

## Review

## Pre-mRNA splicing and its cotranscriptional connections

Hossein Shenasa<sup>1</sup> and David L. Bentley<sup>1,\*</sup>

**Transcription of eukaryotic genes by RNA polymerase II (Pol II) yields RNA precursors containing introns that must be spliced out and the flanking exons ligated together. Splicing is catalyzed by a dynamic ribonucleoprotein complex called the spliceosome. Recent evidence has shown that a large fraction of splicing occurs cotranscriptionally as the RNA chain is extruded from Pol II at speeds of up to 5 kb/minute. Splicing is more efficient when it is tethered to the transcription elongation complex, and this linkage permits functional coupling of splicing with transcription. We discuss recent progress that has uncovered a network of connections that link splicing to transcript elongation and other cotranscriptional RNA processing events.**

### Splicing and synthesis of mRNA precursors are functionally coupled

Almost all protein-coding genes and many long non-coding RNA genes are transcribed by **RNA polymerase II (Pol II)**; see [Glossary](#)) to generate precursors containing non-coding **intron** sequences that are spliced out and the flanking **exons** ligated with exquisite accuracy. The average human gene contains ~10 introns that make up ~90% of its length. Splicing of precursors in alternative ways generates multiple mRNAs from >95% of human genes [1]. Alternative splicing entails the inclusion or skipping of alternative cassette exons, intron retention, and the selection of alternative **5' splice sites (5' SSs)** and **3' splice sites (3' SSs)** that extend or truncate exons [2].

Splicing is catalyzed by the spliceosome, a dynamic complex comprising five non-coding U-rich small nuclear RNAs (snRNAs) and >170 proteins [3–5]. Classic electron microscopy (EM) showed that splicing can occur on growing transcripts still attached to Pol II [6,7], and subsequent nascent RNA sequencing showed that a large fraction of splicing occurs cotranscriptionally [8–11]. This means that the substrate for the spliceosome is 'work in progress' – a transcript that is being extruded through the Pol II exit channel at rates varying from <0.5 kb/minute to >5 kb/minute [12].

Cotranscriptional splicing occurs close to the transcription elongation complex (TEC), which comprises RNA polymerase II, the nascent transcript, and the chromatin template, as evidenced by the genome-wide association of splicing **U-rich small nuclear ribonucleoproteins (snRNPs)** specifically with intron-containing genes [13]. Other factors that 'travel' with the TEC include RNA-binding proteins (RBPs), elongation factors that control the growth of the RNA chain, and processing factors that carry out cotranscriptional capping, base modification, and 3'-end formation. Splicing is more efficient when it is physically tethered to the TEC [14–17], and this linkage permits functional coupling of splicing with transcription. These coupling mechanisms are of two classes: spatial and kinetic [18]. Spatial coupling is mediated by contacts between splicing factors and the transcription machinery, whereas kinetic coupling links splicing outcomes to the rate of RNA chain synthesis. Coupling mechanisms can work in both directions, meaning that transcription can affect splicing and vice versa. In the following we review recent

### Highlights

The Pol II transcription elongation complex (TEC) makes contacts with the splicing apparatus to facilitate cotranscriptional splicing.

Splicing can occur rapidly once the 3' splice site has been synthesized, even for long introns, suggesting that exon definition is not essential.

The splicing of individual introns is highly plastic, with variation in the kinetics, mode of exon:intron recognition, and the use of recursive splice sites.

Nascent RNA folding into alternative structures is sensitive to the speed of transcription and may exert widespread effects on splicing.

Exon-mediated activation of transcription starts (EMATS) is a means by which splicing of an alternative exon feeds back onto transcription of the same gene to activate weak promoters and select alternative mRNA 5' ends.

<sup>1</sup>Department of Biochemistry and Molecular Genetics, RNA Bioscience Initiative, University of Colorado School of Medicine, PO Box 6511, Aurora, CO 80045, USA

\*Correspondence: [david.bentley@ucdenver.edu](mailto:david.bentley@ucdenver.edu) (D.L. Bentley).

research on the connections linking splicing to elongation of the transcript and other cotranscriptional RNA processing events.

### Heterogeneous pathways of cotranscriptional splicing

Spliceosome assembly and activation (Box 1) are highly regulated and vary greatly in efficiency between introns in different biological contexts. In addition to SS and **branch point (BP)** sequences that are recognized by the core spliceosome [19,20], splicing is modulated by diverse RBPs including **serine and arginine-rich (SR) proteins** and **heterogeneous nuclear ribonucleoproteins (hnRNPs)**. These factors exert context-dependent positive and negative effects on spliceosome function through cotranscriptional binding to exonic and intronic splicing enhancer and silencer elements [2,21,22]. How splicing factors interact with the TEC to regulate spliceosome assembly and remodeling is an important unresolved question.

*In vitro* splicing reactions and spliceosomal cryo-electron microscopy (cryo-EM) structures are reconstituted with synthetic RNA substrates that often have a single short intron and flanking exons. Splicing complexes assembled on these substrates are uncoupled from transcription. By contrast, natural pre-mRNA substrates are far more complex and dynamic because they comprise growing RNA chains still attached to Pol II and contain multiple exons and introns that can be many kilobases in length. Commitment to pairing of 5' and 3' SSs is made during the formation of the A complex (Figure 1) [23], but how proper pairing is achieved cotranscriptionally, when multiple 5' and 3' SSs may be present on a nascent transcript, remains poorly understood. At the commitment step, bridging occurs between U1 snRNP at the 5' SS and U2 snRNP and **U2 auxiliary factor (U2AF)** at the 3' SS through both direct and indirect contacts mediated by a network of SR proteins [24–27]. Bridging between 5' and 3' SSs can be made in a single step across an intron to make an intron definition complex, or in a two-step process where initial cross-exon contacts form an exon definition complex that transitions to the intron definition complex [27,28]. U1 snRNP occupancy at both 5' SSs flanking a 3' SS synergistically stimulates splicing, suggesting that cross-exon and cross-intron cooperation can promote a single splicing event [29]. Structural modeling suggests that the yeast A complex is equally compatible with cross-exon and cross-intron contacts between U1 and U2 snRNPs [30] (Figure 2).

It is usually thought that intron definition predominates when introns are short (<250 nt) whereas exon definition predominates when introns are long, which is common in mammals [31,32]. However, recent studies of cotranscriptional splicing have challenged this view. Sequencing methods have been developed for chromatin-associated RNA that can simultaneously identify the 3'-OH end of the RNA, that defines the position of the RNA polymerase, as well as the splicing status indicated by the presence of exon:exon junctions or lariat BPs located upstream of the 3' end. The surprising finding of these studies is that both lariat formation and exon ligation can be completed when Pol II has extended as few as 18–26 bases past the 3' SS (Figure 2, upper panel) [33–35]. Using cotranscriptional lariat sequencing (CoLa-seq) to map the 3' end and BP of lariat–exon2 splicing intermediates, Zeng and colleagues showed that, in human cells with long introns, step 1 often occurs

#### Box 1. Spliceosome assembly and catalysis

Splicing begins with base-pairing of the 5' SS to the U1 snRNA in U1 snRNP, and recognition of the branch point (BP) adenosine and the 3' SS by SF1 and U2AF, respectively, to generate the E complex (see Figure 1 in main text). U2 snRNP then binds to the BP in an ATP-dependent step through base-pairing with U2 snRNA and establishes contact with U1 snRNP to form the A complex where 5' and 3' SSs are selected. Subsequent recruitment of the U4/6 U5 tri-snRNP and rearrangement ejects the U1 snRNP which is replaced at the 5' SS by U6 snRNA. Base-pairing between U6 and U2 snRNAs creates an active site which catalyzes the first transesterification reaction, step 1 (Figure 1). Step 1 creates two products: exon1 cleaved at the 5' SS and the intron/exon2 lariat intermediate. Further remodeling converts the active site to the exon ligation conformation that catalyzes step 2 transesterification. This reaction joins the exons and releases the intron as a lariat (Figure 1). Reviewed in [3,5,150].

