CellPress

Opinion

How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing?

Dror Hollander,[1,‡] Shiran Naftelberg,[1,‡] Galit Lev-Maor,[1] Alberto R. Kornblihtt,[2] and Gil Ast[1,*]

The splice sites (SSs) delimiting an intron are brought together in the earliest step of spliceosome assembly yet it remains obscure how SS pairing occurs, especially when introns are thousands of nucleotides long. Splicing occurs *in vivo* in mammals within minutes regardless of intron length, implying that SS pairing can instantly follow transcription. Also, factors required for SS pairing, such as the U1 small nuclear ribonucleoprotein (snRNP) and U2AF65, associate with RNA polymerase II (RNAPII), while nucleosomes preferentially bind exonic sequences and associate with U2 snRNP. Based on recent publications, we assume that the 5′ SS-bound U1 snRNP can remain tethered to RNAPII until complete synthesis of the downstream intron and exon. An additional U1 snRNP then binds the downstream 5′ SS, whereas the RNAPII-associated U2AF65 binds the upstream 3′ SS to facilitate SS pairing along with exon definition. Next, the nucleosome-associated U2 snRNP binds the branch site to advance splicing complex assembly. This may explain how RNAPII and chromatin are involved in spliceosome assembly and how introns lengthened during evolution with a relatively minimal compromise in splicing.

Trends

The splicing rates of very long mammalian introns and of short ones are similar. It is therefore baffling how spatially distant splice sites (SSs) are rapidly brought into proximity *in vivo* when introns are thousands of nucleotides long.

There is much evidence concerning the functional associations between splicing factors and RNA polymerase II (RNAPII) as well as between splicing factors and chromatin.

Since splicing factors involved in the identification of the 5′ and 3′ SSs associate with RNAPII, SS pairing could potentially occur closely following the synthesis of long introns as they are still attached to chromatin via RNAPII.

Functional associations between splicing factors and chromatin would later promote splicing factor recruitment to the RNA to advance the splicing reaction.

Splicing and Spliceosome Assembly

Splicing is the mRNA maturation reaction in which introns are removed from mRNA precursors and exons are ligated together [1]. Regulation of the splicing process is essential to ensure correct gene expression and for cellular responses to environmental changes [2]. The reaction is governed by four main regulatory consensus sequences: the 5′ SS, the 3′ SS, the branch site sequence, and the polypyrimidine tract [3–5]. It occurs within a multicomponent complex termed the spliceosome. The spliceosome comprises five snRNP complexes – U1, U2, U4, U5, and U6 – and many additional proteins [5]. The initial mechanistic analysis of the splicing reaction was enabled by the establishment of an *in vitro* splicing system 30 years ago [6–9] (Box 1). Based on the fact that introns are spliced less efficiently *in vitro* the larger they are, and the relatively long time it takes to remove an intron from an mRNA precursor in the *in vitro* assay, we can assume that *in vitro* the splicing factors locate their binding sites on the mRNA precursor via diffusion by stochastic interactions. How this binding occurs and what affects its kinetics *in vivo* is much less clear.

The Influence of Intron Length on Splicing Efficiency and Kinetics

Exon–intron structure plays a role in the recognition of exons by the splicing machinery. *In vitro* and *in vivo* studies in flies and vertebrates demonstrate that intron length negatively affects exon inclusion levels [10,11] and that alternatively spliced exons are flanked by longer introns compared with those flanking constitutively spliced exons [11–13]. Nevertheless, a large fraction

[1]Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel
[2]IFIBYNE-UBA-CONICET and Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón II, C1428EHA Buenos Aires, Argentina
‡These authors contributed equally to this work.

*Correspondence: gilast@post.tau.ac.il (G. Ast).

 http://dx.doi.org/10.1016/j.tig.2016.07.003

of introns expanded by thousands of nucleotides during vertebrate evolution without profoundly hindering splicing, with humans having the longest introns among the sequenced vertebrate genomes [13]. The introns examined in the *in vitro* splicing reaction are relatively short (up to a few hundred nucleotides long) while an average intron in vertebrates is often ten times longer [13]. Thus a puzzling issue is how the spliceosome successfully defines exons, which are on average about 150 nucleotides long in humans [13], and distinguishes between the pseudo-SSs found across vast intronic sequences and *bona fide* ones.

Splicing kinetics seems to be more affected by intron length in lower eukaryotes such as yeast and flies, while in mice and humans the rate of intron removal from the pre-mRNA is unrelated to intron length [14–16]. This could plausibly be explained by the evolutionary transition from intron definition to exon definition (Box 2): the removal of the relatively shorter yeast and fly introns [17–19] presumably occurs mainly via intron definition while mammalian exons flanked by longer introns are plausibly mainly identified via exon definition [20]. Since a diffusion-based model for splicing factor interactions with RNA cannot solely explain how very long introns are spliced at the same rate as short ones, we can assume that additional strategies must have developed during evolution to facilitate the identification of exons located between long introns.

Intron Removal Is Generally Rapid *In Vivo*

In vitro splicing products are first detected after roughly 30 min or more of incubation and splicing rates can be modulated depending on various factors [6,21,22]. Although it can be said that the splicing reaction is generally slower *in vitro*, splicing kinetics *in vivo* are less understood. It has been reported that human pre-mRNA is synthesized at 2.9–3.3 kilobases per minute and that the removal of the first intron occurs as RNAPII transcribes the second exon [23]. However, another study estimated the average human transcription rate to be 3.8 kilobases per minute and intron removal to occur within 5–10 min of intron synthesis [16]. Other studies indicate that human intron excision can occur within 20–30 s of their synthesis [24,25]. Further studies of splicing kinetics in relation to transcription in human cells found that intron removal occurs 4.33 min following synthesis of the 3′ SS [26] or 20–30 s after intron transcription [27]. Although *in vivo* splicing rate estimates differ between studies, the splicing reaction is probably completed *in vivo* within seconds to a few minutes, closely following RNA synthesis, whereas *in vitro* splicing product formation is more prolonged. So what difference between the *in vitro* and *in vivo* splicing reactions may explain this discrepancy in splicing kinetics?

