency of mismatch repair observed for exons versus introns. In this capacity, introns can essentially act as a sponge to harbor mutations that would be otherwise detrimental in exonic sequences. Nevertheless, many mutations in intronic elements are linked to diseases, suggesting that there are limits to the number of mutations an intron can withstand. Mutations at splice sites and within introns are known to underscore an array of genetic and developmental disorders, including muscular dystrophy (Dominov et al., 2019) and inherited retinal diseases (Qian et al., 2021). Pathogenic disorders due to mutation of the spliceosome, i.e., spliceosomopathies, include but are not limited to craniofacial defects, myelodysplastic syndromes, and retinitis pigmentosa (Griffin and Saint-Jeannet, 2020). For review of major and minor splicing-associated diseases, see (Anna and Monika, 2018; Olthof et al., 2022).

While introns are seemingly advantageous, prokaryotes show that the absence of introns is not prohibitive to life. This begs the question, to what extent do eukaryotic cells really require introns? In one study, Parenteau et al. (2008), investigated the consequences of intron depletion in *Saccharomyces cerevisiae* (Figure 4A). Introns are far less abundant in *S. cerevisiae* compared to other species, such as vertebrates and land plants, making the yeast genome a strong model for intron depletion studies (Csuros et al., 2011). Indeed, *S. cerevisiae* could survive without introns, however, intron-depleted strains fared variably when subjected to drug-induced and carbon source stresses. However, transcription machinery was found to be capable of responding to expression deficits following intron-depletion by using alternative promoter selections, highlighting the role introns play in expanding the eukaryotic transcriptome (Parenteau et al., 2008). Should one suppose that introns can be leveraged to induce stress-related patterns of gene expression, it then follows that the splicing efficiency of an intron is responsive to stress application. This idea was recently explored by Frumkin et al. (2019), who employed YFP reporter constructs containing known introns with high and low splicing efficiencies embedded and fused to a kanamycin resistance cassette (Figure 4B). To test the capacity of introns and the spliceosome to respond to metabolic pressure, the constructs were expressed in *S. cerevisiae* cells under antibiotic selection and subjected to a lab-evolution paradigm. Growth and transcriptomic analyses of derived cell generations revealed independent, adaptive mutations occurring both *cis*- and *trans*-to improve splicing efficiency and thus antibiotic resistance and cell survivability. The *cis*-mutations were proposed to increase accessibility of splice site sequences, while *trans*-mutations might increase the cellular abundance of splicing machinery. Importantly, *cis*-fitness-inducing mutations could alleviate selection-independent splicing inefficiencies, however, mutations in *trans*-were particularly advantageous during periods of active selection (Frumkin et al., 2019). Though these experiments were performed in *S. cerevisiae*, one can imagine that similar mechanisms may be employed for evolutionary adaptation. For example, in ecotypic Cichlid fish, alternative splicing is a dominant mechanism for rapid changes in gene expression. Specifically, alternative splicing underpins the diversification of jaw morphology as it relates to the food they have evolved to consume in different ecological niches (Singh et al., 2017).

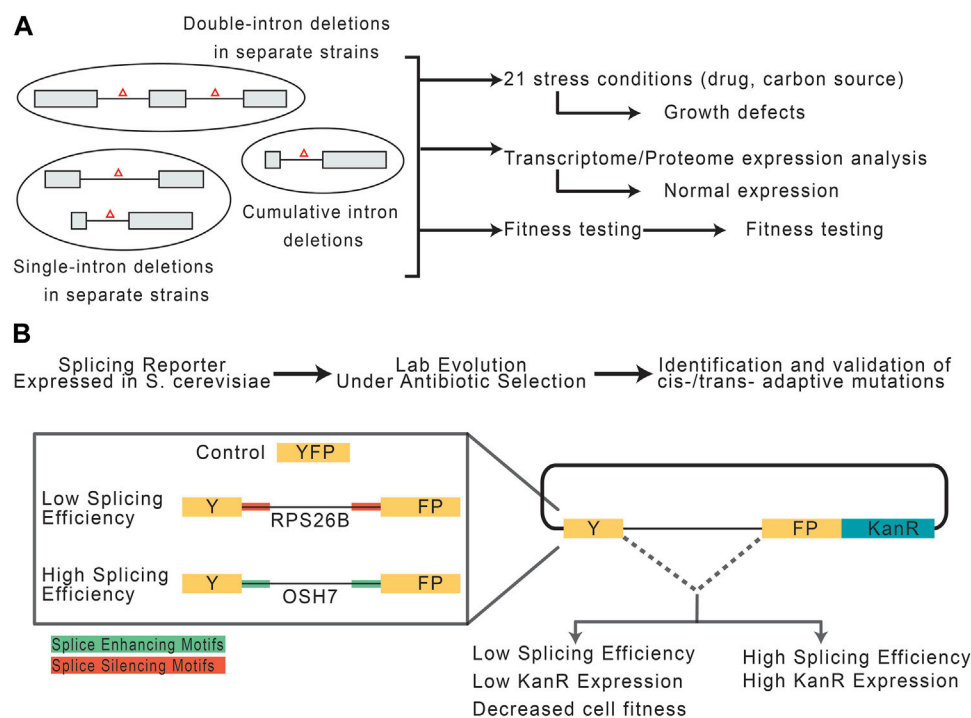


FIGURE 4

Potential role of introns and spliceosomal snRNAs in stress response. (A) Experimental paradigm, adapted from Parenteau et al. (2008), to assess the consequences of intron depletion in *S. cerevisiae*. Yeast with sets of removed intron(s) were grown under normal or stress conditions and assessed for fitness. (B) Experimental paradigm, adapted from Frumkin et al. (2019), to assess the capacity of introns and the splicing machinery to adapt to selective pressures.

The influence of introns on gene expression

In both mammals and plants, the presence of introns is known to enhance gene expression in a phenomenon sometimes referred to as intron-mediated enhancement (Brinster et al., 1988; Furger et al., 2002; Samadder et al., 2008). The recent development of sequencing techniques such as GRO-seq, mNET-seq and long read sequencing have revealed that splicing of neither major nor minor introns occurs in isolation, but rather in a highly active genomic context where splicing and transcription are coupled both kinetically and physically (Nojima et al., 2015; Herzel et al., 2017; Sheridan et al., 2019; Drexler et al., 2020; Reimer et al., 2021; Zhang et al., 2021). In the context of splicing informing transcription, the position of the intron matters, as promoter-proximal introns are especially known to enhance transcription (Furger et al., 2002; Rose et al., 2008). The knowledge that introns may enhance transcriptional output was leveraged to modify the generally used CMV promoter for expression plasmids, whereby introduction of an intron significantly upregulated transcription of downstream coding sequences (Simari et al., 1998).