#### Glossary

**Branch point (BP):** a short sequence usually containing an A base located 15–50 nt upstream of the 3' SS that is recognized by base-pairing with U2 snRNA. The 2'-OH of the BP A base attacks the 5' SS in step 1 to create an intronic lariat intermediate.

**5' Cap:** the 5'-triphosphate end of the primary transcript is converted cotranscriptionally to an  $N^7$ -meG(5')ppp (5)-2'-O-meN cap structure by removing the terminal phosphate, GMP transfer,  $N^7$ -methylation (me) of the terminal G, and 2'-O-methylation of the first transcribed base. The cap-binding complex (CBC) facilitates splicing of the first exon.

**Cleavage/polyadenylation (CPA):** mRNA 3'-end formation occurs by endonucleolytic cleavage (by CPSF73) of the nascent transcript ~20 nt 3' of a polyadenylation signal (AAUAAA) and subsequent addition of a non-templated poly(A) tail by poly(A) polymerase. CPA is functionally coupled to transcription termination.

**Exon:** a sequence (often coding) within the pre-mRNA that is included in the mRNA. Exons are ligated together in step 2 of splicing facilitated by the spliceosome.

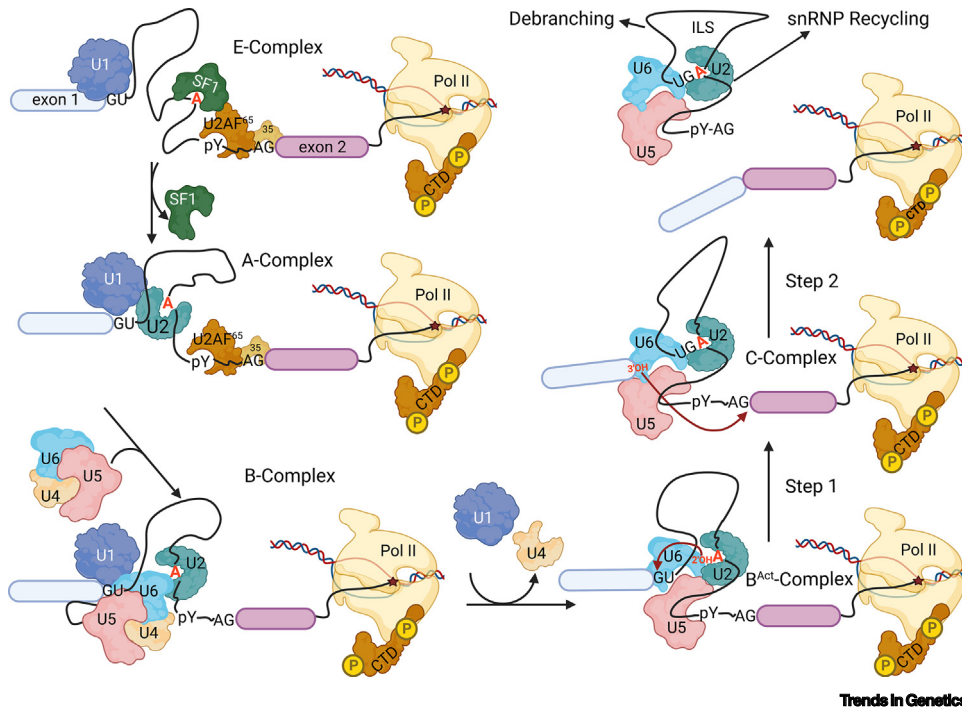
**Heterogeneous nuclear ribonucleoproteins (hnRNPs):** a class of RNA-binding proteins (RBPs) that can affect alternative splicing by binding to exonic splicing silencers and modulating spliceosome assembly.

**Intron:** a transcribed (generally non-coding) sequence between exons that is excised by the spliceosome as a lariat RNA.

**Polypyrimidine tract-binding proteins (PTBPs):** the two PTBPs are RBPs that bind to the pre-mRNA upstream of 3' SSs. They can compete with U2AF and regulate alternative splicing

**RNA polymerase II (Pol II):** one of the three nuclear RNA polymerases. Pol II transcribes all protein-coding and many non-coding genes. It has a signature C-terminal domain (CTD) on its large subunit that comprises intrinsically disordered heptad (YSPTSPS) repeats.

**Serine and arginine-rich (SR) proteins:** a class of conserved RBPs that facilitate early spliceosome assembly and other biological processes. SR proteins have RNA recognition motifs (RRMs) and phosphorylated RS domains with arginine (R)–serine (S) dipeptide repeats.



**3' Splice site (3' SS):** the site of RNA cleavage, usually after an AG at the end of the intron. The AG is preceded by a polypyrimidine-rich sequence (Py). The Py is recognized by U2AF<sup>65</sup> and the AG is contacted by U2AF<sup>35</sup>.

**5' Splice site (5' SS):** the site of RNA cleavage between the end of an exon and beginning of an intron which usually starts with the sequence GU. A loosely conserved 9 nt sequence around the 5' SS is recognized by base-pairing with U1 snRNA.

**U2 auxiliary factor (U2AF):** a heterodimer of 35 kDa and 65 kDa subunits that serves to recognize functional 3' SSs early in spliceosome assembly. U2AF<sup>65</sup> contacts the Py tract and U2AF<sup>35</sup> contacts the AG of the 3' SS.

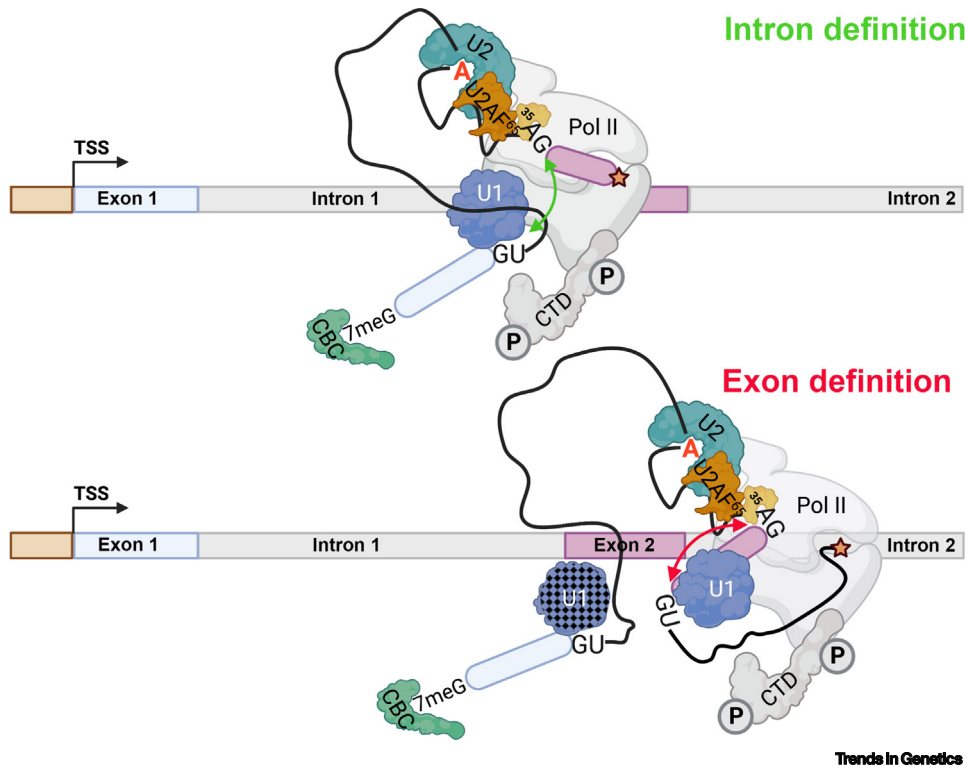
**U-rich small nuclear ribonucleoproteins (snRNPs):** U1, U2, U4, U5, U6 snRNPs are core splicing factors that contain unique non-coding uridine-rich RNAs and a set of common and unique protein subunits.