Box 1. *In Vitro* Splicing and Spliceosome Assembly

In the *in vitro* splicing system, an mRNA precursor comprising two exons separated by an intron is incubated in nuclear extract. After incubation for approximately 30 min or more, the mRNA precursor is spliced and the first nucleotide of the intron forms a covalent bond with the branch site adenosine to form a lariat intermediate. Later, the two exons are ligated, the intron is released, and the final products of the mRNA splicing reaction gradually begin to accumulate. *In vitro* assays have contributed greatly to our understanding of the different stages of splicing complex assembly and the identification of splicing factors involved in each stage [6–9].

The first step in the assembly of the spliceosome is the formation of the commitment complex, also known as the E complex. In this complex, the U1 snRNP binds the 5′ SS of the pre-mRNA and is associated via SR proteins with the U2AF35– U2AF65 heterodimer (also known as U2AF1 and U2AF2, respectively) and SF1, which bind the 3′ SS, the polypyrimidine tract, and the branch site region, respectively [107,108] (Figure I). Formation of the commitment complex is completed within seconds and in it the 5′ and 3′ SSs are brought into close proximity. The SSs are thus defined at this early stage of the reaction. The commitment complex is then advanced into the pre-spliceosome, also known as the A complex, in which the U2 snRNP replaces SF1 on the branch site and interacts with the U1 snRNP [5,109]. The next step is the addition of the U4–U5–U6 tri-snRNP complex and formation of the fully assembled spliceosome, the B complex. Later stages involve activated B complex, catalytically activated B complex, the C complex, and the post-splicing complex. In these stages the U1 and U4 snRNPs detach from the spliceosome, other proteins attach to it, and structural rearrangements occur [5,110].

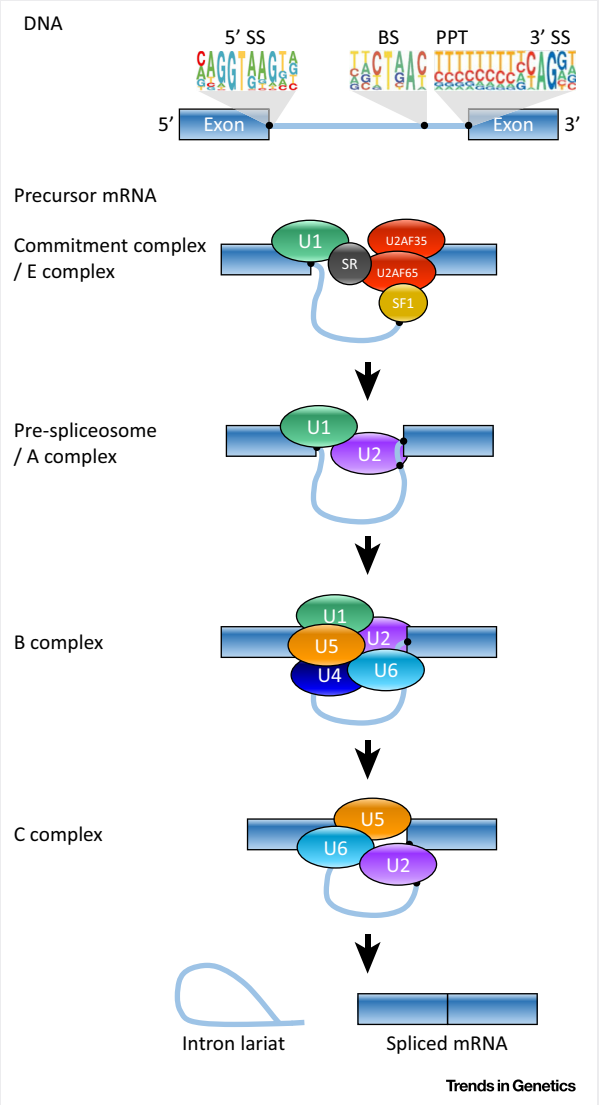


Figure I. The Stepwise Assembly of the Spliceosome.

Numerous studies have demonstrated that in various species splicing can occur co-transcriptionally *in vivo*, while pre-mRNA is being transcribed and RNAPII is still attached to chromatin (recently reviewed in [28]). In a recent yeast study, splicing was shown to be 50% complete in a set of tested genes when RNAPII is only 45 nucleotides downstream of the 3′ SS [29]. The extent of co-transcriptional splicing in the human genome remains unclear, ranging from a conservative estimate that splicing often occurs after transcription has been completed [30] to an opposite one that most pre-mRNAs undergo splicing while being transcribed [31]. Nevertheless, we hypothesize that the kinetic differences observed between splicing *in vitro* and *in vivo* could be related to the fact that the *in vitro* splicing system is completely transcription independent while *in vivo* splicing can occur co-transcriptionally.

The Connections of Splicing with RNAPII and Chromatin Structure *In Vivo*

Various studies demonstrate that RNAPII and chromatin structure are associated with splicing. RNAPII accumulates over intron–exon junctions and exons [32–34] and transcription elongation

Box 2. Exon Definition and Intron Definition

Two models for the recognition of splicing units have been suggested: exon definition and intron definition [20,95,111]. In the intron definition model, the splicing machinery recognizes the intronic unit and places the basal splicing machinery across introns, thereby constraining intron length. This mechanism is proposed to be widespread in lower eukaryotes and is also thought to be the ancestral splicing mechanism [95,112,113].