The mechanism by which 5' introns regulate transcription involves, at least in part, control of the open chromatin signatures H3K4me3 and H3K9ac, which facilitate recruitment of RNA polymerase II and general transcription factors to promoters. These marks are deposited at the first exon-intron boundary of genes, explaining how the distance between transcription start site and the first intron can influence the expression level of a gene (Bieberstein et al., 2012; Lister, 2009). Interestingly, differential

methylation patterns are not unique to protein-coding genes, as revealed through a bioinformatics model which considered the modified human nucleosome library and analysis of splicing efficiency. For example, high nucleosome density was observed in the internal exons of long non-coding RNAs, while high H3K4me3 signals were observed in upstream introns. Importantly, these signatures were often associated with exon skipping and intron retention, particularly around the first intron (Dey and Mattick, 2021). While a tissue-independent model likely obscures some of the nuanced features regulating splicing-dependent gene expression, a genome-wide comparative analysis by Anastasiadi et al. (2018) revealed that correlation between CpG methylation and gene expression is unique to the first exon and intron. As CpG markers of DNA methylation tend to decrease across exons and increase across introns, it is possible that methylation may inform gene expression by mediating intron splice site recognition (Laurent et al., 2010). In fact, removal of promoter-proximal introns altogether reduces levels of H3K4me3 and chromatin-bound RNA polymerase II, reducing transcriptional output (Bieberstein et al., 2012; Laxa et al., 2016). Similarly, reduction in chromatin accessibility was observed when formation of the active spliceosome was inhibited by spliceostatin A. This finding highlights an important role for the spliceosome in regulating transcriptional output. Notably, this effect was not intrinsic to the presence of introns, but dependent on their splicing (Bieberstein et al., 2012).

One caveat to the spliceostatin A experiment is that it inhibits the entire splicing machinery, without revealing the specific interactions between the spliceosome and intron consensus sequences that enhance transcription. In fact, it is not the entire spliceosome that needs to be activated for transcription enhancement, as the formation of stable interactions between U1 snRNA and the promoter-proximal 5'SS can enhance transcription (Engreitz et al., 2014). Recruitment of the U1 snRNP to the first intron enhances transcription initiation through recruitment of general transcription factors, such as TFIID, and stabilization of the first formed phosphodiester bond by RNA polymerase II (Kwek et al., 2002; Damgaard et al., 2008). Notably, this effect is independent of its role in major intron splicing, as mere introduction of a 5'SS sequence is sufficient to enhance transcription (Damgaard et al., 2008). This splicing-independent function of U1 might help explain its constitutive association with the elongating RNA polymerase II and why it is likewise recruited to intronless genes (Spiluttini et al., 2010; Leader et al., 2021).

Beside the role of U1 in transcription initiation, U1 snRNA is also independently involved in preventing pre-mature transcription termination, which can occur if RNA polymerase II encounters a polyadenylation site within an intron. Surmounting 3' end sequencing data has revealed that introns often contain cryptic or pre-mature polyadenylation sites that result in the destabilization of RNA polymerase II, thereby producing truncated transcripts incapable of encoding a protein (Di Giammartino et al., 2011). Remarkably, the production of these truncated transcripts can be blocked by the U1 snRNA in a process called telescripting. In this capacity, U1 is capable of complexing with 3' processing factors to protect the mRNA from premature cleavage and termination (Kaida et al., 2010; Berg et al., 2012). This mechanism occurs alongside the elongating polymerase to allow for U1-mediated suppression of cryptic polyadenylation sites in the intron or 3' UTR (Di et al., 2019). Proper transcription termination is important in regulating the length and structure of the 3' UTR, which in turn promotes formation of the export-competent messenger ribonucleoprotein. Similar to U1, U11 is expressed more highly than is necessary for its function in splicing (Baumgartner et al., 2015). Given that U11 is more abundant than U12 though they present at the same stoichiometric ratio within the minor spliceosome, U11 may similarly have splicing-independent functions. We speculate that U11 may either function in a mechanism like telescripting or participate in an alternative function, such as the subnuclear clustering of expressed minor intron-containing genes.

Localization of spliceosome components

Genes, chromatin, and RNA polymerase II have a subnuclear organization around topologically-associated domains to phase-separate euchromatic regions of active transcription (Szentirmay and Sawadogo, 2000; Ulianov et al., 2016; Szabo et al., 2020). Alongside this, it would be reasonable to hypothesize that splicing machinery is also organized to support efficient gene expression. In fact, major and minor spliceosome snRNPs display similar partiality for nuclear localization, except for U6 and U6atac snRNPs (Spiller et al., 2007; Pessa et al., 2008; Steitz et al., 2008). In the nucleus, matured snRNPs of the major spliceosome accumulate in phase-separated

speckles that serve to organize spliceosome components adjacent to perichromatin regions of active transcription. This was concluded following nonradioactive and fluorescence *in situ* hybridization analyses, as well as RNA and protein blotting of subcellular compartment extracts (Pessa et al., 2008). While this model is an enticing way to interpret speckles as a regulatory mechanism over major intron splicing, it does not necessarily extend to that of minor introns. Given that the major and minor spliceosomes are known to interact with each other in the splicing of minor intron-containing genes, the model does not encompass all mechanisms of splicing (Akinyi and Frilander, 2021; Olthof et al., 2021). Punctate subcellular localization of spliceosome machinery is not specific to core snRNP components, but also includes some of the auxiliary splicing factors that contribute to spliceosome stability, conformational changes, and catalytic activity during splicing. These non-snRNP factors are integral to spliceosome assembly and the coordinated action of snRNPs during splicing (Bindereif and Green, 1990). For example, a new model supposes that the unequal phasic separation of SR proteins and heterogeneous nuclear ribonucleoproteins (hnRNP) at nuclear speckles can contribute to splice site selection. Specifically, the positional distribution of SR proteins and hnRNPs around a splice site generally determines the positive or negative regulatory effect of their binding, and taken with their distinct subnuclear distributions, can dictate the use of splice sites (Liao and Regev, 2021).

In all, here through an intron-centric lens, we focus our attention on the myriad of regulatory and functional consequences that have emerged by the presence of introns in the genome. Thus, we hope that future studies will begin to shed light on this “dark matter” of the eukaryotic genome to uncover the secrets buried within. Importantly, the advent of next-generation sequencing and computational analysis will invariably play a critical role in uncovering some of these mysteries. Throughout this article, we have described several of these methods, and here we point readers to other reviews (Halperin et al., 2021; Lorenzi et al., 2021; Gondane and Itkonen, 2023).

Author contributions

KG was responsible for curation of literature, organizing, writing text, and generating figures. AO was responsible for editing, structural organization, and help with literature curation. RK was responsible for the vision, writing, figures, and structure of the document. All authors contributed to the article and approved the submitted version.

Funding

Funding for this study comes from the National Institute of Neurological Disorders and Stroke (R01NS102538 to RK).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

