MRNA Splicing Machinery

Subjects: Oncology
Contributor: Jing Li

Dysregulation of messenger RNA (mRNA) processing—in particular mRNA splicing—is a hallmark of cancer. Compared to normal cells, cancer cells frequently present aberrant mRNA splicing, which promotes cancer progression and treatment resistance. This hallmark provides opportunities for developing new targeted cancer treatments. Splicing of precursor mRNA into mature mRNA is executed by a dynamic complex of proteins and small RNAs called the spliceosome. Spliceosomes are part of the supraspliceosome, a macromolecular structure where all co-transcriptional mRNA processing activities in the cell nucleus are coordinated. Here we review the biology of the mRNA splicing machinery in the context of other mRNA processing activities in the supraspliceosome.

Keywords: alternative splicing ; splicing dysregulation ; splicing factors

1. Introduction

A central dogma of eukaryotic cell biology describes the transcription of genes from DNA in the cell nucleus into messenger RNA (mRNA), which is transported to the cytosol where it is translated into proteins. Hence, mRNA represents the crucial intermediate between “the script” of the cell or its genetic code and “the workers” of the cell, the proteins. In 1977, Nobel laureates Phillip Sharp and Richard Roberts changed our view on how genetic codes are used to produce proteins. They discovered that when hybridized, the major late mRNA transcript from an adenovirus and its DNA template did not form a continuous hybrid DNA-RNA double strand. Instead, electron microscope analysis visualized single strand loops of DNA protruding from different sections of the hybrid double strand [1][2]. Their important finding showed that the genetic message on the DNA is discontinuous and suggested that in eukaryotes the transfer of genetic information from DNA to mRNA is not just by transcription, but also includes removal of at that time considered “invalid” genetic information from a pre-mRNA intermediate transcript. The genetic information without translation function was called introns, while the genetic information that is retained in the mRNA and is subsequently translated by ribosomes into protein was called exons. We now know that the mRNA splicing process, in which introns are removed from the primary precursor transcript to yield mature mRNA, takes place in a large complex of proteins and small RNA molecules called the spliceosome. In recent years, spliceosomes in the process of catalyzing mRNA splicing have been captured and analyzed successfully [3][4][5][6], providing the possibility for in-depth study of the splicing process and splicing disorders. The mRNA splicing process includes highly dynamic spliceosome formation, rearrangement and catalytic steps, which are discussed below.

The mRNA splicing machinery starts to assemble onto pre-mRNA while it is being transcribed in the nucleus [7]. Regardless of the number of introns in a pre-mRNA, four spliceosomes assemble with the pre-mRNA in a complex coined the supraspliceosome, allowing coordinated simultaneous splicing of four introns [8][9]. Emerging evidence has shown that the supraspliceosome, in addition to the mRNA splicing machinery, also harbors other mRNA processing activities [10][11]. Many of the proteins that carry out or initiate these reactions are recruited to the pre-mRNA through interaction with the carboxy-terminal domain of RNA polymerase II [7], allowing immediate processing of the nascent pre-mRNA. Since RNA processing reactions are all enzymatic activities, with kinetics and efficacies dependent on the conditions, they influence each other. The vast majority of mRNA splicing occurs co-transcriptionally, but the process can continue after release of the transcript from its DNA template to complete splicing of—in particular—the last transcribed 3′ introns [12]. The supraspliceosome is considered the master coordinator of quality-controlled mRNA processing in the cell nucleus, producing correctly spliced mature mRNA for subsequent export to the cytosol [13].

The boundaries between exons and introns on the pre-mRNA are delineated by splice (donor and acceptor) sites. When these sites are effectively recognized by the spliceosome, it completely removes the introns and anneals the adjacent exons to yield the mature mRNA. The efficacy at which splice sites are recognized by the spliceosome is, however, not absolute. In contrast, the utilization of splice donor and acceptor sites by the mRNA splicing machinery is context dependent [14][15]. As will be discussed in detail later, splice site utilization is regulated by cis-acting RNA elements and trans-acting proteins. Consequently, many single pre-mRNAs can be processed into multiple mature mRNAs carrying

joined exon combinations that encode different protein variants, often with distinct functions. The vast majority of transcribed human genes are subject to this so-called alternative splicing (AS) [16]. On average, human gene transcripts are processed into three or more alternatively spliced mature mRNA transcripts [17]. This contributes to protein diversity and facilitates cell differentiation and tissue development [18][19][20].