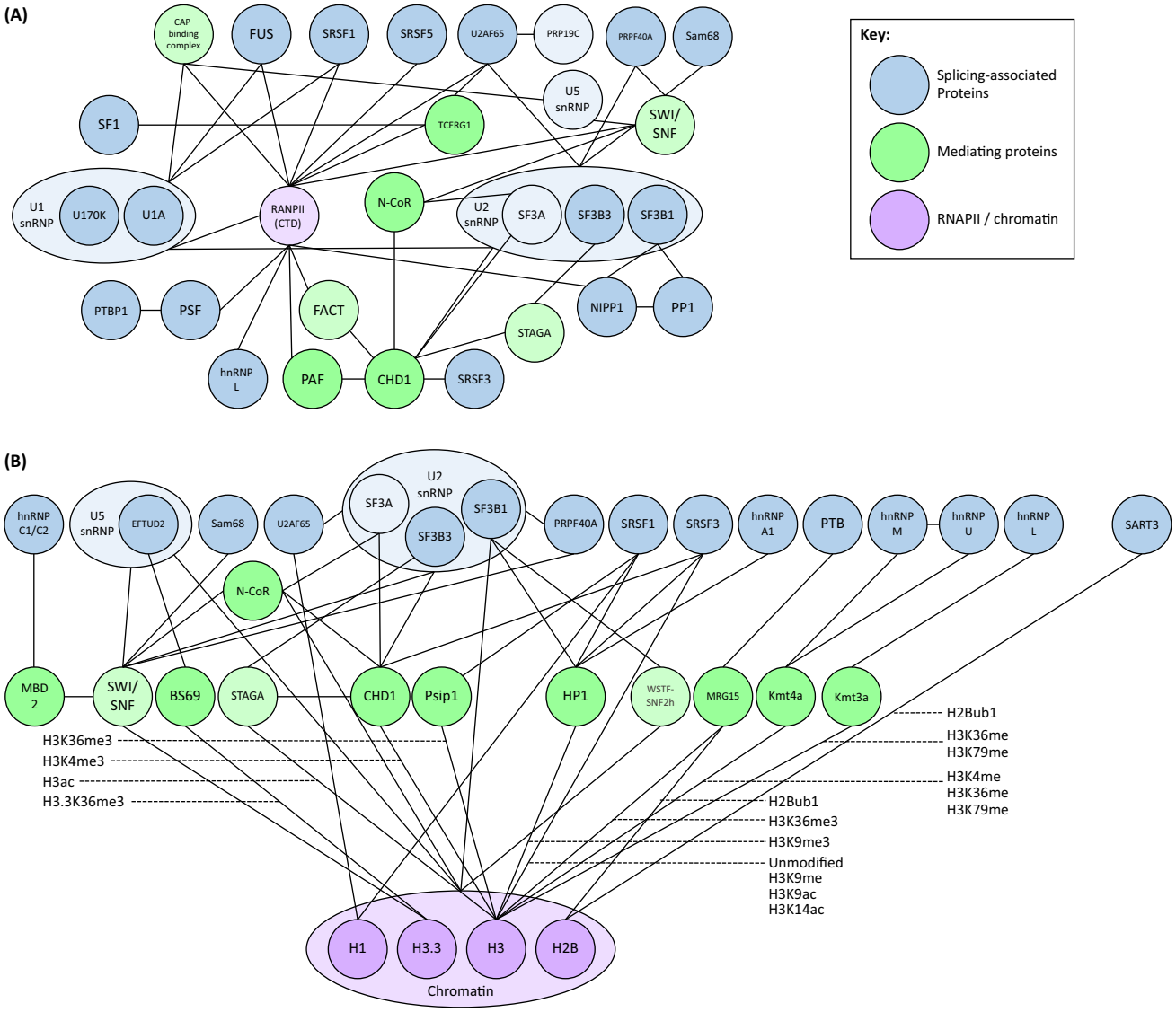
Exon definition is thought to be widespread in vertebrate species, which have a large fraction of long introns [113,114]. Exon definition occurs when the basal machinery is placed across exons, thus constraining their length and allowing their recognition in the context of substantially longer intronic sequences [11,95,111]. During exon definition, splicing-enhancing *cis* sequences within the exon recruit SR proteins that concurrently interact with the U2AF proteins and U1 snRNP located at the two ends of the exon [115,116]. The exon definition model is supported by the finding that mutating the 5′ SS of an exon affects U2 snRNP binding at the branch site upstream of that exon [99]. It is hypothesized that cross-exon interactions are later converted to a cross-intron complex connecting the U2 snRNP with the U1 snRNP bound to an upstream 5′ SS [107]. Such a conversion can presumably occur when an exon-bound U4–U5–U6 tri-snRNP interacts with an upstream 5′ SS without prior formation of a cross-intron A complex [117].

is thought to slow or pause at the 3′ end of introns [35,36]. High RNAPII elongation rates are positively correlated with certain epigenetic and gene features in human cell lines, among which is low density of exons in the genes [35]. Two similar methods of native elongating-transcript sequencing have been developed. The results of one study suggest that cleaved upstream-exon intermediate transcripts are associated with RNAPII that is accumulated over downstream exons when the RNAPII C-terminal domain (CTD) is specifically phosphorylated at the serine 5 position [37]; another study showed RNAPII accumulation at both ends of constitutive exons [38]. These findings are in accordance with evidence demonstrating that phosphorylation of the RNAPII CTD stimulates RNA splicing both *in vivo* and *in vitro* [39,40]. Finally, changes in RNAPII elongation rate influence the splicing of certain alternative exons, leading to either their increased inclusion or exclusion [41–45].

Less studied are the connections between splicing and chromatin structure. However, nucleosome occupancy, specific histone modifications, and CpG methylation levels differ between exons and introns [46–51]. Additionally, variations in chromatin organization and epigenetic marks are linked with and can directly or indirectly lead to alterations in splicing patterns [52–59,123]. In light of these findings, it is of interest to understand how RNAPII transcription elongation and chromatin structure affect the assembly of spliceosome complexes. Here we attempt to explain how long-intron removal from the pre-mRNA occurs co-transcriptionally *in vivo* via exon definition. Below, we describe a model in which splicing factor interactions with the RNAP CTD and chromatin during synthesis of the pre-mRNA can define exons as the splicing unit and promote the formation of the commitment complex. First, we examine and consolidate relevant findings pertaining to co-transcriptional spliceosome assembly.