- Abebrese, E. L., Ali, S. H., Arnold, Z. R., Andrews, V. M., Armstrong, K., Burns, L., et al. (2017). Identification of human short introns. *PLoS ONE* 12, e0175393. doi:10.1371/journal.pone.0175393
- Abel, S., Kiss, T., and Solymosy, F. (1989). Molecular analysis of eight U1 RNA gene candidates from tomato that could potentially be transcribed into U1 RNA sequence variants differing from each other in similar regions of secondary structure. *Nucleic Acids Res.* 17, 6319–6337. doi:10.1093/nar/17.15.6319
- Abou Alezz, M., Celli, L., Belotti, G., Lisa, A., and Bione, S. (2020). GC-AG introns features in long non-coding and protein-coding genes suggest their role in gene expression regulation. *Front. Genet.* 11. doi:10.3389/fgene.2020.00488
- Agirre, E., Oldfield, A. J., Bellora, N., Segelle, A., and Luco, R. F. (2021). Splicing-associated chromatin signatures: A combinatorial and position-dependent role for histone marks in splicing definition. *Nat. Commun.* 12, 682. doi:10.1038/s41467-021-20979-x
- Akinyi, M. V., and Frilander, M. J. (2021). At the intersection of major and minor spliceosomes: Crosstalk mechanisms and their impact on gene expression. *Front. Genet.* 12, 700744. doi:10.3389/fgene.2021.700744
- Alioto, T. S. (2007). U12DB: A database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res.* 35, D110–D115. doi:10.1093/nar/gkl796
- Alvarez, M. E. V., Chivers, M., Borovska, I., Monger, S., Giannoulou, E., Kralovicova, J., et al. (2021). Transposon clusters as substrates for aberrant splice-site activation. *RNA Biol.* 18, 354–367. doi:10.1080/15476286.2020.1805909
- Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., et al. (2012). Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep.* 1, 543–556. doi:10.1016/j.celrep.2012.03.013
- Anastasiadi, D., Esteve-Codina, A., and Piferrer, F. (2018). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* 11, 37. doi:10.1186/s13072-018-0205-1
- Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., and Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* 19, 1732–1741. doi:10.1101/gr.092353.109
- Anjos, A., Ruiz-Ruano, F. J., Camacho, J. P. M., Loreto, V., Cabrero, J., de Souza, M. J., et al. (2015). U1 snDNA clusters in grasshoppers: Chromosomal dynamics and genomic organization. *Heredity* 114, 207–219. doi:10.1038/hdy.2014.87
- Anna, A., and Monika, G. (2018). Splicing mutations in human genetic disorders: Examples, detection, and confirmation. *J. Appl. Genet.* 59, 253–268. doi:10.1007/s13353-018-0444-7
- Bai, R., Wan, R., Wang, L., Xu, K., Zhang, Q., Lei, J., et al. (2021). Structure of the activated human minor spliceosome. *Science* 371, eabg0879. doi:10.1126/science.abg0879
- Balachandran, P., Walawalkar, I. A., Flores, J. I., Dayton, J. N., Audano, P. A., and Beck, C. R. (2022). Transposable element-mediated rearrangements are prevalent in human genomes. *Nat. Commun.* 13, 7115. doi:10.1038/s41467-022-34810-8
- Baralle, F. E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi:10.1038/nrm.2017.27
- Baumgartner, M., Drake, K., and Kanadia, R. N. (2019). An integrated model of minor intron emergence and conservation. *Front. Genet.* 10, 1113. doi:10.3389/fgene.2019.01113
- Baumgartner, M., Lemoine, C., Al Seesi, S., Karunakaran, D. K. P., Sturrock, N., Banday, A. R., et al. (2015). Minor splicing snRNAs are enriched in the developing mouse CNS and are crucial for survival of differentiating retinal neurons. *Dev. Neurobiol.* 75, 895–907. doi:10.1002/dneu.22257
- Berg, M. G., Singh, L. N., Younis, I., Liu, Q., Pinto, A. M., Kaida, D., et al. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64. doi:10.1016/j.cell.2012.05.029
- Berget, S. M. (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270, 2411–2414. doi:10.1074/jbc.270.6.2411
- Bieberstein, N. I., Carrillo Oesterreich, F., Straube, K., and Neugebauer, K. M. (2012). First exon length controls active chromatin signatures and transcription. *Cell Rep.* 2, 62–68. doi:10.1016/j.celrep.2012.05.019
- Bindereif, A., and Green, M. R. (1990). Identification and functional analysis of mammalian splicing factors. *Genet. Eng. (N. Y.)* 12, 201–224. doi:10.1007/978-1-4613-0641-2_11
- Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336. doi:10.1146/annurev.biochem.72.121801.161720
- Blankvoort, S., Witter, M. P., Noonan, J., Cotney, J., and Kentros, C. (2018). Marked diversity of unique cortical enhancers enables neuron-specific tools by enhancer-driven gene expression. *Curr. Biol. CB* 28, 2103–2114. doi:10.1016/j.cub.2018.05.015
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. U. S. A.* 75, 4853–4857. doi:10.1073/pnas.75.10.4853
- Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50, 349–383. doi:10.1146/annurev.bi.50.070181.002025
- Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Lüthmann, R. (1984). Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J.* 3, 1357–1363. doi:10.1002/j.1460-2075.1984.tb01977.x
- Bringmann, P., and Lüthmann, R. (1986). Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* 5, 3509–3516. doi:10.1002/j.1460-2075.1986.tb04676.x
- Bringmann, P., Rinke, J., Appel, B., Reuter, R., and Lüthmann, R. (1983). Purification of snRNPs U1, U2, U4, U5 and U6 with 2,2,7-trimethylguanosine-specific antibody and definition of their constituent proteins reacting with anti-Sm and anti-(U1)RNP antisera. *EMBO J.* 2, 1129–1135. doi:10.1002/j.1460-2075.1983.tb01557.x
- Brinster, R. L., Allen, J. M., Behringer, R. R., Gelin, R. E., and Palmiter, R. D. (1988). Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* 85, 836–840. doi:10.1073/pnas.85.3.836
- Burge, C. B., Padgett, R. A., and Sharp, P. A. (1998). Evolutionary fates and origins of U12-type introns. *Mol. Cell* 2, 773–785. doi:10.1016/s1097-2765(00)80292-0
- Cáceres, J. F., McKenzie, D., Thimmapaya, R., Lund, E., and Dahlberg, J. E. (1992). Control of mouse U1a and U1b snRNA gene expression by differential transcription. *Nucleic Acids Res.* 20, 4247–4254. doi:10.1093/nar/20.16.4247
- Carmel, L., Wolf, Y. I., Rogozin, I. B., and Koonin, E. V. (2007). Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Res.* 17, 1034–1044. doi:10.1101/gr.6438607
- Carranza, F., Shenasa, H., and Hertel, K. J. (2022). Splice site proximity influences alternative exon definition. *RNA Biol.* 19, 829–840. doi:10.1080/15476286.2022.2089478
- Catania, F., Gao, X., and Scofield, D. G. (2009). Endogenous mechanisms for the origins of spliceosomal introns. *J. Hered.* 100, 591–596. doi:10.1093/jhered/esp062
- Chen, L., Luillo, D. J., Ma, E., Celniker, S. E., Rio, D. C., and Doudna, J. A. (2005). Identification and analysis of U5 snRNA variants in *Drosophila*. *RNA* 11, 1473–1477. doi:10.1261/rna.2141505
- Ciavarella, J., Perea, W., and Greenbaum, N. L. (2020). Topology of the U12–U6^{atc} snRNA complex of the minor spliceosome and binding by NTC-related protein RBM22. *ACS Omega* 5, 23549–23558. doi:10.1021/acsomega.0c01674
- Crick, F. (1979). Split genes and RNA splicing. *Science* 204, 264–271. doi:10.1126/science.373120
- Csuros, M., Rogozin, I. B., and Koonin, E. V. (2011). A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Comput. Biol.* 7, e1002150. doi:10.1371/journal.pcbi.1002150
- Damgaard, C. K., Kahns, S., Lykke-Andersen, S., Nielsen, A. L., Jensen, T. H., and Kjems, J. (2008). A 5' splice site enhances the recruitment of basal transcription initiation factors *in vivo*. *Mol. Cell* 29, 271–278. doi:10.1016/j.molcel.2007.11.035
- De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *WIREs RNA* 4, 49–60. doi:10.1002/wrna.1140
- de Melo Costa, V. R., Pfeuffer, J., Louloui, A., Örom, U. A. V., and Piro, R. M. (2021). SPLICE-Q: A Python tool for genome-wide quantification of splicing efficiency. *BMC Bioinforma.* 22, 368. doi:10.1186/s12859-021-04282-6
- de Wolf, B., Oghabian, A., Akinyi, M. V., Hanks, S., Tromer, E. C., van Hooff, J. J. E., et al. (2021). Chromosomal instability by mutations in the novel minor spliceosome component CENATAC. *EMBO J.* 40, e106536. doi:10.15252/embj.202106536
- Denison, R. A., Van Arsdell, S. W., Bernstein, L. B., and Weiner, A. M. (1981). Abundant pseudogenes for small nuclear RNAs are dispersed in the human genome. *Proc. Natl. Acad. Sci.* 78, 810–814. doi:10.1073/pnas.78.2.810