Trends in Genetics

**Figure 1. Cotranscriptional spliceosome assembly on nascent pre-mRNA attached to RNA polymerase II (Pol II).** U1 snRNP binds to the 5' SS (GU), SF1 binds to the branch point A (red), and U2AF<sup>65/35</sup> subunits bind to the polypyrimidine tract (pY) and the 3' SS (AG) to form the E complex on the nascent transcript. A star indicates the 3'-OH growing end of the RNA. Note that interactions between Pol II and splicing components including helicases that mediate rearrangements within the spliceosome are poorly understood and many are not shown in this simplified depiction. U2 snRNP replaces SF1 to form the A complex. Recruitment of the U4/U6.U5-tri-snRNP forms the pre-catalytic B complex. Structural rearrangements then release U1 and U4 snRNPs to form the activated B complex (B<sup>Act</sup>). B<sup>Act</sup> catalyzes step 1 in which the 2'-OH of the bulged branch point A attacks the 5' SS, creating an upstream exon with a free 3'-OH and an intron lariat attached to the second exon. Further rearrangements form the C complex that catalyzes step 2 in which exons are ligated and the intron lariat structure (ILS) is excised and is subsequently debranched and degraded. Putative protein:protein contacts between Pol II and splicing factors are not shown for clarity. Abbreviations: CTD, C-terminal domain; P, phosphorylation; snRNP, small nuclear ribonucleoprotein; SS, splice site; U2AF, U2 auxiliary factor.

in an 'ultrafast' fashion before transcription of the downstream exon is complete, thereby precluding cross-exon interactions required for exon definition. This work suggests that exon definition, which predominates in the splicing of long introns *in vitro* uncoupled from transcription, may be less favored when splicing occurs cotranscriptionally *in vivo*. Additional work will be necessary to establish exactly how important exon definition is in the context of cotranscriptional splicing.

Although 'ultrafast' splicing can occur at many yeast and human introns, the average cotranscriptional splicing event is probably much slower, and is not completed until Pol II has transcribed hundreds or thousands of bases beyond the end of the intron [34,36,37]. Indeed, live cell imaging of individual genes indicates that, before they are removed, unspliced introns can persist for many minutes after transcription of the intron is complete – by which time the polymerase has presumably traveled far downstream [38]. That most splicing occurs after polymerase has transcribed well past the 3' SS is also supported by metabolic labeling of RNA for short periods in budding yeast followed by sequencing to measure the kinetics of accumulation of unspliced precursors, step 1 lariat intermediates, and fully spliced products [37]. This study showed rates of splicing vary from 30 s to 15 minutes for different introns, and that the median



Trends in Genetics

Figure 2. Models for the recognition of 5' and 3' SS by intron and exon definition with U1 snRNP bound to RNA polymerase II (Pol II). (Upper panel) An intron looping/scanning model of cotranscriptional spliceosome assembly and intron definition [53]. Tethering of the 5' SS by U1 snRNP bound to Pol II causes intron looping that facilitates recognition of the BP and 3' SS by U2 snRNP and U2AF<sup>65/35</sup> by 'scanning', thereby promoting the formation of cross-intron contacts (green arrows) and 'ultrafast' splicing [33,34] before the synthesis of exon 2 is complete. Pol II-associated U2AF<sup>65/35</sup> could participate in scanning. (Lower panel) Pol II bound U1 snRNP could function in exon definition through the formation of cross-exon contacts (red arrows) with U2 snRNP and U2AF. An exchange of U2 snRNP and/or U2AF contacts between the downstream U1 snRNP (blue) and the upstream U1 snRNP (stippled) on the Pol II surface could mediate the transition to an intron definition complex (upper panel). In an alternative pathway, this transition may occur by direct binding of U6 to the upstream 5' SS [27,28]. The Cap-binding complex (CBC) is bound to the 7meG cap that is added cotranscriptionally. Abbreviations: BP, branch point; CTD, C-terminal domain; 7meG, N<sup>7</sup>-meG(5')ppp(5')-2'-O-meN cap structure; P, phosphorylation; snRNP, small nuclear ribonucleoprotein; SS, splice site; U2AF, U2 auxiliary factor.

half-life for conversion of a precursor to fully spliced product is ~2 minutes (~80 s for step 1 and 40 s for step 2), during which time polymerase, at its normal speed, would reach the end of most yeast genes. Interestingly, splicing is faster for introns with longer distances between the 3' SS and the end of the gene, consistent with the idea that it is facilitated by longer engagement of the spliceosome with the TEC [37].

Not only is the timing of cotranscriptional splicing highly variable, but so are the actual SSs used to remove a given intron. This surprising conclusion came from imaging splicing in live cells where individual introns were marked by insertion of hairpin loops at known positions to permit visualization using fluorescently tagged RBPs. Loss of the fluorescent signal at the site of transcription acts as a surrogate for excision of the intron. Remarkably, the labeled intron elements often only persisted at the site of transcription for short periods that were not long enough for transcription of the whole intron to be completed [38]. The explanation of this conundrum is that introns can be excised piecemeal by stochastic splicing to different zero-length pseudoexons comprising juxtaposed 3' and 5' SSs (AGGT) within introns. This process regenerates SSs at the junctions that are

subsequently used in another round of recursive splicing (RS) until the entire intron is removed [39]. Live-cell imaging suggests that cotranscriptional intron removal by RS occurs far more frequently than was previously suspected [38], probably because RS intermediates are unstable and are therefore rarely detected by RNA-seq. On the other hand, if RS is in fact a common event, it remains unclear how it evades suppression of 5' SSs by the exon junction complex that is rapidly deposited at splice junctions [40,41]. In summary, splicing of individual introns does not follow a unique predetermined or hard-wired pathway. Instead, splicing is achieved in multiple ways: cotranscriptionally or post-transcriptionally, cotranscriptionally immediately after 3' SS synthesis or with a delay, using either exon or intron definition, and by classical two-step splicing or multiple recursive steps. Cotranscriptional splicing therefore appears to be highly non-uniform, raising the question of whether regulatory mechanisms modulate the choices between alternative pathways that can be used to remove a given intron.

### The U1 snRNP–Pol II connection and spatial coupling

Functional coupling of splicing with transcription likely involves a physical interaction between Pol II and splicing factors including U2AF, Prp19 [42], and SR proteins [43,44]. Such interactions involve contacts with both the Pol II body and its C-terminal domain (CTD) – a unique feature of Pol II that is not found on other RNA polymerases. The CTD is an intrinsically disordered region (IDR) composed of conserved heptad repeats (Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub>) that are reversibly phosphorylated at multiple positions during transcription [45]. Truncation of the CTD inhibits splicing [46] *in vivo*, and *in vitro* the phosphorylated CTD can enhance splicing [47]. Furthermore, inhibition of the Ser5 CTD kinase CDK7 causes widespread disruption of splicing [48]. Splicing factors including U2AF and U1 snRNP copurify preferentially with Pol II bearing the Ser5 phosphorylated isoform of the CTD [42,49,50] which can form phase-separated condensates where splicing factors are concentrated [51]. Ser5 phosphorylation is strongly enriched at the 5' ends of genes, but how this bias affects splicing is unknown. The CTD could serve as a type of 'antenna' that traps splicing factors at the site of transcription, or it could allosterically stimulate specific steps in spliceosome assembly or both. Although the inhibition of splicing by CTD truncation and other Pol II mutations demonstrates that cotranscriptional splicing is functionally important, it remains to be determined what fraction of introns normally require cotranscriptional splicing for proper expression of the genes in which they reside.