Splicing Factor Recruitment to the RNAPII CTD *In Vivo*

The mammalian RNAPII CTD is necessary for efficient pre-mRNA processing *in vivo* [60–63]. In the absence of the CTD or when it cannot be phosphorylated, splicing is impaired [64,65], suggesting that the CTD is important for the recruitment of splicing factors to SSs. In accordance with this, the U1 snRNP associates with the CTD of RNAPII [66–72] (Figure 1A). The association can occur in the absence of transcription [66,69] but U1 snRNP is mainly localized to the chromatin of actively expressed genes in mammalian cells [73] regardless of whether these transcripts undergo splicing [69,70]. The association between the U1 snRNP and RNAPII may facilitate the uploading of the U1 snRNP not only onto authentic 5′ SSs but also onto cryptic signals to prevent premature cleavage and polyadenylation of the pre-mRNA and ensure transcript integrity [74]. It is currently unclear whether the U1 snRNP directly binds the RNAPII CTD or associates with it via mediating proteins. In fused-in-sarcoma protein (FUS)-knockdown nuclear extracts of human cells, the U1 snRNP can no longer interact with RNAPII, suggesting that FUS acts as a mediating protein in this interaction [71]. In accordance with U1 snRNP



Trends in Genetics

Figure 1. Protein Interaction Maps Connecting Human Splicing-Associated Proteins to (A) the RNA Polymerase II (RNAPII) C-Terminal Domain (CTD) and (B) Chromatin. Spliceosome-associated proteins and splicing mediators are colored blue and green, respectively. RNAPII and chromatin components are colored purple in (A) and (B), respectively. Protein complexes are marked by lighter colors. Interactions that are not mentioned in the text were observed in additional studies [94,118–122].

binding of the 5′ SS, FUS CLIP-seq read density is highest near the 5′ end of long intronic regions [75].

Additionally, it was shown that U2AF65 binds to the phosphorylated RNAPII CTD in humans and proposed that this association is maintained throughout transcription elongation [76] (Figure 1A). The CTD is recognized by U2AF65 while it is associated with the PRP19 complex, which is important for activation of the spliceosome [76]. This is in accordance with previous affinity chromatography and co-immunoprecipitation studies showing that U2AF65 co-purifies with RNAPII in HeLa extract [77,78]. The association between U2AF65 and the RNAPII CTD may explain a report that on calcium treatment and formation of euchromatin structure, the RNAPII elongation rate increases and U2AF65 is depleted from 3′ SSs in human cells [79]. Moreover, the

transcription elongation regulator TCERG1 can interact with U2AF65 [80] and SF1 [81] as well as with the RNAPII CTD simultaneously phosphorylated on the serine 2, 5, and 7 positions [82]. TCERG1 was found to bind independently to elongation and splicing complexes, suggesting that it could couple transcription elongation and splicing by transient interactions [83].

Splicing Factor Recruitment to Chromatin *In Vivo*

Direct and mediated associations between splicing factors and chromatin *in vivo* may also contribute to the spliceosomal identification of short exons located between long introns. An *in vivo* study in HeLa cells has revealed that U2 snRNP associates with chromatin via the chromatin remodeler CHD1, which binds H3K4me3 [84] (Figure 1B). CHD1 and GCN5 are both components of the human histone acetyltransferase STAGA complex [85], which associates with SF3B3, a U2 snRNP component [86]. The STAGA complex promotes the formation of euchromatin structure and both euchromatin and H3K4me3 are active transcription marks. Thus, it could be hypothesized that these interactions play a part in U2 snRNP functional recruitment to transcribed areas. In support of this hypothesis, siRNA-mediated reduction of either CHD1 or H3K4me3 in HeLa cells hinders the association of U2 snRNP with chromatin and decreases splicing efficiency *in vivo* [84].

Nucleosomes associate with exonic sequences more than they do with introns [47,48] and even more so when these exons are flanked by long introns [87]. It is currently unknown whether U2 snRNP, in its entirety, preferentially binds exonic nucleosomes. However, SF3B1, a U2 snRNP component, does preferentially associate with exonic nucleosomes and is highly enriched at exons adjacent to long introns [88] (Figure 1B). The binding of SF3B1 to nucleosomes is also functionally important for the splicing of some alternatively spliced exons [88]. Interestingly, SF3B1 phosphorylation influences pre-mRNA processing [89,90] and is associated with active transcription [90,91]. U2 snRNP components are also associated with heterochromatin protein 1 (HP1) [57] (Figure 1B). HP1 variants are important for maintaining chromatin structure and integrity and enrichment of HP1 at methylated DNA regions has a regulatory effect on splicing [57].

Other splicing factors also associate with chromatin or chromatin remodelers. For example, the nucleosome remodeling complex SWI/SNF interacts with U5 snRNP in human cells [92]. Additionally, BS69, a U5 component, selectively recognizes the H3K36me3 modification (enriched at exonic nucleosomes) and promotes intron retention [93]. Interestingly, mass spectrometry analysis for mononucleosomes also detected an association with U5 snRNP [88]. Finally, U2AF65, which binds the RNAPII CTD [76], also associates with the histone H1 linker [94] as well as with H3K4me3 in HeLa nuclear extracts [84]. Below we hypothesize how chromatin organization and transcription elongation may contribute to exonic recognition by the splicing machinery, especially when exons are located between vastly larger introns.

Model for Co-transcriptional SS Pairing Over Long Introns *In Vivo*

Introns recognized through an intron definition mechanism are under selection to remain short [14,95], presumably as splicing factors find their targets on the mRNA precursor via diffusion. The exon definition model theoretically allows intron lengthening. However, intron lengthening is likely to introduce cryptic SSs that compete with functional ones and longer introns can have RNA secondary structures that may hinder spliceosome assembly [96]. Long introns also reduce the likelihood of stochastic bridging between U1 snRNP bound to the 5′ SS and U2AF35 and U2AF65 bound to the 3′ SS and the polypyrimidine tract, respectively. Finally, several minutes may pass between the synthesis of the 5′ and 3′ SSs of long introns. It is therefore puzzling that the splicing reaction can occur within seconds to minutes after RNA synthesis *in vivo* and that the splicing rate of mammalian introns does not generally decrease the longer the introns are. A mediating mechanism is therefore needed to bring the SSs delimiting a long intron closer for the splicing reaction to occur.