- Dergai, O., Cousin, P., Gouge, J., Satia, K., Praz, V., Kuhlman, T., et al. (2018). Mechanism of selective recruitment of RNA polymerases II and III to snRNA gene promoters. *Genes Dev.* 32, 711–722. doi:10.1101/gad.314245.118
- Dey, P., and Mattick, J. S. (2021). High frequency of intron retention and clustered H3K4me3-marked nucleosomes in short first introns of human long non-coding RNAs. *Epigenetics Chromatin* 14, 45. doi:10.1186/s13072-021-00419-2
- Di, C., So, B. R., Cai, Z., Arai, C., Duan, J., and Dreyfuss, G. (2019). U1 snRNP telescripting roles in transcription and its mechanism. *Cold Spring Harb. Symp. Quant. Biol.* 84, 115–122. doi:10.1101/sqb.2019.84.040451
- Di Giammartino, D. C., Nishida, K., and Manley, J. L. (2011). Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* 43, 853–866. doi:10.1016/j.molcel.2011.08.017
- Dietrich, R. C., Inorvaia, R., and Padgett, R. A. (1997). Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol. Cell* 1, 151–160. doi:10.1016/s1097-2765(00)80016-7
- Dominov, J. A., Uyan, Ö., McKenna-Yasek, D., Nallamilli, B. R. R., Kergourlay, V., Bartoli, M., et al. (2019). Correction of pseudoexon splicing caused by a novel intronic dysferlin mutation. *Ann. Clin. Transl. Neurol.* 6, 642–654. doi:10.1002/acn3.738
- Domitrovich, A. M., and Kunkel, G. R. (2003). Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies. *Nucleic Acids Res.* 31, 2344–2352. doi:10.1093/nar/gkg331
- Doucet, A. J., Droc, G., Siol, O., Audoux, J., and Gilbert, N. (2015). U6 snRNA pseudogenes: Markers of retrotransposition dynamics in mammals. *Mol. Biol. Evol.* 32, 1815–1832. doi:10.1093/molbev/msv062
- Drexler, H. L., Choquet, K., and Churchman, L. S. (2020). Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* 77, 985–998. doi:10.1016/j.molcel.2019.11.017
- Ederly, P., Marcaillou, C., Sahbatou, M., Labalme, A., Chastang, J., Touraine, R., et al. (2011). Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science* 332, 240–243. doi:10.1126/science.1202205
- Effenberger, K. A., Urabe, V. K., and Jurica, M. S. (2017). Modulating splicing with small molecular inhibitors of the spliceosome: Modulating splicing with small molecular inhibitors. *Wiley Interdiscip. Rev. RNA* 8, e1381. doi:10.1002/wrna.1381
- Elliott, B., Richardson, C., and Jasin, M. (2005). Chromosomal translocation mechanisms at intronic Alu elements in mammalian cells. *Mol. Cell* 17, 885–894. doi:10.1016/j.molcel.2005.02.028
- Elsaid, M. F., Chalhoub, N., Ben-Omran, T., Kumar, P., Kamel, H., Ibrahim, K., et al. (2017). Mutation in noncoding RNA RNU12 causes early onset cerebellar ataxia. *Ann. Neurol.* 81, 68–78. doi:10.1002/ana.24826
- Emera, D., Yin, J., Reilly, S. K., Gockley, J., and Noonan, J. P. (2016). Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc. Natl. Acad. Sci.* 113, E2617–E2626. doi:10.1073/pnas.1603718113
- Engreitz, J. M., Sirokman, K., McDonel, P., Shishkin, A. A., Surka, C., Russell, P., et al. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* 159, 188–199. doi:10.1016/j.cell.2014.08.018
- Fox-Walsh, K. L., Dou, Y., Lam, B. J., Hung, S., Baldi, P. F., and Hertel, K. J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl. Acad. Sci.* 102, 16176–16181. doi:10.1073/pnas.0508489102
- Franchini, L. F., López-Leal, R., Nasif, S., Beati, P., Gelman, D. M., Low, M. J., et al. (2011). Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retrotransposons. *Proc. Natl. Acad. Sci. USA* 108, 15270–15275. doi:10.1073/pnas.1104997108
- Friend, K., Lovejoy, A. F., and Steitz, J. A. (2007). U2 snRNP binds intronless histone pre-mRNAs to facilitate U7-snRNP-Dependent 3'-end formation. *Mol. Cell* 28, 240–252. doi:10.1016/j.molcel.2007.09.026
- Frigola, J., Sabarinathan, R., Mularoni, L., Muiños, F., Gonzalez-Perez, A., and López-Bigas, N. (2017). Reduced mutation rate in exons due to differential mismatch repair. *Nat. Genet.* 49, 1684–1692. doi:10.1038/ng.3991
- Frumkin, I., Yofe, I., Bar-Ziv, R., Gurvich, Y., Lu, Y.-Y., Voichek, Y., et al. (2019). Evolution of intron splicing towards optimized gene expression is based on various Cis- and Trans-molecular mechanisms. *PLoS Biol.* 17, e3000423. doi:10.1371/journal.pbio.3000423
- Furger, A., O'Sullivan, J. M., Binnie, A., Lee, B. A., and Proudfoot, N. J. (2002). Promoter proximal splice sites enhance transcription. *Genes Dev.* 16, 2792–2799. doi:10.1101/gad.983602
- Gehring, N. H., and Roignant, J.-Y. (2021). Anything but ordinary – emerging splicing mechanisms in eukaryotic gene regulation. *Trends Genet.* 37, 355–372. doi:10.1016/j.tig.2020.10.008
- Gelfman, S., Burstein, D., Penn, O., Savchenko, A., Amit, M., Schwartz, S., et al. (2011). Changes in exon–intron structure during vertebrate evolution affect the splicing pattern of exons. *Genome Res.* 22, 35–50. doi:10.1101/gr.119834.110