The functional impact of AS requires for it to be tightly regulated and controlled. Hence, not surprisingly, dysregulated mRNA processing, such as mRNA splicing, is an important cause of disease, including cancer [21]. A pan-cancer analysis [22] showed that tumors harbor up to 30% more AS events than normal tissues, with many tumors harboring thousands of AS events that are not detectable in normal samples. As will be discussed herein, abnormal splicing of tumor-related genes is associated with cancer development and progression, as well as with resistance to therapy.

2. Structure and Function of the (Supra) Spliceosome

The supraspliceosome comprises a single pre-mRNA in association with four spliceosomes [9][23]. The RNA splicing machinery, splicing reaction and alternative splicing are described in more detail in Section 2.1. The presence of intact pre-mRNA is essential for the integrity of the supraspliceosome [8]. A “rolling model” was proposed in which the pre-mRNA “rolls” through the supraspliceosome, thereby making new introns available to the splicing machinery. The supraspliceosome is co-transcriptionally assembled and does not disassemble during the splicing reaction. Its composition remains intact as snRNA:pre-mRNA basepairing and protein:RNA interactions are reformed on each subsequent intron [24]. Electron microscopic studies of the structure of the supraspliceosome revealed that the four native spliceosomes consist of a large and a small subunit, of which the first faces towards the outside and the latter faces the inside of the supraspliceosome, forming contacts with the neighboring spliceosomes. The small subunit contains a hole that leads into the cavity between its neighboring small and large subunits, presumably to facilitate the pre-mRNA to connect all four spliceosomes [25]. An in-silico study proposed that the snRNPs reside within the large subunit, leaving the non-snRNPs to be attributed to the small subunit [26]. A schematic depiction of the supraspliceosome is given in Figure 1a.

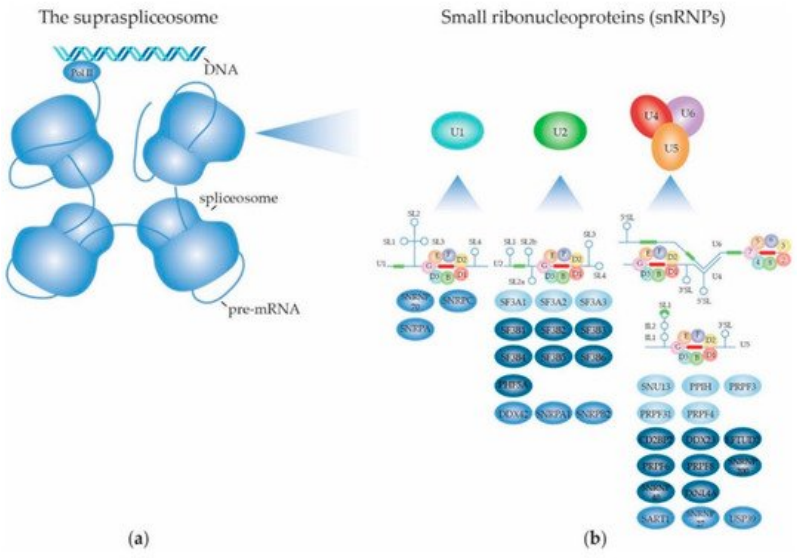

Figure 1. Structure and composition of the supraspliceosome. (**a**) Schematic impression of the macromolecular structure of the supraspliceosome, consisting of four native spliceosomes assembled onto a pre-mRNA being transcribed by RNA Pol II. In addition to pre-mRNA splicing carried out by the spliceosomes, the supraspliceosome provides a platform for all other co-transcriptional mRNA processing activities (by proteins that are not shown in the illustration). (**b**) Major components of the spliceosome U1, U2, and U4/U6.U5 snRNP subunits, with their Sm-snRNA core complexes and subunit-specific splicing factor proteins (U2: SF3a complex in light blue, SF3b complex in dark blue; U4/U6.U5: U4/U6-specific proteins in light blue, U5-specific proteins in dark blue). SL, snRNA stem-loop; red box, snRNA Sm site; green box, other RNA-binding domains (U1: 5′ss recognition sequence, U2: BPS recognition sequence, U6: U2/U6 duplex-forming sequence).