The following description of such a mechanism incorporates two previously proposed models. It has been hypothesized that RNAPII-associated U1 and SR proteins recognize a transcribed exon, resulting in its tethering to RNAPII, and that later the RNAPII-bound U2AF65 interacts with the 3′ SS [76]. Additionally, it has been suggested that the cleaved 5′ SS intermediate RNA of an upstream exon remains tethered to RNAPII that is accumulated over the downstream exon [37]. Combining these models with additional information relating to the associations between chromatin and splicing factors, we propose that the following could be important for the removal of long mammalian introns from the pre-mRNA via an exon definition mechanism (Figure 2). We can assume that elongating RNAPII remains associated with the 5′ SS during the synthesis of a long intron through its interaction with the U1 snRNP, probably until synthesis of the downstream

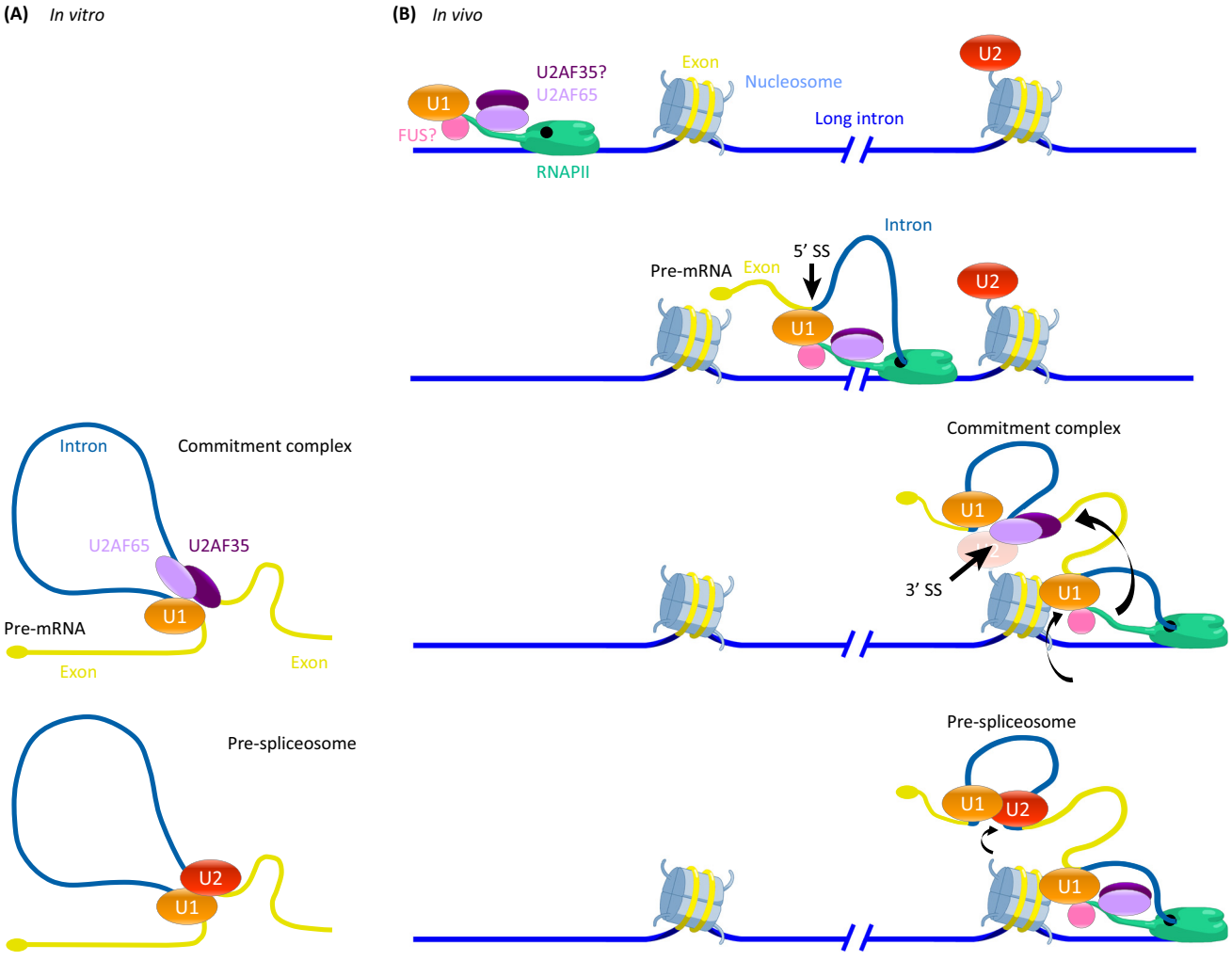


Figure 2. Models for Splicing Complex Assembly *In Vitro* and *In Vivo*. (A) In the *in vitro* system, an mRNA precursor is incubated in HeLa nuclear extract. Splicing in the *in vitro* system is therefore free of chromatin and does not occur concomitantly with transcription by RNA polymerase II (RNAPII). In the commitment complex, the U1 small nuclear ribonucleoprotein (snRNP) binds the 5′ splice site (SS) and is associated via SR proteins (not shown in the panel) with U2AF35, U2AF65, and SF1 (not shown in the panel), which bind the 3′ SS, the polypyrimidine tract, and the branch site region, respectively. In the pre-spliceosome, U2 snRNP binds the branch site. (**B**) Model for SS pairing of long introns *in vivo*. U1 snRNP binds the 5′ SS of the pre-mRNA and remains attached to elongating RNAPII until synthesis of the downstream 5′ SS. RNAPII slows as it nears an exonic nucleosome. An additional U1 snRNP binds the newly transcribed downstream 5′ SS and the RNAPII C-terminal domain (CTD). Concurrently, RNAPII-bound U2AF65 attaches to the 3′ end of the intron, thus bringing the 5′ and 3′ SSs into proximity and allowing the formation of the commitment complex. Next, U2 snRNP detaches from the exonic nucleosome and binds the pre-mRNA branch site to advance the formation of the pre-spliceosome. Other components of the commitment complex and the pre-spliceosome besides the U1 and U2 snRNPs, U2AF65, and U2AF35 are not depicted in the panel.