The best-characterized interaction of a splicing factor with the TEC is between U1 snRNP and Pol II [52] that was described in a landmark cryo-EM structure [53]. U1 snRNP contacts the back side of Pol II, near the RNA exit channel where the U1-70K subunit approaches conserved residues in the Pol II Rpb2 and Rpb12 subunits. Genetic studies will be necessary to establish the physiological significance of the protein:protein contacts between Pol II and U1 snRNP in the structure, but notably it predicts that U1 snRNA can base-pair with a newly synthesized 5' SS (Figure 2). However, formation of the U1 snRNP–Pol II complex does not require an RNA transcript. Direct contact of U1 snRNP with Pol II is consistent with the recently reported stimulation of transcription elongation by this snRNP [54]. The U1 snRNP–Pol II complex is also consistent with the localization of U1 (unlike other snRNPs) at transcribed genes independently of splicing [55–57] – both at start sites and within gene bodies [57–60]. Furthermore, if Pol II is stalled part way through an intron by using targeted Cas9 as a roadblock, then the 5' SS of that intron copurifies with Pol II [58], consistent with tethering of the 5' SS–U1 snRNP complex to Pol II as in the cryo-EM structure of Zhang *et al.* [53].

U2 snRNP can be modeled into the U1 snRNP–Pol II complex [53], suggesting that cross-intron or cross-exon interactions could form in close proximity to the TEC (Figure 2). An intriguing possibility suggested by Zhang and colleagues is that, as an intron is transcribed, it forms an

expanding loop in which the growing end of the transcript in the Pol II active site is at one end, and the 5' SS captured by U1 snRNP bound to Pol II is at the other end [53] (Figure 2). In this scenario the TEC might function as an intron 'scanner' that promotes formation of cross-intron complexes. It would presumably be advantageous for the 'scanner' to recognize 3' SSs as well as 5' SSs, and U2AF might function in this role. U2AF copurifies with Pol II and the U2AF65 subunit appears to be handed off from Pol II to the RNA transcript [42,61,62] (Figure 2). Rapid intron definition facilitated by intron looping/scanning may permit 'ultrafast' splicing before the exon is fully transcribed (Figure 2). Whether exon and intron definition are equally feasible in the cotranscriptional context, and whether U1 snRNP bound to Pol II participates in both recognition pathways, remains to be seen. Given that only one U1 snRNP can contact Pol II at a time, cotranscriptional exon definition and subsequent SS pairing would presumably require an exchange on the Pol II surface between U1 snRNP bound to the downstream 5' SS (non-stippled in Figure 2) and U1 snRNP bound to the upstream 5' SS (stippled in Figure 2).

In summary, the U1 snRNP–Pol II structures have galvanized thinking about cotranscriptional splicing and have raised interesting questions for future investigation (see [Outstanding questions](#)). For example, is there a mechanism, perhaps a transcriptional pause, that ensures U1 snRNP is in place on the Pol II surface ready to capture a 5' SS when it emerges from the RNA exit channel, and could U1 snRNP association with Pol II affect its interaction with the 5' SS? If so, this could have significant consequences for the function of U1 snRNPs in discriminating between genuine and cryptic 5' SSs, as well as in regulating 5' SS recognition by enhancer and silencer elements [63–65].

### Determinants of cotranscriptional splicing

Transcription imposes an order and timing on the synthesis of SS and RBP binding sites, whereas on a full-length transcript they are all presented simultaneously. This distinction means that different regulatory mechanisms may operate on cotranscriptional and post-transcriptional splicing. An important challenge is to uncover how cotranscriptional splicing efficiency [i.e., the fraction of transcripts where splicing of a particular intron is completed before release from the polymerase by cleavage at the poly(A) site] is controlled. Nascent RNA sequencing has identified features that correlate with efficient cotranscriptional splicing, including slow transcription which lengthens the window of opportunity for splicing before transcription terminates [66,67], structured RNA at SSs [67], and strong U2AF binding sites [34]. Cotranscriptional and post-transcriptional splicing take place in distinct subnuclear environments where regulatory proteins could operate differently. This possibility is demonstrated by recent work on the **polypyrimidine tract-binding protein (PTBP)** 1 that is best known as a repressor of splicing at specific introns which was discovered by examination of mature mRNAs [68,69]. Analysis of nascent RNA, however, revealed that PTBP1 has an opposite function in promoting cotranscriptional splicing of a distinct group of introns [70]. PTBP1 therefore appears to have a predominantly negative effect on splicing when acting post-transcriptionally and a positive effect when acting cotranscriptionally.

Ongoing transcription is an important influence on the order in which introns are removed, as well as on coordination between the splicing of different introns in the nascent transcript. In general, introns near 5' ends are removed before those at 3' ends – as predicted by the 'first come first served' model [71]; however, the order in which introns are removed is by no means strictly determined by the order in which they are transcribed [72]. Among adjacent pairs of human introns, upstream introns are preferentially spliced out first only about half the time [36], and, within a pair, one is often 'always first' [72]. Adjacent splicing events are often interdependent, resulting in clusters of introns with similar cotranscriptional splicing efficiencies [32,36]. In the most extreme cases, nascent transcripts from a given gene have all-or-none splicing of the introns [73,74].

The all-unspliced transcripts are also poorly processed at poly(A) sites, reflecting the coupling between these processing steps (see following section). It is not known whether the all-unspliced transcripts are eventually processed into mRNA or are dead-end products. In summary, cotranscriptional splicing efficiency is determined by mechanisms that operate on individual introns, on groups of neighboring introns, and even on whole transcripts.

### The influence of cotranscriptional RNA folding on splicing

Folding of a growing RNA chain is a highly dynamic process that has recently been revealed *in vivo* by chemical probing methods [75]. Nascent RNA folding is sensitive to transcription speed: slow transcription favors base-pairing of more proximal elements whereas fast transcription favors pairing of more distal elements to form more open structures [76]. The alternative RNA conformations assumed by a growing transcript are most accurately described as ensembles of multiple local energy minima that govern their probabilities of formation [77,78]. The ensemble of alternative structures for a particular sequence element changes as the RNA chain grows in a way that is sensitive to its rate of growth and to transcriptional pausing [79,80]. Even the addition of a single nucleotide to a growing RNA chain can instigate a dramatic change in structure through cotranscriptional strand displacement [81]. Different RNA structures assumed by the pre-mRNA influence constitutive and alternative splicing by modulating the proximity and accessibility of SSs and RBP binding sites [82–85]. SSs and binding sites for RBPs must be presented for recognition in a suitable conformation that is usually single-stranded [86]. Indeed, extensive RNA unfolding of structures around SSs is predicted to occur within the spliceosome [87], and sequestration of SSs within structures has long been known to reduce splicing efficiency [88–90]. The relevance of nascent RNA structure was recently highlighted by a study of disease-causing mutations in the *MAPT* gene encoding the microtubule-associated Tau protein. These mutations change inclusion of alternative exon 10 in ways that correlate with their effects on the ensemble of structures around the 5' SS of that exon [87].

Alternative splicing is slower and more frequently completed post-transcriptionally than constitutive splicing [8,9,72], in part because of poorly defined mechanisms that delay the splicing of introns that flank alternative exons [72]. However, splicing reactions that are completed post-transcriptionally may still be affected by earlier cotranscriptional events. One example occurs at alternative cassette exons that are sensitive to transcription speed. When slow transcription favors inclusion of these exons, it correlates with reduced RNA structure at the 3' SS downstream of the cassette exon; conversely, when slow transcription favors exon skipping, RNA structure at the downstream 3' SS is elevated [67]. Hence cotranscriptional RNA folding appears to influence the outcome of alternative splicing that is completed post-transcriptionally. In summary, the outcome of both cotranscriptional and post-transcriptional splicing is likely to depend on the rapidly changing alternative structures assumed by the nascent transcript as it grows.

### Splicing and cotranscriptional RNA modifications

Splicing is one of several interdependent cotranscriptional pre-mRNA processing steps that include 5' capping, covalent nucleotide modification, and **cleavage/polyadenylation (CPA)**. The **5' cap** is added shortly after the 5' end emerges from the RNA exit channel by capping enzymes that directly contact Pol II [91–93]. The cap is recognized by nuclear cap-binding complex (CBC, Figure 2) which promotes recognition of the exon 1 5' SS by U1 snRNP and subsequent handoff to U6 [94–97].