2.1. The RNA Splicing Machinery and the Splicing Reaction

A core task of the supraspliceosome is catalyzing the mRNA splicing reaction in its four spliceosomes. The splicing reaction itself consists of two subsequent transesterification reactions that together yield two ligated exons and excise the intronic sequence. The splicing process is initiated by recognition of splice sites (ss) in intronic sequences. The vast majority of introns in eukaryotes are U2-type introns that contain highly conserved GT and AG dinucleotides at their 5′ss

and 3′ss, respectively. The remaining part of the introns are U12-type introns, accounting for ~0.01-0.02% of eukaryotic introns, that have AT-AC boundary sequences [27]. Although there are some exceptions, U2-type introns undergo so-called canonical splicing by the major spliceosome, whereas U12-type introns are spliced by the minor spliceosome through non-canonical splicing. As the majority of human introns are U2-type introns, this review will only focus on the major spliceosome. The major spliceosome consists of the five snRNPs U1, U2, U4, U5 and U6; DExD/H-type RNA-dependent ATPases/helicases that facilitate structural remodeling of the snRNPs at different steps in the splicing process, as well as many other splicing factors that regulate splice site usage and the mRNA splicing reaction. The snRNPs are the core units of the spliceosome. In this section, each snRNP's structure and function will be described in more detail. An overview of all the snRNP-specific and -associated RNA splicing factors according to the KEGG, amiGO and Reactome databases is given in Supplementary Table S1 and the most important ones are shown in Figure 1b.

2.1.1. Biogenesis of the Spliceosome: Assembly and Transport of Sm-snRNA Complexes

Biogenesis of snRNPs for assembly of spliceosomes has been reviewed more extensively elsewhere [28][29][30][31] and is illustrated in Figure 2. The U1, U2, U4 and U5 snRNAs are transcribed in the nucleus and exported to the cytoplasm, where they associate with the seven Sm proteins that form a doughnut-shaped circular structure around the snRNA. In the case of the U6 snRNA, the LSm proteins form the typical heptameric ring structure around its Sm site and its biogenesis presumably occurs within the nucleus. The important (L)Sm-snRNA structure in the heart of the spliceosome is discussed in detail below. Complete Sm rings incorporating their snRNA (termed Sm-snRNA complexes here) need to translocate back to the nucleus where they reside in Cajal bodies. There, further snRNP biogenesis takes place, by binding of U snRNP particle-specific proteins to the Sm-snRNA complexes. Mature snRNPs are primarily located in nuclear membrane-less organelles termed nuclear speckles that serve as a reservoir for spliceosome components. Nuclear speckles are found in interchromatin regions close to actively transcribed genes (reviewed by [32][33]). Many proteins involved in transcription, epigenetic regulation, and RNA processing, modification and packaging are present in these speckles, making them nuclear gene expression hubs. The localization of proteins in nuclear speckles is regulated by post-translational modifications such as phosphorylation, addition of phosphoinositol derivates, ubiquitination and SUMOylation [32]. The localization in speckles is important, as increasing the levels of U1 snRNPs did not achieve enhancement of mRNA production when no speckles were present [34]. For use in the mRNA splicing reaction, snRNPs are recruited from the speckles to active transcription sites at the interface of speckles and chromatin. The Sm and LSm proteins have been postulated to be the earliest spliceosomal components, with their gene family nearly achieving its current composition already by the time the last eukaryotic common ancestor emerged approximately two and a half billion years ago [35]. The seven Sm proteins that together form the heteroheptameric ring structure are denoted SmB/B', SmE, SmF, SmG, SmD1, SmD2 and SmD3. Additionally, SmN is a tissue-specific substitute for SmB/B' expressed primarily in the brain and heart that affects mRNA splicing through downregulation of mature U2 snRNP when it is incorporated in its Sm ring [36]. The structurally highly similar (L)Sm proteins form a heptameric ring structure around each snRNA, which presumably functions as a platform for other snRNP proteins to assemble onto. Sm proteins are crucial for the assembly, stability and nuclear import of snRNPs and hence for proper functioning of the spliceosome.