exon to allow exon definition. Transcription elongation is reduced or rather RNAPII accumulates at the 3′ SS and into the start of the downstream exon while the RNAPII CTD serine 5 is phosphorylated. Next, on synthesis of the downstream 5′ SS an additional U1 snRNP can attach to it. The downstream U1 snRNP could already be bound to the RNAPII CTD or perhaps disassociation of the upstream U1 snRNP from RNAPII at this stage leads to the downstream U1 snRNP taking the upstream U1 snRNP's place on the CTD. Concurrently or immediately after that, RNAPII CTD-bound U2AF65 attaches to the upstream polypyrimidine tract, essentially defining the downstream exon as the splicing unit, bringing the two intron ends closer together, and facilitating the formation of the commitment complex over the intron ends. SR proteins further define the downstream exon by concurrently binding to U1 snRNP and U2AF at both exon ends [97,98]. Subsequently, U2 snRNP, bound to the exonic nucleosome via SF3B1, detaches and binds the pre-mRNA branch site (formation of the pre-spliceosome/A complex). At this stage U2 and U1 snRNPs are located at both exon ends. Thus, RNAPII and nucleosomes plausibly assist in defining exons located between much larger introns. Finally, binding of U5 snRNP components to chromatin may facilitate recruitment of U4 and U6. Binding of the U4–U5–U6 tri-snRNP to the pre-spliceosome complex results in the formation of the B complex.

Concluding Remarks

The aforementioned model can clarify how introns could have expanded by thousands of nucleotides during vertebrate evolution without profoundly hindering splicing [13]. The distance between the 5′ and 3′ SSs would presumably be of lesser importance when splicing factors are uploaded onto nucleosomes that mark exons and when factors bound to RNAPII bring the 5′ and 3′ SSs into close proximity during RNA synthesis. Additionally, such a model may explain how an elongation rate increase during transcription of a gene can have a negative effect on the inclusion level of some of the gene's alternative exons, as high elongation may lead to the synthesis of a downstream exon before the relevant splicing complexes assemble around an upstream exon. Moreover, alternative exons with suboptimal 5′ SSs could potentially be skipped, with the U1 snRNP detaching from RNAPII only once an additional U1 snRNP attaches to the stronger 5′ SS of the downstream constitutive exon. The model could thus provide an explanation for the manner by which 5′ SS mutations at an exon's end may lead to suboptimal binding of the U2 snRNP upstream of the exon [99] and to exon skipping [100] via exon definition. Finally, U1 snRNP attachment to RNAPII throughout transcription elongation could facilitate U1 uploading onto cryptic polyadenylation signals [74] and also explain the observation that U1 snRNP is detected on the transcripts of intronless genes and on transcripts of genes with mutated SSs [69,70].

It should be noted that alternative mechanisms could also facilitate the splicing of exons adjacent to long introns. An undetermined fraction of fly and human splicing events occur via recursive splicing, in which long introns are removed in a stepwise fashion via the use of intermediate intronic sites [101–103]. Recursive splicing has been hypothesized to occur co-transcriptionally for some introns [104]. Furthermore, the alternative splicing factor hnRNP C was found to compact pre-mRNA in a manner that can affect alternative splicing and bring SSs closer [105]. Its extensive binding across introns also competes with U2AF65 and blocks U2AF65 binding to cryptic 3′ SSs [106]. How these mechanisms occur either in opposition to or in accordance with the influence of transcription elongation and chromatin structure on splicing remains to be determined.

In summary, a body of evidence suggests that RNAPII and chromatin organization can constitute scaffolds that facilitate splicing complex assembly over long introns both spatially and temporally. Since the rate of intron removal from the pre-mRNA is unrelated to intron length in mice and humans, we speculate that the proposed mechanism could exist in mammals. In addition, besides being flanked by long introns, we hypothesize that exons identified through this

Outstanding Questions

What specific characteristics do exons identified through the mechanism described here exhibit other than being flanked by long introns, if any?

Generally, we hypothesize that long introns are removed through the aforementioned mechanism, but how many introns are actually spliced in this manner?

Since splicing unit recognition probably occurs via various mechanisms, including the one we propose here, what could lead to one being employed by the cellular machinery and not the other?

When did this splicing mechanism specifically develop during evolution?

What promotes U1 snRNP association with the RNAPII CTD and what would lead to its disassociation?

How is binding of the upstream U2AF65 to the RNA connected with attachment of the downstream U1 snRNP to the 5′ SS, if at all?

Does RNAPII CTD serine 5 phosphorylation play a part in splicing regulation or spliceosome assembly in higher eukaryotes? Are levels of the modification affected by the splicing reaction? Or, perhaps, do co-transcriptional splicing and RNAPII CTD serine 5 phosphorylation simply co-occur?

How do RNAPII interactions with splicing factors affect, if at all, associations between chromatin and splicing factors?

mechanism also display high levels of co-transcriptional splicing [31] and are highly associated with nucleosomes. As an association between exon definition and a specific exon–intron GC content architecture that is enriched for nucleosome occupancy has been detected [87], we hypothesize that exons exhibiting these properties could also be identified via this mechanism. It is important to note that this mechanism may explain the splicing of a subset of exons, whereas a diffusion-based model better fits the splicing of introns of short or intermediate length. Further studies are needed to better characterize the splicing units that are specifically identified through the co-transcriptional exon definition mechanism proposed here, a diffusion-based mechanism, or other mechanisms, what could lead to each mechanism being employed, and the various interactions each mechanism entails (see Outstanding Questions).