As the nascent transcript grows it becomes partially modified at specific nucleotides in introns and exons, and some of these covalent marks affect splicing. One of the most widely distributed covalent modifications in nascent RNA is pseudouridine. Knockout of pseudouridine synthase

PUS1 alters the inclusion of >2000 alternative cassette exons [98]. This widespread effect of pseudouridylation on splicing is mediated at least in part by modification of RBP binding sites [98], but it could also work by enhancing the stability of RNA duplexes [99].

$N^6$ -methyladenosine (m6A) – a modification deposited cotranscriptionally by METTL3/METTL14 [100] – alters splicing by multiple mechanisms that affect RBP binding and RNA folding. m6A deposition in exons close to splice junctions is associated with rapid constitutive splicing, whereas in introns it is associated with slower splicing typical of alternative cassette exons [100]. m6A destabilizes RNA structures in a way that facilitates binding of the splicing regulators hnRNPC and hnRNPG, and thereby affects the inclusion of many alternative exons [101,102]. RNA binding by the m6A reader YTHDC1 can also influence splicing by selective recruitment of SR proteins to the transcript [103].

Folding of the nascent transcript generates duplexes, often generated by pairing between oppositely oriented repeat elements in introns, and these structures serve as substrates for ADAR1 and ADAR2 (adenosine deaminase acting on RNA) that cotranscriptionally convert adenosine to inosine [104]. A–I conversions can affect SSs and more commonly regulatory elements, and thereby influence numerous alternative splicing decisions [105]. Conversely, reduced splicing enhances intronic A–I editing probably by lengthening the window of opportunity for making these cotranscriptional modifications [106]. Enzymatically inactive ADAR1 also affects splicing, presumably by binding to duplexes and stabilizing them or excluding other factors [105].

### Connections between splicing and cleavage/polyadenylation

There is a complex relationship between cotranscriptional splicing and 3'-end processing by CPA. Splicing of the last intron and processing at the poly(A) site are mutually interdependent, although the mechanisms responsible for this coupling are not well understood. Poly(A) site recognition facilitates splicing of the last intron, probably because it defines the 3' end of the last exon, and mutation of the poly(A) site inhibits terminal intron splicing [107,108]. Conversely, recognition of the last intron by splicing factors facilitates processing at the poly(A) site that defines the end of the last exon [109–112]. Coupling between splicing and 3' processing at the end of the gene likely involves cooperative interactions between U2snRNP, U2AF, and CPA factors [113,114].

By contrast, at cryptic poly(A) sites within introns, there is strong antagonism between splicing and CPA that plays a major role in preventing premature truncation of mRNAs [115,116]. There are two models to account for repression of intronic poly(A) sites by splicing: kinetic competition and direct interference. According to the kinetic competition model, if splicing removes an intron harboring a cryptic poly(A) site before that site is processed, then premature 3'-end formation will be averted. This model is supported by the finding that lengthening of an intron and hence increasing the delay until it is spliced out can enhance use of an intronic alternative poly(A) site [117]. By the same token, accelerating splicing by strengthening a 5' SS can suppress use of an intronic poly(A) site [118]. Similarly, inhibition of splicing by pladienolide B (Plad B), which blocks U2 snRNP function, can enhance intronic polyadenylation and coupled transcription termination [35]. The interference model postulates that U1 snRNP bound to a 5' SS downstream of a poly(A) site can prevent cleavage at that site via inhibitory interactions with the CPA machinery [119,120]. Depletion of U1 snRNA stimulates premature CPA globally, suggesting that U1 snRNP normally safeguards transcripts against truncation of their 3' ends in a process dubbed 'telescripting' [116]. It is also possible that U1 depletion might derepress premature CPA by slowing down splicing, as predicted by the kinetic competition model. Another potential interference mechanism might be that U1 snRNP binding to Pol II competes with the recruitment of CPA factors that are thought to perform poly(A) site cleavage in very close proximity to Pol II [121,122].



Whether a poly(A) site is subject to suppression by the splicing machinery appears to depend on whether it is situated in an intron or not. The importance of an intronic versus exonic context for poly(A) site recognition is demonstrated by examples in which an entire gene is embedded within an intron of a host gene. In these cases the same poly(A) site is processed when it is transcribed as the last exon of the nested gene, but processing is prevented when it is transcribed as one of the host gene introns [123]. This example suggests that the TEC somehow 'knows' whether the poly(A) site is located in an intron, and is unable to support CPA, or it is located in an exon, and is competent for CPA. The distinction might correspond to different factors associated with the TEC when it is transcribing an exon compared to an intron, and this might depend on its cotranscriptional splicing history.

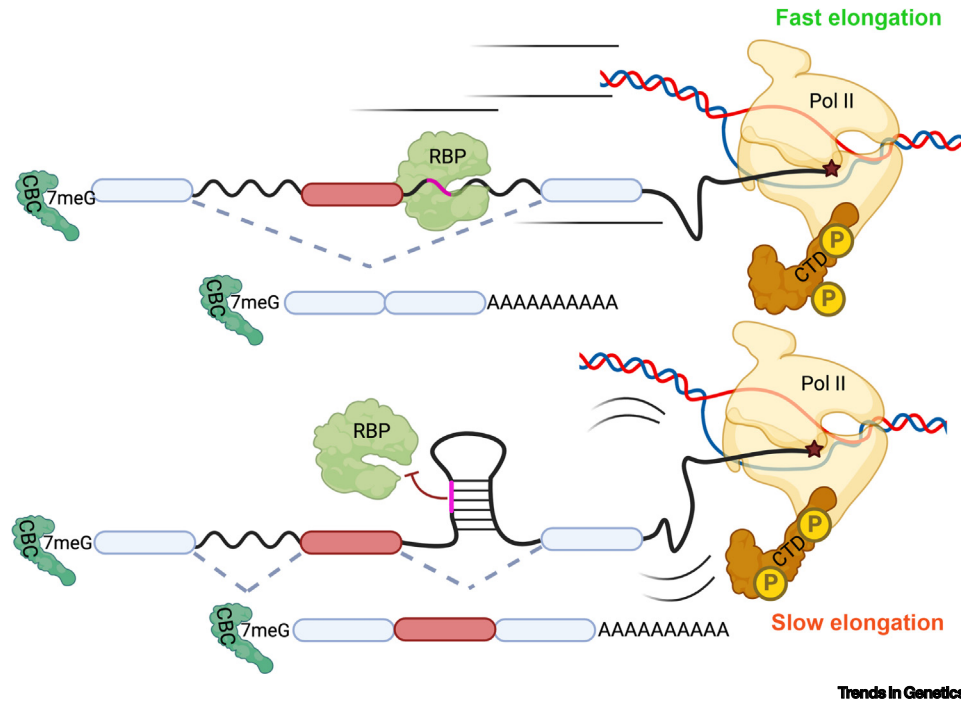
### Splicing connections with transcription initiation, elongation, and chromatin modification

The average speed of transcript elongation by Pol II varies from <0.5 kb/minute to >5.0 kb/minute both between genes and within genes [12]. Transcript elongation comprises short intervals of relatively rapid growth punctuated by frequent pauses where the polymerase may even backtrack. Pausing is more frequent, and transcription is slower, in exons that have higher GC content than introns [124,125]. Slower elongation in exons, or pausing at specific positions in the vicinity of SSs, could affect cotranscriptional splicing, possibly by facilitating timely recruitment of splicing factors to the TEC [126]. High-resolution mapping of Pol II shows that pausing does not generally occur at 3' SSs, at least in metazoans [74,127]. Whether pausing occurs near 5' SSs remains unresolved, in part because contaminating step1 intermediates with a free 3'-OH end at the SS confound the analysis. It is intriguing to note, however, that the consensus sequence of mammalian Pol II pause sites is GT at the pause and +1 positions, in common with the first two nucleotides of most introns [127].