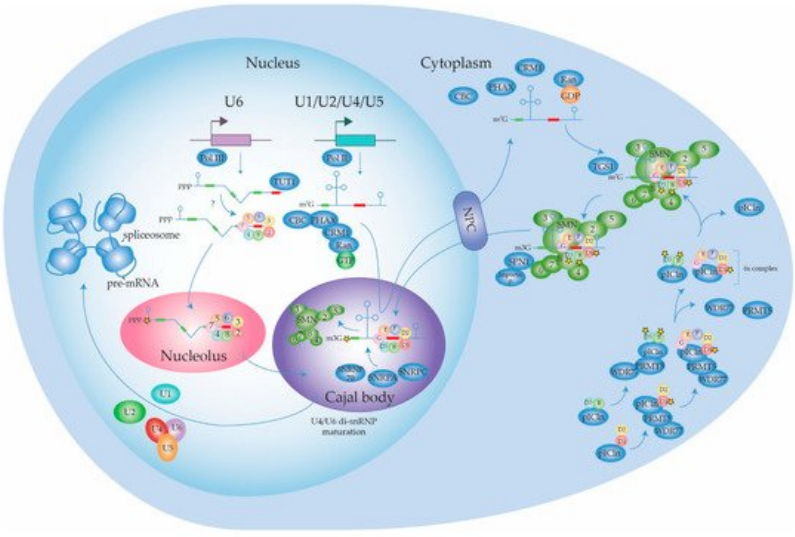

Figure 2. Spliceosome snRNP subunit biogenesis. Details are given in the main text. Since U1, U2, U4 and U5 snRNPs are compiled in the same manner, the biogenesis of U1 snRNP is shown as example. The snRNAs are transcribed by either RNA polymerase II (U1, U2, U4, U5) or RNA polymerase III (U6). RNA polymerase II-transcribed snRNA is exported to the cytoplasm after quality control in Cajal bodies (CBs). SmD1, SmD3 and SmB/B' are dimethylated (displayed by stars) in the cytoplasm by the methylosome complex. The Sm ring is assembled around the snRNA by the SMN complex

and the Sm-snRNA core complex is imported back into the nucleus where snRNP maturation takes place in the CBs. The exact LSm assembly route onto the U6 snRNA is not known but presumably takes place in the nucleus. The LSm-U6 snRNA core complex probably moves to the nucleolus where it is 2-O'-methylated (displayed as a star) and subsequently transported to the CBs where U6-specific proteins are assembled and U4/U6.U5 tri-snRNP maturation can take place. Mature snRNPs are recruited into the supraspliceosome to take part in the RNA splicing reaction.

Each Sm protein consists of a short N-terminal α helix and five anti-parallel β strands, representing the highly conserved Sm fold. β strands 1–3 represent Sm motif 1 which is involved in the protein-snRNA interaction. β4 and β5 represent Sm motif 2 which is involved in the protein–protein interactions within the Sm heptameric ring structure. β4 interacts with the β5 of its neighbor through the formation of hydrogen bonds. The Sm motifs are highly conserved [37]. The Sm ring structure is further stabilized by hydrophobic residues that point towards its center and make contacts with other Sm proteins [38].

The association of the Sm ring with the U snRNA represents the first phase of the snRNP assembly [39]. The stepwise assembly of the Sm-snRNA complex takes place in the cytoplasm (Figure 2) and is mediated by the methylsome complex (consisting of protein arginine N-methyltransferase 5 (PRMT5), methylosome protein 50 (WDR77) and methylosome subunit pICln) and the SMN complex (consisting of Gem-associated proteins Gemin 2–8 and SMN1 and 2) [40][41]. SmD1-SmD2 initially forms a dimer bound by pICln that is assembled onto PRMT5/WDR77 where SmD1 is symmetrically dimethylated. SmD3/SmB/B', in parallel, is bound by another pICln molecule and recruited to a second PRMT5/WDR77 complex where SmD3 and SmB are both symmetrically dimethylated. In plants, Sm protein dimethylation by PRMT5 was shown to be required for recruitment of the NineTeen/Prp19 complex (NTC) and therefore proper functioning of the spliceosome [42]. In addition to PRMT5, PRMT7 was also found to be required for symmetrically dimethylating Sm proteins [43]. Subsequently, the SmF-SmE-SmG trimer binds to SmD1-SmD2-pICln on the methylosome complex and forms the 6S complex [40]. The 6S complex is an Sm ring intermediate in which pICln functions as an Sm protein mimic reserving space for SmD3-SmB/B' [40][44]. Of note, pICln was postulated to not only function as structural chaperone in the formation of the Sm protein ring, but also to prevent the formation of aggregates by unassembled Sm proteins [41]. In the final steps of the Sm ring assembly, the 6S complex and SmD3-SmB/B' complex release their pICln subunits and are loaded onto the SMN complex through interactions with Gemin2, where the final heptameric Sm ring is formed [44][45]. Interestingly, Sm proteins are not the only proteins that adopt the Sm fold, as this structural arrangement was also observed for Gemin6 and Gemin7 in the SMN complex. Gemin6 and Gemin7 form a dimer through their β4 and β5 strands, respectively. The β5 of Gemin6 and β4 of Gemin7 are involved in binding Sm proteins, thereby facilitating the interaction between Sm proteins and the SMN complex [46]. Defects in the association of Sm proteins with the SMN complex can have drastic phenotypic consequences, as malfunction of the SMN complex due to loss of SMN1 leads to spinal muscular atrophy (reviewed by [47]). Additionally, the F22S mutation in SmE leads to the disruption of its interaction with the SMN complex and is associated with microcephaly [48]. Above-mentioned SMN defects give rise to aberrant splicing patterns.