References

1. Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336
2. Kalsotra, A. and Cooper, T.A. (2011) Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* 12, 715–729
3. Cartegni, L. *et al.* (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3, 285–298
4. Sibley, C.R. *et al.* (2016) Lessons from non-canonical splicing. *Nat. Rev. Genet.* 17, 407–421
5. Wahl, M.C. and Luhrmann, R. (2015) SnapShot: spliceosome dynamics I. *Cell* 161, 1474–1481
6. Krainer, A.R. *et al.* (1984) Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. *Cell* 36, 993–1005
7. Luhrmann, R. *et al.* (1990) Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim. Biophys. Acta* 1087, 265–292
8. Black, D.L. *et al.* (1985) U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell* 42, 737–750
9. Mount, S.M. *et al.* (1983) The U1 small nuclear RNA–protein complex selectively binds a 5′ splice site *in vitro*. *Cell* 33, 509–518
10. Bell, M.V. *et al.* (1998) Influence of intron length on alternative splicing of CD44. *Mol. Cell. Biol.* 18, 5930–5941
11. Fox-Walsh, K.L. *et al.* (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl Acad. Sci. U. S.A.* 102, 16176–16181
12. Kim, E. *et al.* (2007) Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* 35, 125–131
13. Gelfman, S. *et al.* (2012) Changes in exon–intron structure during vertebrate evolution affect the splicing pattern of exons. *Genome Res.* 22, 35–50
14. Khodor, Y.L. *et al.* (2012) Cotranscriptional splicing efficiency differs dramatically between *Drosophila* and mouse. *RNA* 18, 2174–2186
15. Shcherbakova, I. *et al.* (2013) Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy. *Cell Rep.* 5, 151–165
16. 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
17. Schwartz, S.H. *et al.* (2008) Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* 18, 88–103
18. Irimia, M. *et al.* (2007) Coevolution of genomic intron number and splice sites. *Trends Genet.* 23, 321–325
19. Hawkins, J.D. (1988) A survey on intron and exon lengths. *Nucleic Acids Res.* 16, 9893–9908
20. De Conti, L. *et al.* (2013) Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* 4, 49–60
21. Reed, R. and Maniatis, T. (1986) A role for exon sequences and splice-site proximity in splice-site selection. *Cell* 46, 681–690
22. Reichert, V. and Moore, M.J. (2000) Better conditions for mammalian *in vitro* splicing provided by acetate and glutamate as potassium counterions. *Nucleic Acids Res.* 28, 416–423
23. Wada, Y. *et al.* (2009) A wave of nascent transcription on activated human genes. *Proc. Natl Acad. Sci. U.S.A.* 106, 18357–18361
24. Huranova, M. *et al.* (2010) The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. *J. Cell Biol.* 191, 75–86
25. Schmidt, U. *et al.* (2011) Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *J. Cell Biol.* 193, 819–829
26. Coulon, A. *et al.* (2014) Kinetic competition during the transcription cycle results in stochastic RNA processing. *Elife.* 3, e03939
27. Martin, R.M. *et al.* (2013) Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell Rep.* 4, 1144–1155
28. Saldi, T. *et al.* (2016) Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J. Mol. Biol.* 428, 2623–2635
29. Carrillo Oesterreich, F. *et al.* (2016) Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* 165, 372–381
30. Bhatt, D.M. *et al.* (2012) Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* 150, 279–290
31. Tilgner, H. *et al.* (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 22, 1616–1625
32. Brodsky, A.S. *et al.* (2005) Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol.* 6, R64
33. Adelman, K. and Lis, J.T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* 13, 720–731
34. Kwak, H. *et al.* (2013) Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950–953
35. Veloso, A. *et al.* (2014) Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* 24, 896–905
36. Jonkers, I. *et al.* (2014) Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* 3, e02407
37. Nojima, T. *et al.* (2015) Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* 161, 526–540
38. Mayer, A. *et al.* (2015) Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* 161, 541–554
39. Bird, G. *et al.* (2004) RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3′-end formation. *Mol. Cell. Biol.* 24, 8963–8969
40. Millhouse, S. and Manley, J.L. (2005) The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent

splicing activator in a heterologous protein. *Mol. Cell. Biol.* 25, 533–544

41. de la Mata, M. *et al.* (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* 12, 525–532
42. Fong, N. *et al.* (2014) Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* 28, 2663–2676
43. Kornblihtt, A.R. *et al.* (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165
44. Dujardin, G. *et al.* (2014) How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell* 54, 683–690
45. Jimeno-Gonzalez, S. *et al.* (2015) Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing. *Proc. Natl Acad. Sci. U. S.A.* 112, 14840–14845
46. Gelfman, S. *et al.* (2013) DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon–intron structure. *Genome Res.* 23, 789–799
47. Schwartz, S. *et al.* (2009) Chromatin organization marks exon–intron structure. *Nat. Structural Mol. Biol.* 16, 990–995
48. Tilgner, H. *et al.* (2009) Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* 16, 996–1001
49. Maunakea, A.K. *et al.* (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* 23, 1256–1269
50. Kolasinska-Zwierz, P. *et al.* (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41, 376–381
51. Fuchs, G. *et al.* (2014) Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. *Genome Res.* 24, 1572–1583
52. Luco, R.F. *et al.* (2010) Regulation of alternative splicing by histone modifications. *Science* 327, 996–1000
53. Schor, I.E. *et al.* (2013) Intragenic epigenetic changes modulate NCAM alternative splicing in neuronal differentiation. *EMBO J.* 32, 2264–2274
54. Schor, I.E. *et al.* (2009) Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl Acad. Sci. U.S.A.* 106, 4325–4330
55. Simon, J.M. *et al.* (2014) Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. *Genome Res.* 24, 241–250
56. Zhou, H.L. *et al.* (2014) Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Res.* 42, 701–713
57. Yearim, A. *et al.* (2015) HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Rep.* 10, 1122–1134
58. Iannone, C. *et al.* (2015) Relationship between nucleosome positioning and progesterone-induced alternative splicing in breast cancer cells. *RNA* 21, 360–374
59. Agirre, E. *et al.* (2015) A chromatin code for alternative splicing involving a putative association between CTCF and HP1α proteins. *BMC Biol.* 13, 31
60. Rosonina, E. and Blencowe, B.J. (2004) Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3′-end cleavage. *RNA* 10, 581–589
61. de la Mata, M. and Kornblihtt, A.R. (2006) RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat. Struct. Mol. Biol.* 13, 973–980
62. Hsin, J.P. and Manley, J.L. (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* 26, 2119–2137
63. Bentley, D.L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* 15, 163–175
64. Misteli, T. and Spector, D.L. (1999) RNA polymerase II targets pre-mRNA splicing factors to transcription sites *in vivo*. *Mol. Cell* 3, 697–705
65. Gu, B. *et al.* (2013) CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II *in vivo*. *Nucleic Acids Res.* 41, 1591–1603
66. Tian, H. (2001) RNA ligands generated against complex nuclear targets indicate a role for U1 snRNP in co-ordinating transcription and RNA splicing. *FEBS Lett.* 509, 282–286
67. Das, R. *et al.* (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* 26, 867–881
68. Jobert, L. *et al.* (2009) Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO Rep.* 10, 494–500
69. Spiluttini, B. *et al.* (2010) Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *J. Cell Sci.* 123, 2085–2093
70. Brody, Y. *et al.* (2011) The *in vivo* kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol.* 9, e1000573
71. Yu, Y. and Reed, R. (2015) FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP. *Proc. Natl Acad. Sci. U.S.A.* 112, 8608–8613
72. Harlen, K.M. *et al.* (2016) Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep.* 15, 2147–2158
73. Engreitz, J.M. *et al.* (2014) RNA–RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* 159, 188–199
74. Berg, M.G. *et al.* (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64
75. Masuda, A. *et al.* (2016) FUS-mediated regulation of alternative RNA processing in neurons: insights from global transcriptome analysis. *Wiley Interdiscip. Rev. RNA* 7, 330–340
76. David, C.J. *et al.* (2011) The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65–Prp19 complex. *Genes Dev.* 25, 972–983
77. Robert, F. *et al.* (2002) A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly. *J. Biol. Chem.* 277, 9302–9306
78. Ujvari, A. and Luse, D.S. (2004) Newly initiated RNA encounters a factor involved in splicing immediately upon emerging from within RNA polymerase II. *J. Biol. Chem.* 279, 49773–49779
79. Schor, I.E. *et al.* (2012) Perturbation of chromatin structure globally affects localization and recruitment of splicing factors. *PLoS One* 7, e48084
80. Sanchez-Alvarez, M. *et al.* (2006) Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions. *Mol. Cell. Biol.* 26, 4998–5014
81. Goldstrohm, A.C. *et al.* (2001) The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell. Biol.* 21, 7617–7628
82. Liu, J. *et al.* (2013) Specific interaction of the transcription elongation regulator TCERG1 with RNA polymerase II requires simultaneous phosphorylation at Ser2, Ser5, and Ser7 within the carboxyl-terminal domain repeat. *J. Biol. Chem.* 288, 10890–10901
83. Sanchez-Hernandez, N. *et al.* (2016) The *in vivo* dynamics of TCERG1, a factor that couples transcriptional elongation with splicing. *RNA* 22, 571–582
84. Sims, R.J., III *et al.* (2007) Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol. Cell* 28, 665–676
85. Pray-Grant, M.G. *et al.* (2005) Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* 433, 434–438
86. Martinez, E. *et al.* (2001) Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors *in vivo*. *Mol. Cell. Biol.* 21, 6782–6795
87. Amit, M. *et al.* (2012) Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep.* 1, 543–556