Changes in average transcription speed can have profound effects on alternative and constitutive splicing [66,128–130]. Pol II mutants that slow down or speed up elongation affect numerous alternative splicing decisions, but we cannot predict whether decelerating or accelerating transcription will promote exon inclusion or skipping [128,131,132]. It is likely that transcription speed affects splicing in multiple ways that include changing the time delay between the synthesis of competing SSs or RBP binding sites, as well as altering how the transcript folds (Figure 3) and how it is covalently modified. Both cotranscriptional m6A modification and A-I editing are sensitive to transcriptional speed, and slow elongation increases the level of these modifications [67,133].

In addition to transcriptional effects on splicing, recent studies suggest that splicing feeds back on both transcription elongation and initiation. U1 snRNP contacts with Pol II clash with Rtf1, an important allosteric activator of elongation [134,135], suggesting that U1 snRNP binding might slow down transcription. Inhibition of U2 snRNP by Plad B results in the accumulation of Pol II near the 5' end of the gene at the promoter-proximal pause site, consistent with impaired elongation [136,137]. It is unclear how inhibition of U2 snRNP engaged at the BP of an intron transcribed by one Pol II could affect elongation by another Pol II near the TSS [136], but notably the TAT-SF1 (Cus2 in yeast) subunit of U2 can regulate transcription elongation [138,139]. One possibility is that Plad B affects transcription by inhibiting the release of TAT-SF1 from the U2 snRNP when it engages the BP [140].

Remarkably, inclusion of an alternative exon near the 5' end of a gene can activate transcription initiation at an alternative start site within a few kilobases upstream [141], thereby creating an alternative first exon which is a major generator of transcript diversity [142]. This phenomenon of



**Figure 3. Potential mechanisms by which transcription speed affects cotranscriptional splicing.** RNA polymerase II (Pol II) speed influences cotranscriptional RNA folding, which can in turn promote or repress the binding of RNA-binding protein (RBPs). The RNA structural landscape and RBP interactome in turn influence constitutive and alternative splicing decisions. Fast elongation (upper panel) favors less local RNA folding and could thereby favor binding of RBPs with single-stranded binding sites. In this example, the RBP enhances skipping of the alternative exon (red). Slow elongation (lower panel) favors more local RNA structure that can reduce RBP binding, resulting in exon inclusion. Abbreviations: CBC, cap-binding complex; CTD, C-terminal domain; 7meG, *N*<sup>7</sup>-meG(5')ppp(5')-2'-O-meN cap structure; P, phosphorylation;

exon-mediated activation of transcription start sites (EMATS) is thought to be mediated through splicing-dependent recruitment of transcription initiation factors [141,143,144]. The link between alternative splicing and transcription initiation may have significant consequences for therapeutics that affect alternative exon inclusion, such as the spinal muscular atrophy drug nusinersen – an antisense oligonucleotide that increases inclusion of *SMN2* exon 7 by masking a splicing silencer. Increased exon inclusion may therefore not only alter mRNA coding capacity but also stimulate transcription through EMATS [144].

Splicing factors operate in the vicinity of chromatin where they have the opportunity to engage in crosstalk with histones; indeed, changes in splicing correlate with the deposition of specific histone variants [145] and covalent histone marks [146,147]. In particular, histone H3 trimethylated on lysine 36 (H3K36me3) is cotranscriptionally deposited by SETD2 preferentially on nucleosomes within exons, in a way that correlates with splicing activity [146,148]. An unintentional side effect of altering splicing with antisense oligonucleotides such as nusinersen can be the establishment of a repressive chromatin mark, histone H3 dimethylated on lysine 9 (H3K9me2), close to the target exon. This observation has suggested new therapeutic strategies that target both splicing and chromatin modification in a synergistic way [149]. In summary, recent research has revealed a web of interactions that link splicing of the nascent transcript with chromatin modification and with transcription initiation and elongation. Working out exactly how these interactions function in regulated mRNA biogenesis remains an important challenge for the future.

### Concluding remarks and future perspectives

Our understanding of cotranscriptional splicing has advanced significantly in recent years through the development of sophisticated methods to capture and sequence nascent pre-mRNAs as they are being synthesized and processed by the spliceosome. These studies suggest some new rules that apply specifically to cotranscriptional splicing of nascent RNAs which were not revealed by work on splicing uncoupled from transcription, or by analysis of mature mRNAs. The relative importance of exon definition for splicing long introns, and of intron definition for splicing short introns, has been questioned in light of nascent RNA studies that uncovered ultrafast splicing [33,34]. Hard-wired or predetermined pairing of the 5' and 3' SSs has also been questioned in light of evidence that it can be stochastic in the cotranscriptional context and that the timing of cotranscriptional splicing is very heterogeneous [37,38]. We are also beginning to learn about splicing regulatory mechanisms that work differently in cotranscriptional versus post-transcriptional contexts [70]. Structural probing of nascent transcripts is starting to yield information about the relationship between RNA folding and cotranscriptional splicing [67,87]. Nascent RNA studies are currently limited by sequencing read length and depth and by our ability to relate different datasets with one another. When these limitations are overcome many new insights will be revealed by relating the splicing of different introns to RNA structures, to the binding of RBPs, and to transcriptional pauses. The structures of complexes between components of the splicing machinery and the transcription machinery promise to revolutionize our understanding of cotranscriptional splicing. Indeed, this revolution has already been sparked by the suggestion that U1 snRNP binds to Pol II in a way that is compatible with early 5' SS recognition [53] (Figure 2).

### Acknowledgments

We thank R. Zhao, R. Luhrmann, E. Makeyev, and T. Blumenthal for helpful discussions. Figures were created with [BioRender.com](https://BioRender.com). Our work is supported by the University of Colorado Denver RNA Bioscience Initiative and National Institutes of Health (NIH) grants R35GM118051 and R01HD100935 to D.L.B.

### Declaration of Interests

The authors declare no conflicts of interest.

### References

- Pan, Q. *et al.* (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415
- Ule, J. and Blencowe, B.J. (2019) Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol. Cell* 76, 329–345
- Wilkinson, M.E. *et al.* (2020) RNA splicing by the spliceosome. *Annu. Rev. Biochem.* 89, 359–388
- Fica, S.M. (2020) Cryo-EM snapshots of the human spliceosome reveal structural adaptations for splicing regulation. *Curr. Opin. Struct. Biol.* 65, 139–148
- Wahl, M.C. *et al.* (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718
- Beyer, A.L. and Osheim, Y.N. (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2, 754–765
- Bauren, G. and Wieslander, L. (1994) Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* 76, 183–192
- Pandya-Jones, A. and Black, D.L. (2009) Co-transcriptional splicing of constitutive and alternative exons. *RNA* 15, 1896–1908
- 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
- Khodor, Y.L. *et al.* (2011) Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev.* 25, 2502–2512
- Churchman, L.S. and Weissman, J.S. (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368–373
- Muniz, L. *et al.* (2021) RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *EMBO J.* 40, e105740
- Tardiff, D.F. *et al.* (2006) A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol. Cell* 24, 917–929
- 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
- Bird, G. *et al.* (2005) Ribozyme cleavage reveals connections between mRNA release from the site of transcription and pre-mRNA processing. *Mol. Cell* 20, 747–758
- 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
- Hicks, M.J. *et al.* (2006) Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. *PLoS Biol.* 4, e147
- Bentley, D.L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* 15, 163–175
- Pineda, J.M.B. and Bradley, R.K. (2018) Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes Dev.* 32, 577–591
- Roca, X. *et al.* (2012) Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev.* 26, 1098–1109

### Outstanding questions

Are there functional differences between mRNA RNPs produced by cotranscriptional versus post-transcriptional splicing?

What is the general significance of stochastic recursive splicing in the removal of long introns, and more generally how is pairing of 5' and 3' SSs determined in a cotranscriptional setting?

What is the significance of ultrafast splicing with intron definition versus delayed splicing with exon definition for regulated gene expression?

Does a Pol II-U1 snRNP-5'SS complex facilitate scanning of a looped intron for candidate 3' SSs?