Gemin5 is an RNA-binding protein (RBP) that is part of the SMN complex and is involved in the recruitment of the snRNA towards the Sm proteins through recognition of the Sm site and the 7-methylguanylate (m^7G) cap of the snRNA [49][50][51][52]. The Sm site of U1 has a weaker binding affinity than the Sm sites of other U snRNAs for Gemin5. Here, U1 snRNP 70 kDa (SNRNP70) in the cytoplasm recruits Gemin2-Sm complexes directly to U1 snRNA, independent of Gemin5 [53]. The Sm proteins are organized around the snRNA in the following consecutive order: SmE, SmG, SmD3, SmB, SmD1, SmD2 and SmF, with each Sm protein directly interacting with a single nucleotide of the Sm site [54]. The Sm site consists of a 5′ adenosine preceding five consecutive uridines (with U1 being the exception that contains a guanine instead of the fourth uridine), followed by a guanine (with U5 and U6 snRNAs being the exceptions, which contain a uridine or a 2′,3′-cyclicphosphate group, respectively) [54][55][56][57]. The 5′ adenosine and the 2′-OH groups of the sugar backbone of the Sm site are important for stable Sm-snRNA complex formation. Additionally, flanking nucleotides of the Sm site determine the rate of Sm protein assembly onto the snRNA [39]. SmE, SmF and SmG make the initial contact with the Sm site, which is stabilized by SmD1 and SmD2, of which the contact between SmG and the first uridine is highly conserved [57]. Every Sm protein clamps its corresponding Sm site nucleotide in between their loops L3 and L5 [38][54]. The positioning of the Sm ring relative to the Sm site is not the same for every U snRNA. In the human U1 snRNA, the last two nucleotides of the Sm site interact with SmD2 and SmF [54], whereas SmD1 and SmD2 interact with the last two nucleotides of the U4 snRNA Sm site [38]. A study in yeast demonstrated that mutagenesis of conserved Sm motif 1 amino acids in individual Sm proteins did not compromise cell viability, but simultaneous mutation in two Sm proteins was lethal [58]. Thus, at least 6 intact RNA binding sites in the heptameric Sm ring seem sufficient for Sm-snRNA complex formation. During the assembly of the Sm ring around the snRNA by the SMN complex, the snRNA cap is hypermethylated by trimethylguanosine synthase 1 (TGS1) into 2,2,7-trimethylguanosine (m3G) [59].

The m3G-cap is recognized by snurportin-1, which together with importin-β [60] transports the fully assembled Sm-snRNA core complexes back into the nucleus where they are loaded into membrane-less nuclear Cajal bodies. Here, the additional snRNP-specific proteins are loaded onto Sm-snRNA core complexes and snRNP maturation takes place [61]. snRNAs are 2′-O-methylated by small Cajal body-specific RNAs (scaRNAs) and pseudouridylated [28][29][31]. Additionally, snRNAs can be methylated at the N6-position of 2′-O-methylated adenosine residues, which affect spliceosomal function. For instance, the U6 snRNA is methylated at its 2′-O-methyl adenosine at position 43 by RNA N6-adenosine-methyltransferase METTL16 and this affects 5′ss recognition [62]. In the U2 snRNA, the adenosine at position 30 is 2′-O-methylated and subsequently N6-methylated by N(6)-adenine-specific methyltransferase METTL4, which affected 3′ss usage [63]. Both N6-methylated and 2′-O-methylated forms can coexist, as the N6-methylation can be removed by the RNA demethylase alpha-ketoglutarate-dependent dioxygenase FTO [64]. As the adenosines are 2-O′-methylated before they are N6-methylated, it is assumed that the latter also occurs in Cajal bodies upon re-uptake of the Sm-snRNA complex into the nucleus.