- Georgomanolis, T., Sofiadis, K., and Papantonis, A. (2016). Cutting a long intron short: Recursive splicing and its implications. *Front. Physiol.* 7, 598. doi:10.3389/fphys.2016.00598
- Gilbert, W. (1978). Why genes in pieces? *Nature* 271, 501. doi:10.1038/271501a0
- Gondane, A., and Itkonen, H. M. (2023). Revealing the history and mystery of RNA-seq. *Curr. Issues Mol. Biol.* 45, 1860–1874. doi:10.3390/cimb45030120
- Gozashti, L., Roy, S. W., Thornlow, B., Kramer, A., Ares, M., and Corbett-Detig, R. (2022). Transposable elements drive intron gain in diverse eukaryotes. *Proc. Natl. Acad. Sci.* 119, e2209766119. doi:10.1073/pnas.2209766119
- Grabowski, P. J., Seiler, S. R., and Sharp, P. A. (1985). A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* 42, 345–353. doi:10.1016/S0092-8674(85)80130-6
- Griffin, C., and Saint-Jeannet, J.-P. (2020). Spliceosomopathies: Diseases and mechanisms. *Dev. Dyn.* 249, 1038–1046. doi:10.1002/dvdy.214
- Hall, S. L., and Padgett, R. A. (1994). Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* 239, 357–365. doi:10.1006/jmbi.1994.1377
- Hall, S. L., and Padgett, R. A. (1996). Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* 271, 1716–1718. doi:10.1126/science.271.5256.1716
- Halperin, R. F., Hegde, A., Lang, J. D., Raupach, E. A., Legendre, C., Liang, W. S., et al. (2021). Improved methods for RNAseq-based alternative splicing analysis. *Sci. Rep.* 11, 10740. doi:10.1038/s41598-021-89938-2
- Henry, R. W., Mittal, V., Ma, B., Kobayashi, R., and Hernandez, N. (1998). SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. *Genes Dev.* 12, 2664–2672. doi:10.1101/gad.12.17.2664
- Herzel, L., Ottow, D. S. M., Alpert, T., and Neugebauer, K. M. (2017). Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* 18, 637–650. doi:10.1038/nrm.2017.63
- Hesselberth, J. R. (2013). Lives that introns lead after splicing. *Wiley Interdiscip. Rev. RNA* 4, 677–691. doi:10.1002/wrna.1187
- Hoffman, M. M., and Birney, E. (2006). Estimating the neutral rate of nucleotide substitution using introns. *Mol. Biol. Evol.* 24, 522–531. doi:10.1093/molbev/msl179
- Huang, Y., Gu, L., and Li, G.-M. (2018). H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *J. Biol. Chem.* 293, 7811–7823. doi:10.1074/jbc.RA118.002839
- Hudson, A. J., McWatters, D. C., Bowser, B. A., Moore, A. N., Larue, G. E., Roy, S. W., et al. (2019). Patterns of conservation of spliceosomal intron structures and spliceosome divergence in representatives of the diplomonad and parabasalid lineages. *BMC Evol. Biol.* 19, 162. doi:10.1186/s12862-019-1488-y
- Huff, J. T., Zilberman, D., and Roy, S. W. (2016). Mechanism for DNA transposons to generate introns on genomic scales. *Nature* 538, 533–536. doi:10.1038/nature20110
- Inorvaia, R., and Padgett, R. A. (1998). Base pairing with U6atac snRNA is required for 5' splice site activation of U12-dependent introns *in vivo*. *RNA* 4, 709–718. doi:10.1017/s1355838298980207
- Jackson, I. J. (1991). A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* 19, 3795–3798. doi:10.1093/nar/19.14.3795
- Jakt, L. M., Dubin, A., and Johansen, S. D. (2022). Intron size minimisation in teleosts. *BMC Genomics* 23, 628. doi:10.1186/s12864-022-08760-w
- Jawdekar, G., and Henry, R. (2008). Transcriptional regulation of human small nuclear RNA genes. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 1779, 295–305. doi:10.1016/j.bbagr.2008.04.001
- Jiang, M., Zhang, S., Yin, H., Zhuo, Z., and Meng, G. (2023). A comprehensive benchmarking of differential splicing tools for RNA-seq analysis at the event level. *Brief. Bioinform.*, bbad121. doi:10.1093/bib/bbad121
- Joseph, B., Scala, C., Kondo, S., and Lai, E. C. (2022). Molecular and genetic dissection of recursive splicing. *Life Sci. Alliance* 5, e202101063. doi:10.26508/lsa.202101063
- Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., et al. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668. doi:10.1038/nature09479
- Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., et al. (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.* 3, 576–583. doi:10.1038/nchembio.2007.18
- Kandul, N. P., and Noor, M. A. (2009). Large introns in relation to alternative splicing and gene evolution: A case study of *Drosophila bruno-3*. *BMC Genet.* 10, 67. doi:10.1186/1471-2156-10-67
- Kapustin, Y., Chan, E., Sarkar, R., Wong, F., Vorechovsky, I., Winston, R. M., et al. (2011). Cryptic splice sites and split genes. *Nucleic Acids Res.* 39, 5837–5844. doi:10.1093/nar/gkr203
- Kazanian, H. H., and Moran, J. V. (1998). The impact of L1 retrotransposons on the human genome. *Nat. Genet.* 19, 19–24. doi:10.1038/ng0598-19
- Keane, P. A., and Seoighe, C. (2016). Intron length coevolution across mammalian genomes. *Mol. Biol. Evol.* 33, 2682–2691. doi:10.1093/molbev/msw151
- Kolosova, I., and Padgett, R. A. (1997). U11 snRNA interacts *in vivo* with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA* 3, 227–233.

- Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: A compromise solution to the introns-early versus introns-late debate? *Biol. Direct* 1, 22. doi:10.1186/1745-6150-1-22
- Krchňáková, Z., Thakur, P. K., Krausová, M., Bieberstein, N., Haberman, N., Müller-McNicoll, M., et al. (2019). Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5' splice-site sequences due to weak interactions with SR proteins. *Nucleic Acids Res.* 47, 911–928. doi:10.1093/nar/gky1147
- Kumari, A., Sedehzadeh, S., Brook, J. D., Kozlowski, P., and Wojciechowska, M. (2022). Differential fates of introns in gene expression due to global alternative splicing. *Hum. Genet.* 141, 31–47. doi:10.1007/s00439-021-02409-6
- Kwek, K. Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., et al. (2002). U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat. Struct. Biol.* 9, 800–805. doi:10.1038/nsb862
- Lambowitz, A. M., and Zimmerly, S. (2011). Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* 3, a003616. doi:10.1101/cshperspect.a003616
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi:10.1038/35057062
- Larue, G. E., Eliáš, M., and Roy, S. W. (2021). Expansion and transformation of the minor spliceosomal system in the slime mold *Physarum polycephalum*. *Curr. Biol.* 31, 3125–3131.e4. doi:10.1016/j.cub.2021.04.050
- Laurent, L., Wong, E., Li, G., Huynh, T., Tsigiris, A., Ong, C. T., et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res.* 20, 320–331. doi:10.1101/gr.101907.109
- Laxa, M., Müller, K., Lange, N., Doering, L., Pruscha, J. T., and Peterhänsel, C. (2016). The 5'UTR intron of arabidopsis GGT1 aminotransferase enhances promoter activity by recruiting RNA polymerase II. *Plant Physiol.* 172, 313–327. doi:10.1104/pp.16.00881
- Leader, Y., Lev Maor, G., Sorek, M., Shayevitch, R., Hussein, M., Hameiri, O., et al. (2021). The upstream 5' splice site remains associated to the transcription machinery during intron synthesis. *Nat. Commun.* 12, 4545. doi:10.1038/s41467-021-24774-6
- Lee, S., and Stevens, S. W. (2016). Spliceosomal intronogenesis. *Proc. Natl. Acad. Sci.* 113, 6514–6519. doi:10.1073/pnas.1605113113
- Lev-Maor, G., Sorek, R., Levanon, E. Y., Paz, N., Eisenberg, E., and Ast, G. (2007). RNA-editing-mediated exon evolution. *Genome Biol.* 8, R29. doi:10.1186/gb-2007-8-2-r29
- Li, X., Liu, S., Zhang, L., Issaian, A., Hill, R. C., Espinosa, S., et al. (2019). A unified mechanism for intron and exon definition and back-splicing. *Nature* 573, 375–380. doi:10.1038/s41586-019-1523-6
- Li, Y., Xu, Y., and Ma, Z. (2017). Comparative analysis of the exon-intron structure in eukaryotic genomes. *Yangetze Med.* 01, 50–64. doi:10.4236/ym.2017.11006
- Liao, S. E., and Regev, O. (2021). Splicing at the phase-separated nuclear speckle interface: A model. *Nucleic Acids Res.* 49, 636–645. doi:10.1093/nar/gkaa1209
- Lin, C.-F., Mount, S. M., Jarmolowski, A., and Makalowski, W. (2010). Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol. Biol.* 10, 47. doi:10.1186/1471-2148-10-47
- Liu, J., and Maxwell, E. S. (1990). Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. *Nucleic Acids Res.* 18, 6565–6571. doi:10.1093/nar/18.22.6565
- Lo, P. C., and Mount, S. M. (1990). *Drosophila melanogaster* genes for U1 snRNA variants and their expression during development. *Nucleic Acids Res.* 18, 6971–6979. doi:10.1093/nar/18.23.6971
- Lorenzi, C., Barriere, S., Arnold, K., Luco, R. F., Oldfield, A. J., and Ritchie, W. (2021). IRFinder-S: A comprehensive suite to discover and explore intron retention. *Genome Biol.* 22, 307. doi:10.1186/s13059-021-02515-8
- Lu, Z., and Matera, A. G. (2014). Developmental analysis of spliceosomal snRNA isoform expression. *G3 Bethesda Md* 5, 103–110. doi:10.1534/g3.114.015735
- Lynch, M., and Richardson, A. O. (2002). The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* 12, 701–710. doi:10.1016/s0959-437x(02)00360-x
- Mabin, J. W., Lewis, P. W., Brow, D. A., and Dvinge, H. (2021). Human spliceosomal snRNA sequence variants generate variant spliceosomes. *RNA* 27, 1186–1203. doi:10.1261/rna.078768.121
- Malca, H., Shomron, N., and Ast, G. (2003). The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. *Am. Soc. Microbiol.* 3, 3442–3455. doi:10.1128/MCB.23.10.3442-3455.2003
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U. S. A.* 36, 344–355. doi:10.1073/pnas.36.6.344
- Meng, F., Zhao, H., Zhu, B., Zhang, T., Yang, M., Li, Y., et al. (2021). Genomic editing of intronic enhancers unveils their role in fine-tuning tissue-specific gene expression in *Arabidopsis thaliana*. *Plant Cell* 33, 1997–2014. doi:10.1093/plcell/koab093
- Mercer, T. R., Clark, M. B., Andersen, S. B., Brunck, M. E., Haerty, W., Crawford, J., et al. (2015). Genome-wide discovery of human splicing branchpoints. *Genome Res.* 25, 290–303. doi:10.1101/gr.182899.114
- Michel, F., Umesono, K., and Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns—a review. *Gene* 82, 5–30. doi:10.1016/0378-1119(89)90026-7
- Mittal, V., Ma, B., and Hernandez, N. (1999). SNAP(c): A core promoter factor with a built-in DNA-binding damper that is deactivated by the oct-1 POU domain. *Genes Dev.* 13, 1807–1821. doi:10.1101/gad.13.14.1807
- Montzka, K. A., and Steitz, J. A. (1988). Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc. Natl. Acad. Sci. U. S. A.* 85, 8885–8889. doi:10.1073/pnas.85.23.8885
- Morales, J., Borrero, M., Sumerel, J., and Santiago, C. (1997). Identification of developmentally regulated sea urchin U5 snRNA genes. *DNA Seq. J. DNA Seq. Mapp.* 7, 243–259. doi:10.3109/10425179709034044
- Moyer, D. C., Larue, G. E., Hershberger, C. E., Roy, S. W., and Padgett, R. A. (2020). Comprehensive database and evolutionary dynamics of U12-type introns. *Nucleic Acids Res.* 48, 7066–7078. doi:10.1093/nar/gkaa464
- Neugebauer, K. M. (2019). Nascent RNA and the coordination of splicing with transcription. *Cold Spring Harb. Perspect. Biol.* 11, a032227. doi:10.1101/cshperspect.a032227
- Nilsen, T. W. (2003). The spliceosome: The most complex macromolecular machine in the cell? *Bioessays* 25, 1147–1149. doi:10.1002/bies.10394
- Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., et al. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* 161, 526–540. doi:10.1016/j.cell.2015.03.027
- Norppa, A. J., and Frilander, M. J. (2021). The integrity of the U12 snRNA 3' stem-loop is necessary for its overall stability. *Nucleic Acids Res.* 49, 2835–2847. doi:10.1093/nar/gkab048
- Olthof, A. M., Hyatt, K. C., and Kanadia, R. N. (2019). Minor intron splicing revisited: Identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. *BMC Genomics* 20, 686. doi:10.1186/s12864-019-6046-x
- Olthof, A. M., White, A. K., and Kanadia, R. N. (2022). The emerging significance of splicing in vertebrate development. *Development* 149, dev200373. doi:10.1242/dev.200373
- Olthof, A. M., White, A. K., Mieruszynski, S., Doggett, K., Lee, M. F., Chakroun, A., et al. (2021). Disruption of exon-bridging interactions between the minor and major spliceosomes results in alternative splicing around minor introns. *Nucleic Acids Res.* 49, 3524–3545. doi:10.1093/nar/gkab118
- O'Reilly, D., Dienstbier, M., Cowley, S. A., Vazquez, P., Drozd, M., Taylor, S., et al. (2013). Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. *Genome Res.* 23, 281–291. doi:10.1101/gr.142968.112
- Pai, A. A., Henriques, T., McCue, K., Burkholder, A., Adelman, K., and Burge, C. B. (2017). The kinetics of pre-mRNA splicing in the *Drosophila* genome and the influence of gene architecture. *eLife* 6, e32537. doi:10.7554/eLife.32537
- Parenteau, J., Durand, M., Véronneau, S., Lacombe, A.-A., Morin, G., Guérin, V., et al. (2008). Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol. Biol. Cell* 19, 1932–1941. doi:10.1091/mbc.E07-12-1254
- Pessa, H. K. J., Will, C. L., Meng, X., Schneider, C., Watkins, N. J., Perälä, N., et al. (2008). Minor spliceosome components are predominantly localized in the nucleus. *Proc. Natl. Acad. Sci.* 105, 8655–8660. doi:10.1073/pnas.0803646105
- Piovesan, A., Caracausi, M., Ricci, M., Strippoli, P., Vitale, L., and Pelleri, M. C. (2015). Identification of minimal eukaryotic introns through GeneBase, a user-friendly tool for parsing the NCBI Gene databank. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 22, 495–503. doi:10.1093/dnares/dsv028
- Pitollì, C., Marini, A., Sette, C., and Pagliarini, V. (2022). Non-canonical splicing and its implications in brain physiology and cancer. *Int. J. Mol. Sci.* 23, 2811. doi:10.3390/ijms23052811
- Qian, X., Wang, J., Wang, M., Igelman, A. D., Jones, K. D., Li, Y., et al. (2021). Identification of deep-intronic splice mutations in a large cohort of patients with inherited retinal diseases. *Front. Genet.* 12. doi:10.3389/fgene.2021.647400
- Reddy, R., Henning, D., Das, G., Harless, M., and Wright, D. (1987). The capped U6 small nuclear RNA is transcribed by RNA polymerase III. *J. Biol. Chem.* 262, 75–81. doi:10.1016/s0021-9258(19)75890-6
- Reed, R., and Maniatis, T. (1986). A role for exon sequences and splice-site proximity in splice-site selection. *Cell* 46, 681–690. doi:10.1016/0092-8674(86)90343-0
- Reimer, K. A., Mimoso, C. A., Adelman, K., and Neugebauer, K. M. (2021). Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell* 81, 998–1012.e7. doi:10.1016/j.molcel.2020.12.018
- Resch, A. M., Carmel, L., Mariño-Ramírez, L., Ogurtsov, A. Y., Shabalina, S. A., Rogozin, I. B., et al. (2007). Widespread positive selection in synonymous sites of mammalian genes. *Mol. Biol. Evol.* 24, 1821–1831. doi:10.1093/molbev/msm100
- Robberson, B. L., Cote, G. J., and Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell Biol.* 10, 84–94. doi:10.1128/mcb.10.1.84