How does the regulation of Pol II transcription elongation and pausing affect cotranscriptional splicing?

How does coordinated splicing of neighboring introns work?

How do the splicing and cleavage polyadenylation machineries communicate to achieve positive and negative regulation of 3'-end processing?

What are the mechanisms that make transcription initiation and elongation sensitive to splicing?

21. Corley, M. *et al.* (2020) How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell* 78, 9–29
22. Shenasa, H. and Hertel, K.J. (2019) Combinatorial regulation of alternative splicing. *Biochim. Biophys. Acta Gene Regul. Mech.* 1862, 194392
23. Lim, S.R. and Hertel, K.J. (2004) Commitment to splice site pairing coincides with A complex formation. *Mol. Cell* 15, 477–483
24. Wu, J.Y. and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75, 1061–1070
25. Howard, J.M. and Sanford, J.R. (2015) The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip. Rev. RNA* 6, 93–110
26. Shama, 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
27. Schneider, M. *et al.* (2010) Exon definition complexes contain the tri-snRNP and can be directly converted into B-like pre-catalytic splicing complexes. *Mol. Cell* 38, 223–235
28. Bergset, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270, 2411–2414
29. Braun, J.E. *et al.* (2018) Synergistic assembly of human pre-spliceosomes across introns and exons. *eLife* 7, e37751
30. Li, X. *et al.* (2019) A unified mechanism for intron and exon definition and back-splicing. *Nature* 573, 375–380
31. 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
32. Pai, A.A. *et al.* (2017) The kinetics of pre-mRNA splicing in the *Drosophila* genome and the influence of gene architecture. *eLife* 6, e32537
33. Oesterreich, F.C. *et al.* (2016) Splicing of nascent RNA coincides with intron exit from rna polymerase II. *Cell* 165, 372–381
34. Zeng, Y. *et al.* (2022) Profiling lariat intermediates reveals genetic determinants of early and late co-transcriptional splicing. *Mol. Cell* 82, 4681–4699
35. Sousa-Luis, R. *et al.* (2021) POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. *Mol. Cell* 81, 1935–1950
36. Drexler, H.L. *et al.* (2020) Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* 77, 985–998
37. Gildea, M.A. *et al.* (2022) Transcript-specific determinants of pre-mRNA splicing revealed through in vivo kinetic analyses of the 1st and 2nd chemical steps. *Mol. Cell* 82, 2967–2981
38. Wan, Y. *et al.* (2021) Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell* 184, 2878–2895
39. Sibley, C.R. *et al.* (2015) Recursive splicing in long vertebrate genes. *Nature* 521, 371–375
40. Blazquez, L. *et al.* (2018) Exon junction complex shapes the transcriptome by repressing recursive splicing. *Mol. Cell* 72, 496–509
41. Boehm, V. *et al.* (2018) Exon junction complexes suppress spurious splice sites to safeguard transcriptome integrity. *Mol. Cell* 72, 482–495
42. 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
43. 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
44. Yuryev, A. *et al.* (1996) The CTD of RNA polymerase II interacts with a novel set of SR-like proteins. *Proc. Natl. Acad. Sci.* 93, 6975–6980
45. Eick, D. and Geyer, M. (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* 113, 8456–8490
46. McCracken, S. *et al.* (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–361
47. Hirose, Y. *et al.* (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.* 13, 1234–1239
48. Rimmel, J.K. *et al.* (2020) Selective inhibition of CDK7 reveals high-confidence targets and new models for TFIIH function in transcription. *Genes Dev.* 34, 1452–1473
49. Nojima, T. *et al.* (2018) RNA polymerase II phosphorylated on CTD serine 5 interacts with the spliceosome during co-transcriptional splicing. *Mol. Cell* 72, 369–379
50. 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
51. Guo, Y.E. *et al.* (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572, 543–548
52. Das, R. *et al.* (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* 26, 867–881
53. Zhang, S. *et al.* (2021) Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science* 371, 305–309
54. Mimoso, C.A. and Adelman, K. (2023) U1 snRNP increases RNA Pol II elongation rate to enable synthesis of long genes. *Mol. Cell* 83, 1264–1279
55. Spiluttini, B. *et al.* (2010) Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *J. Cell Sci.* 123, 2085–2093
56. Brody, Y. *et al.* (2011) The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol.* 9, e1000573
57. Lacadie, S.A. and Rosbash, M. (2005) Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5' ss base pairing in yeast. *Mol. Cell* 19, 65–75
58. Leader, Y. *et al.* (2021) The upstream 5' splice site remains associated to the transcription machinery during intron synthesis. *Nat. Commun.* 12, 4545
59. Listerman, I. *et al.* (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.* 13, 815–822
60. Kotovic, K.M. *et al.* (2003) Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol. Cell Biol.* 23, 5768–5779
61. 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
62. 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
63. Roca, X. *et al.* (2013) Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev.* 27, 129–144
64. Hansen, S.R. *et al.* (2022) Multi-step recognition of potential 5' splice sites by the *Saccharomyces cerevisiae* U1 snRNP. *eLife* 11, e70534
65. Shenasa, H. *et al.* (2020) Allosteric regulation of U1 snRNP by splicing regulatory proteins controls spliceosomal assembly. *RNA* 26, 1389–1399
66. Aslanzadeh, V. *et al.* (2018) Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. *Genome Res.* 28, 203–213
67. Saldi, T. *et al.* (2021) Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. *Mol. Cell* 81, 1789–1801
68. Llorian, M. *et al.* (2010) Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat. Struct. Mol. Biol.* 17, 1114–1123
69. Sharma, S. *et al.* (2005) Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol. Cell* 19, 485–496
70. Iannone, C. *et al.* (2023) Ptbp1-activated co-transcriptional splicing controls epigenetic status of pluripotent stem cells. *Mol. Cell* 83, 203–218
71. Aebi, M. and Weissman, C. (1987) Precision and orderliness in splicing. *Trends Genet.* 3, 102–107
72. Kim, S.W. *et al.* (2017) Widespread intra-dependencies in the removal of introns from human transcripts. *Nucleic Acids Res.* 45, 9503–9513
73. Herzel, L. *et al.* (2018) Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res.* 28, 1008–1019
74. Reimer, K.A. *et al.* (2021) Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell* 81, 998–1012