88. Kfir, N. *et al.* (2015) SF3B1 association with chromatin determines splicing outcomes. *Cell Rep.* 11, 618–629
89. Wang, C. *et al.* (1998) Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes Dev.* 12, 1409–1414
90. Bessonov, S. *et al.* (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* 16, 2384–2403
91. Girard, C. *et al.* (2012) Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* 3, 994
92. Batsche, E. *et al.* (2006) The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.* 13, 22–29
93. Guo, R. *et al.* (2014) BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. *Mol. Cell* 56, 298–310
94. Kalashnikova, A.A. *et al.* (2013) Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus. *Nucleic Acids Res.* 41, 4026–4035
95. Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270, 2411–2414
96. Hiller, M. *et al.* (2007) Pre-mRNA secondary structures influence exon recognition. *PLoS Genet.* 3, e204
97. Chen, M. and Manley, J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754
98. Fu, X.D. and Ares, M., Jr (2014) Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701
99. Kuo, H.C. *et al.* (1991) Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science* 251, 1045–1050
100. Krawczak, M. *et al.* (2007) Single base-pair substitutions in exon–intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.* 28, 150–158
101. Duff, M.O. *et al.* (2015) Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature* 521, 376–379
102. Kelly, S. *et al.* (2015) Splicing of many human genes involves sites embedded within introns. *Nucleic Acids Res.* 43, 4721–4732
103. Sibley, C.R. *et al.* (2015) Recursive splicing in long vertebrate genes. *Nature* 521, 371–375
104. Dye, M.J. *et al.* (2006) Exon tethering in transcription by RNA polymerase II. *Mol. Cell* 21, 849–859
105. Konig, J. *et al.* (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* 17, 909–915
106. Zarnack, K. *et al.* (2013) Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* 152, 453–466
107. Reed, R. (2000) Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 12, 340–345
108. Wahl, M.C. and Luhrmann, R. (2015) SnapShot: spliceosome dynamics II. *Cell* 162, 456–461
109. Sharma, S. *et al.* (2014) Stem–loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes Dev.* 28, 2518–2531
110. Will, C.L. and Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, a003707
111. Robberson, B.L. *et al.* (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10, 84–94
112. Romfo, C.M. *et al.* (2000) Evidence for splice site pairing via intron definition in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* 20, 7955–7970
113. Ast, G. (2004) How did alternative splicing evolve? *Nat. Rev. Genet.* 5, 773–782
114. Collins, L. and Penny, D. (2006) Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. Investigating the intron recognition mechanism in eukaryotes. *Mol. Biol. Evol.* 23, 901–910
115. Hoffman, B.E. and Grabowski, P.J. (1992) U1 snRNP targets an essential splicing factor, U2AF65, to the 3′ splice site by a network of interactions spanning the exon. *Genes Dev.* 6, 2554–2568
116. Singh, R. and Valcarcel, J. (2005) Building specificity with non-specific RNA-binding proteins. *Nat. Struct. Mol. Biol.* 12, 645–653
117. Schneider, M. *et al.* (2010) Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol. Cell* 38, 223–235
118. Eick, D. and Geyer, M. (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* 113, 8456–8490
119. Wu, J. *et al.* (2011) Chfr and RNF8 synergistically regulate ATM activation. *Nat. Struct. Mol. Biol.* 18, 761–768
120. Jerebtsova, M. *et al.* (2011) Mass spectrometry and biochemical analysis of RNA polymerase II: targeting by protein phosphatase-1. *Mol. Cell. Biochem.* 347, 79–87
121. Lin, S. *et al.* (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.* 15, 819–826
122. Loomis, R.J. *et al.* (2009) Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol. Cell* 33, 450–461
123. Fiszbein, A. *et al.* (2016) Alternative splicing of G9a regulates neuronal differentiation. *Cell Rep.* 14, 2797–2808