- Roca, X., Sachidanandam, R., and Krainer, A. R. (2003). Intrinsic differences between authentic and cryptic 5' splice sites. *Nucleic Acids Res.* 31, 6321–6333. doi:10.1093/nar/gkg830
- Rodriguez-Galindo, M., Casillas, S., Weghorn, D., and Barbadilla, A. (2020). Germline de novo mutation rates on exons versus introns in humans. *Nat. Commun.* 11, 3304. doi:10.1038/s41467-020-17162-z
- Rogozin, I. B., Carmel, L., Csuros, M., and Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. *Biol. Direct* 7, 11. doi:10.1186/1745-6150-7-11
- Romfo, C. M., Alvarez, C. J., van Heeckeren, W. J., Webb, C. J., and Wise, J. A. (2000). Evidence for splice site pairing via intron definition in *Schizosaccharomyces pombe*. *Mol. Cell Biol.* 20, 7955–7970. doi:10.1128/mcb.20.21.7955-7970.2000
- Rose, A. B., Elfers, T., Parra, G., and Korf, I. (2008). Promoter-proximal introns in *Arabidopsis thaliana* are enriched in dispersed signals that elevate gene expression. *Plant Cell* 20, 543–551. doi:10.1105/tpc.107.057190
- Russell, A. G., Charette, J. M., Spencer, D. F., and Gray, M. W. (2006). An early evolutionary origin for the minor spliceosome. *Nature* 443, 863–866. doi:10.1038/nature05228
- Ryll, J., Rothering, R., and Catania, F. (2022). Intronzation signatures in coding exons reveal the evolutionary fluidity of eukaryotic gene architecture. *Microorganisms* 10, 1901. doi:10.3390/microorganisms10101901
- Sadowski, C. L., Henry, R. W., Lobo, S. M., and Hernandez, N. (1993). Targeting TBP to a non-TATA box cis-regulatory element: A TBP-containing complex activates transcription from snRNA promoters through the PSE. *Genes Dev.* 7, 1535–1548. doi:10.1101/gad.7.8.1535
- Sakharkar, M. K., Chow, V. T. K., and Kanguane, P. (2004). Distributions of exons and introns in the human genome. *Silico Biol.* 4, 387–393.
- Samadder, P., Sivamani, E., Lu, J., Li, X., and Qu, R. (2008). Transcriptional and post-transcriptional enhancement of gene expression by the 5' UTR intron of rice rubi3 gene in transgenic rice cells. *Mol. Genet. Genomics MGG* 279, 429–439. doi:10.1007/s00438-008-0323-8
- Sánchez-Escabias, E., Guerrero-Martínez, J. A., and Reyes, J. C. (2022). Co-transcriptional splicing efficiency is a gene-specific feature that can be regulated by TGFβ. *Nat. Commun.* 5, 277. doi:10.1038/s42003-022-03224-z
- SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., et al. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274, 765–768. doi:10.1126/science.274.5288.765
- Seal, R. L., Chen, L.-L., Griffiths-Jones, S., Lowe, T. M., Mathews, M. B., O'Reilly, D., et al. (2020). A guide to naming human non-coding RNA genes. *EMBO J.* 39, e103777. doi:10.15252/embj.2019103777
- Sharp, P. A. (1991). Five easy pieces. *Science* 254, 663. doi:10.1126/science.1948046
- Shen, H., Zheng, X., Luecke, S., and Green, M. R. (2010). The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev.* 24, 2389–2394. doi:10.1101/gad.1974810
- Shepard, S., McCreary, M., and Fedorov, A. (2009). The peculiarities of large intron splicing in animals. *PLOS ONE* 4, e7853. doi:10.1371/journal.pone.0007853
- Sheridan, R. M., Fong, N., D'Alessandro, A., and Bentley, D. L. (2019). Widespread backtracking by RNA pol II is a major effector of gene activation, 5' pause release, termination, and transcription elongation rate. *Mol. Cell* 73, 107–118. doi:10.1016/j.molcel.2018.10.031
- Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R., and Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* 34, 3955–3967. doi:10.1093/nar/gkl556
- Shiau, C.-K., Huang, J.-H., Liu, Y.-T., and Tsai, H.-K. (2022). Genome-wide identification of associations between enhancer and alternative splicing in human and mouse. *BMC Genomics* 22, 919. doi:10.1186/s12864-022-08537-1
- Shukla, G. C., and Padgett, R. A. (2002). A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol. Cell* 9, 1145–1150. doi:10.1016/S1097-2765(02)00505-1
- Siebert, A. E., Corll, J., Paige Gronevelt, J., Levine, L., Hobbs, L. M., Kenney, C., et al. (2022). Genetic analysis of human RNA binding motif protein 48 (RBM48) reveals an essential role in U12-type intron splicing. *Genetics* 222, iyac129. doi:10.1093/genetics/iyac129
- Sierra-Montes, J. M., Pereira-Simon, S., Smail, S. S., and Herrera, R. J. (2005). The silk moth *Bombyx mori* U1 and U2 snRNA variants are differentially expressed. *Gene* 352, 127–136. doi:10.1016/j.gene.2005.02.013
- Simari, R. D., Yang, Z.-Y., Ling, X., Stephan, D., Perkins, N. D., Nabel, G. J., et al. (1998). Requirements for enhanced transgene expression by untranslated sequences from the human cytomegalovirus immediate-early gene. *Mol. Med.* 4, 700–706. doi:10.1007/BF03401764
- Singh, J., and Padgett, R. A. (2009). Rates of *in situ* transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 16, 1128–1133. doi:10.1038/nsmb.1666
- Singh, P., Börger, C., More, H., and Sturmbauer, C. (2017). The role of alternative splicing and differential gene expression in Cichlid adaptive radiation. *Genome Biol. Evol.* 9, 2764–2781. doi:10.1093/gbe/evx204
- Smathers, C. M., and Robart, A. R. (2019). The mechanism of splicing as told by group II introns: Ancestors of the spliceosome. *Biochim. Biophys. Acta Gene Regul. Mech.* 1862, 194390. doi:10.1016/j.bbagr.2019.06.001
- So, B. R., Di, C., Cai, Z., Venters, C. C., Guo, J., Oh, J.-M., et al. (2019). A complex of U1 snRNP with cleavage and polyadenylation factors controls telescripting, regulating mRNA transcription in human cells. *Mol. Cell* 76, 590–599. doi:10.1016/j.molcel.2019.08.007
- Sontheimer, E. J., Gordon, P. M., and Piccirilli, J. A. (1999). Metal ion catalysis during group II intron self-splicing: Parallels with the spliceosome. *Genes Dev.* 13, 1729–1741. doi:10.1101/gad.13.13.1729
- Sontheimer, E. J., and Steitz, J. A. (1992). Three novel functional variants of human U5 small nuclear RNA. *Mol. Cell Biol.* 12, 734–746. doi:10.1128/mcb.12.2.734
- Sorek, R., Ast, G., and Graur, D. (2002). Alu-containing exons are alternatively spliced. *Genome Res.* 12, 1060–1067. doi:10.1101/gr.229302
- Spiller, M. P., Boon, K.-L., Reijns, M. A. M., and Beggs, J. D. (2007). The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. *Nucleic Acids Res.* 35, 923–929. doi:10.1093/nar/gkl1130
- Spiluttini, B., Gu, B., Belagal, P., Smirnova, A. S., Nguyen, V. T., Hebert, C., et al. (2010). Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *J. Cell Sci.* 123, 2085–2093. doi:10.1242/jcs.061358
- Steitz, J. A., Dreyfuss, G., Krainer, A. R., Lamond, A. I., Matera, A. G., and Padgett, R. A. (2008). Where in the cell is the minor spliceosome? *Proc. Natl. Acad. Sci. U. S. A.* 105, 8485–8486. doi:10.1073/pnas.0804024105
- Sun, H., and Chasin, L. A. (2000). Multiple splicing defects in an intronic false exon. *Mol. Cell Biol.* 20, 6414–6425. doi:10.1128/MCB.20.17.6414-6425.2000
- Sun, J., Li, X., Hou, X., Cao, S., Cao, W., Zhang, Y., et al. (2022). Structural basis of human SNAPc recognizing proximal sequence element of snRNA promoter. *Nat. Commun.* 13, 6871. doi:10.1038/s41467-022-34639-1
- Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G. L., Cheutin, T., Bonev, B., et al. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* 52, 1151–1157. doi:10.1038/s41588-020-00716-8
- Szentirmay, M. N., and Sawadogo, M. (2000). Spatial organization of RNA polymerase II transcription in the nucleus. *Nucleic Acids Res.* 28, 2019–2025. doi:10.1093/nar/28.10.2019
- Tammer, L., Hameiri, O., Keydar, I., Roy, V. R., Ashkenazy-Titelman, A., Custódio, N., et al. (2022). Gene architecture directs splicing outcome in separate nuclear spatial regions. *Mol. Cell* 82, 1021–1034.e8. doi:10.1016/j.molcel.2022.02.001
- Tarn, W. Y., and Steitz, J. A. (1996a). A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell* 84, 801–811. doi:10.1016/S0092-8674(00)81057-0
- Tarn, W. Y., and Steitz, J. A. (1996b). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* 273, 1824–1832. doi:10.1126/science.273.5283.1824
- Tarn, W. Y., and Steitz, J. A. (1997). Pre-mRNA splicing: The discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* 22, 132–137. doi:10.1016/S0968-0004(97)01018-9
- Tellier, M., Maudlin, I., and Murphy, S. (2020). Transcription and splicing: A two-way street. *WIREs RNA* 11, e1593. doi:10.1002/wrna.1593
- Theissen, H., Rinke, J., Traver, C. N., Lüthmann, R., and Appel, B. (1985). Novel structure of a human U6 snRNA pseudogene. *Gene* 36, 195–199. doi:10.1016/0378-1119(85)90086-1
- Thompson, P. J., Macfarlan, T. S., and Lorincz, M. C. (2016). Long terminal repeats: From parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* 62, 766–776. doi:10.1016/j.molcel.2016.03.029
- Tichelaar, J. W., Wieben, E. D., Reddy, R., Vrabel, A., and Camacho, P. (1998). *In vivo* expression of a variant human U6 RNA from a unique, internal promoter. *Biochemistry* 37, 12943–12951. doi:10.1021/bi9811361
- Tronçère, H., Wang, J., and Fu, X. D. (1997). A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. *Nature* 388, 397–400. doi:10.1038/41137
- Ulianov, S. V., Khrameeva, E. E., Gavrillov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., et al. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* 26, 70–84. doi:10.1101/gr.196006.115
- Vakirlis, N., Vance, Z., Duggan, K. M., and McLysaght, A. (2022). De novo birth of functional microproteins in the human lineage. *Cell Rep.* 41, 111808. doi:10.1016/j.celrep.2022.111808
- Van Arsdell, S. W., and Weiner, A. M. (1984). Human genes for U2 small nuclear RNA are tandemly repeated. *Mol. Cell Biol.* 4, 492–499. doi:10.1128/mcb.4.3.492