75. Spitale, R.C. and Incarnato, D. (2022) Probing the dynamic RNA structurome and its functions. *Nat. Rev. Genet.* 24, 178–196
76. Pan, T. and Sosnick, T. (2006) RNA folding during transcription. *Annu. Rev. Biophys. Biomol. Struct.* 35, 161–175
77. Ganser, L.R. *et al.* (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* 20, 474–489
78. Bushhouse, D.Z. *et al.* (2022) How does RNA fold dynamically? *J. Mol. Biol.* 434, 167665
79. Pan, T. *et al.* (1999) Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9545–9550
80. Wong, T.N. *et al.* (2007) Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc. Natl. Acad. Sci. U. S. A.* 104, 17995–18000
81. Yu, A.M. *et al.* (2021) Computationally reconstructing cotranscriptional RNA folding from experimental data reveals rearrangement of non-native folding intermediates. *Mol. Cell* 81, 870–883
82. Goguel, V. and Rosbash, M. (1993) Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* 72, 893–901
83. Graveley, B.R. (2005) Mutually exclusive splicing of the insect *Dscam* pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123, 65–73
84. Warf, M.B. and Berglund, J.A. (2010) Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* 35, 169–178
85. Kainina, M. *et al.* (2021) Multiple competing RNA structures dynamically control alternative splicing in the human *ATE1* gene. *Nucleic Acids Res.* 49, 479–490
86. Taliaferro, J.M. *et al.* (2016) RNA sequence context effects measured in vitro predict in vivo protein binding and regulation. *Mol. Cell* 64, 294–306
87. Kumar, J. *et al.* (2022) Quantitative prediction of variant effects on alternative splicing in MAPT using endogenous pre-messenger RNA structure probing. *eLife* 11, e73888
88. Eperon, L.P. *et al.* (1988) Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell* 54, 393–401
89. Goguel, V. *et al.* (1993) Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. *Mol. Cell. Biol.* 13, 6841–6848
90. Buratti, E. and Baralle, F.E. (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.* 24, 10505–10514
91. Martinez-Rucobo, F.W. *et al.* (2015) Molecular basis of transcription-coupled pre-mRNA capping. *Mol. Cell* 58, 1079–1089
92. Cho, E.J. *et al.* (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* 11, 3319–3326
93. McCracken, S. *et al.* (1997) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* 11, 3306–3318
94. Lewis, J.D. *et al.* (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice-site. *Genes Dev.* 10, 1683–1698
95. O'Mullane, L. and Eperon, I.C. (1998) The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. *Mol. Cell. Biol.* 18, 7510–7520
96. Rambout, X. and Maquat, L.E. (2020) The nuclear cap-binding complex as choreographer of gene transcription and pre-mRNA processing. *Genes Dev.* 34, 1113–1127
97. Gornemann, J. *et al.* (2005) Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol. Cell* 19, 53–63
98. Martinez, N.M. *et al.* (2022) Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing. *Mol. Cell* 82, 645–659
99. Kierzek, E. *et al.* (2014) The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* 42, 3492–3501
100. Louloup, A. *et al.* (2018) Transient N6-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency. *Cell Rep.* 23, 3429–3437
101. Liu, N. *et al.* (2015) N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* 518, 560–564
102. Zhou, K.I. *et al.* (2019) Regulation of co-transcriptional pre-mRNA splicing by m6A through the low-complexity protein hnRNP G. *Mol. Cell* 76, 70–81
103. Xiao, W. *et al.* (2016) Nuclear m6A reader YTHDC1 regulates mRNA splicing. *Mol. Cell* 61, 507–519
104. Bazak, L. *et al.* (2014) A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* 24, 365–376
105. Kapoor, U. *et al.* (2020) ADAR-deficiency perturbs the global splicing landscape in mouse tissues. *Genome Res.* 30, 1107–1118
106. Licht, K. *et al.* (2019) A high resolution A-to-I editing map in the mouse identifies editing events controlled by pre-mRNA splicing. *Genome Res.* 29, 1453–1463
107. Niwa, M. and Berget, S.M. (1991) Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev.* 5, 2086–2095
108. Rigo, F. and Martinson, H.G. (2008) Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. *Mol. Cell. Biol.* 28, 849–862
109. Niwa, M. *et al.* (1990) In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev.* 4, 1552–1559
110. Dye, M.J. and Proudfoot, N.J. (1999) Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol. Cell* 3, 371–378
111. Antoniou, M. *et al.* (1998) Efficient 3'-end formation of human beta-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res.* 26, 721–729
112. Davidson, L. and West, S. (2013) Splicing-coupled 3' end formation requires a terminal splice acceptor site, but not intron excision. *Nucleic Acids Res.* 41, 7101–7114
113. Millevoi, S. *et al.* (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *EMBO J.* 25, 4854–4864
114. Kyburz, A. *et al.* (2006) Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' End processing and splicing. *Mol. Cell* 23, 195–205
115. Kaida, D. *et al.* (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668
116. Berg, M.G. *et al.* (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64
117. Peterson, M.L. and Perry, R.P. (1986) Regulated production of mu m and mu s mRNA requires linkage of the poly(A) addition sites and is dependent on the length of the mu s-mu m intron. *Proc. Natl. Acad. Sci. U. S. A.* 83, 8883–8887
118. Peterson, M.L. and Perry, R.P. (1989) The regulated production of mu m and mu s mRNA is dependent on the relative efficiencies of mu s poly(A) site usage and the c mu 4-to-M1 splice. *Mol. Cell. Biol.* 9, 726–738
119. Vagner, S. *et al.* (2000) Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP. *RNA* 6, 178–188
120. So, B.R. *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
121. Geisberg, J.V. *et al.* (2022) Nucleotide-level linkage of transcriptional elongation and polyadenylation. *eLife* 11, e83153
122. Carminati, M. *et al.* (2022) A direct interaction between CPF and Pol II links RNA 3'-end processing to transcription. *BioRxiv* Published online July 28, 2022. <https://doi.org/10.1101/2022.07.28.501803>
123. Tikhonov, M.V. *et al.* (2013) Competition within introns: splicing wins over polyadenylation via a general mechanism. *Acta Nat.* 5, 52–61
124. 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

125. Fong, N. *et al.* (2022) The pausing zone and control of RNA polymerase II elongation by Spt5: Implications for the pause-release model. *Mol. Cell* 82, 3632–3645
126. Alexander, R.D. *et al.* (2010) Splicing-dependent RNA polymerase pausing in yeast. *Mol. Cell* 40, 582–593
127. Sheridan, R.M. *et al.* (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
128. Fong, N. *et al.* (2014) Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* 28, 2663–2676
129. de la Mata, M. *et al.* (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* 12, 525–532
130. Braberg, H. *et al.* (2013) From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* 154, 775–788
131. Dujardin, G. *et al.* (2014) How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell* 54, 683–690
132. Maslon, M.M. *et al.* (2019) A slow transcription rate causes embryonic lethality and perturbs kinetic coupling of neuronal genes. *EMBO J.* 38, e101244
133. Slobodin, B. *et al.* (2017) Transcription impacts the efficiency of mRNA translation via co-transcriptional N6-adenosine methylation. *Cell* 169, 326–337
134. Vos, S.M. *et al.* (2020) Structure of complete Pol II–DSIF–PAF–SPT6 transcription complex reveals RTF1 allosteric activation. *Nat. Struct. Mol. Biol.* 27, 668–677
135. Mohamed, A.A. *et al.* (2022) Structural advances in transcription elongation. *Curr. Opin. Struct. Biol.* 75, 102422
136. Caizzi, L. *et al.* (2021) Efficient RNA polymerase II pause release requires U2 snRNP function. *Mol. Cell* 81, 1920–1934
137. Castillo-Guzman, D. *et al.* (2020) SF3B1-targeted splicing inhibition triggers global alterations in transcriptional dynamics and R-Loop metabolism. *BioRxiv* Published online June 8, 2020. <https://doi.org/10.1101/2020.06.08.130583>
138. Chen, Y. *et al.* (2009) DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev.* 23, 2765–2777
139. Chathoth, K.T. *et al.* (2014) A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. *Mol. Cell* 53, 779–790
140. Zhang, Z. *et al.* (2020) Molecular architecture of the human 17S U2 snRNP. *Nature* 583, 310–313
141. Fiszbain, A. *et al.* (2019) Exon-mediated activation of transcription starts. *Cell* 179, 1551–1565
142. Reyes, A. and Huber, W. (2018) Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* 46, 582–592
143. Damgaard, C.K. *et al.* (2008) A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol. Cell* 29, 271–278
144. Uriostegui-Arcos, M. *et al.* (2022) Splicing-dependent transcriptional activation. *BioRxiv* Published online September 17, 2022. <https://doi.org/10.1101/2022.09.16.508316>
145. Neves, L.T. *et al.* (2017) The histone variant H2A.Z promotes efficient cotranscriptional splicing in *S. cerevisiae*. *Genes Dev.* 31, 702–717
146. Kolasinska-Zwiercz, P. *et al.* (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41, 376–381
147. Gomez Acuna, L.I. *et al.* (2013) Connections between chromatin signatures and splicing. *Wiley Interdiscip. Rev. RNA* 4, 77–91
148. de Almeida, S.F. *et al.* (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.* 18, 977–983
149. Marasco, L.E. *et al.* (2022) Counteracting chromatin effects of a splicing-correcting antisense oligonucleotide improves its therapeutic efficacy in spinal muscular atrophy. *Cell* 185, 2057–2070
150. Semlow, D.R. and Staley, J.P. (2012) Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem. Sci.* 37, 263–273