- Vazquez-Arango, P., Vowles, J., Browne, C., Hartfield, E., Fernandes, H. J. R., Mandefro, B., et al. (2016). Variant U1 snRNAs are implicated in human pluripotent stem cell maintenance and neuromuscular disease. *Nucleic Acids Res.* 44, 10960–10973. doi:10.1093/nar/gkw711
- Vosseberg, J., and Snel, B. (2017). Domestication of self-splicing introns during eukaryogenesis: The rise of the complex spliceosomal machinery. *Biol. Direct* 12, 30. doi:10.1186/s13062-017-0201-6
- Wahl, M. C., Will, C. L., and Lührmann, R. (2009). The spliceosome: Design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi:10.1016/j.cell.2009.02.009
- Wallace, E. W. J., and Beggs, J. D. (2017). Extremely fast and incredibly close: Cotranscriptional splicing in budding yeast. *RNA* 23, 601–610. doi:10.1261/rna.060830.117
- Wang, Z., Xiao, X., Van Nostrand, E., and Burge, C. B. (2006). General and specific functions of exonic splicing silencers in splicing control. *Mol. Cell* 23, 61–70. doi:10.1016/j.molcel.2006.05.018
- Wells, J. N., and Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* 54, 539–561. doi:10.1146/annurev-genet-040620-022145
- Will, C. L., Schneider, C., Reed, R., and Lührmann, R. (1999). Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* 284, 2003–2005. doi:10.1126/science.284.5422.2003
- Yan, D., and Ares, M. (1996). Invariant U2 RNA sequences bordering the branchpoint recognition region are essential for interaction with yeast SF3a and SF3b subunits. *Mol. Cell. Biol.* 16, 818–828. doi:10.1128/mcb.16.3.818
- Younis, I., Dittmar, K., Wang, W., Foley, S. W., Berg, M. G., Hu, K. Y., et al. (2013). Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *eLife* 2, e00780. doi:10.7554/eLife.00780
- Zhang, S., Aibara, S., Vos, S. M., Agafonov, D. E., Lührmann, R., and Cramer, P. (2021). Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science* 371, 305–309. doi:10.1126/science.abf1870
- Zhang, X.-O., Fu, Y., Mou, H., Xue, W., and Weng, Z. (2018). The temporal landscape of recursive splicing during Pol II transcription elongation in human cells. *PLoS Genet.* 14, e1007579. doi:10.1371/journal.pgen.1007579
- Zhu, D., Mao, F., Tian, Y., Lin, X., Gu, L., Gu, H., et al. (2020). The features and regulation of Co-transcriptional splicing in arabidopsis. *Mol. Plant* 13, 278–294. doi:10.1016/j.molp.2019.11.004