2.1.2. Biogenesis of the Spliceosome: Structure and Assembly of the U1 snRNP

The different snRNAs incorporated in the Sm-snRNA complexes direct the formation of the different U snRNPs of the spliceosome in the Cajal bodies (Figure 1b and Figure 2). The U1 snRNA forms four stem-loop (SL) structures, which are oriented in a latin cross-like shape, with SL4 representing the stem. The Sm site is located between the stem and the four-helix junction [65]. The U1 snRNA four-helix junction is situated over a flat surface consisting of the N-termini of each Sm protein. The N-terminus of SmD2 is particularly long and extends into the minor groove of the U1 snRNA. SmB interacts with the SL2 backbone [65]. SL1 and SL2 are also bound by the SNRNP70 and U1 snRNP A (SNRPA) proteins, and SNRNP70 helps to guide the snRNA through the cavity of the Sm ring, together with SmD1 and SmD2 [54]. The U1 Sm core has an additional and unique assembly pathway, in which SNRNP70 plays the key role. As mentioned above, it recruits Gemin2-Sm complexes directly to U1 snRNA, independent of Gemin5. Moreover, SNRNP70 inhibits the formation of other snRNP Sm cores, thereby acting as a regulator of the cell's snRNP repository. This extra U1 Sm core assembly pathway could be an explanation as to why the U1 snRNP is the most abundant snRNP in vertebrates [53]. Another protein involved in U1 Sm core assembly is the RBP FUS, which associates with U1-related proteins and SMN complexes. Mutations in FUS that are associated with amyotrophic lateral sclerosis (ALS) were found to dysregulate SMN function, leading to loss of snRNA levels and affected splicing patterns [66].

The U1 snRNP is the first snRNP to be recruited to pre-mRNA to start the splicing reaction. The composition of U1 snRNP is shown in Figure 1b. A cryo-EM study of the yeast U1 snRNP revealed a shape similar to that of a footprint [67]. The U1 snRNP core (or the foot's ball) is composed of the Sm-snRNA complex and the SNRNP70, SNRPA and SNRPC proteins, with SNRNP70 and the stem-loop 1 (SL1) and SL3 of the U1 snRNA sticking out like toes. The auxiliary area (or the foot's heel) consists of Prp42, Luc7/LUC7L2, Snu56, Nam8/TIA-1 and Prp39/PRPF39. There is no human homologue reported of the yeast Prp42. Instead, in humans, PRPF39 forms a homodimer which interacts with SNRPC, connecting the ball with the heel of the foot, and mimicking the Prp39/Prp42 heterodimer observed in yeast. In immunoprecipitation [68] and X-ray crystallography [65][69] studies of the human U1 snRNP, it was revealed that the N-terminal domain of SNRNP70 plays a crucial role in holding the core domain of the U1 snRNP together, specifically through interacting with SmD2 [65][68][69] and SmB/B' [68]. Moreover, A feedback regulatory mechanism has been described between SNRNP70 and SNRPC, effectuating efficient U1 snRNP homeostasis. Specifically, SNRPC promotes alternative splicing of the SNRNP70 transcript through usage of an alternative 3′ss. This introduces a premature termination codon (PTC), resulting in a truncated splice variant of SNRNP70 that is targeted for nonsense-mediated decay (NMD), and therefore in decreased protein expression of SNRNP70. This in turn leads to decreased incorporation of SNRPC in the U1 snRNP, restoring proper splicing of SNRNP70 to produce the functional protein [70].

2.1.3. Biogenesis of the Spliceosome: Structure and Assembly of the Other snRNPs

The U2 snRNP, which is the next spliceosome unit to be recruited, consists of the U2 Sm-snRNA complex, the SF3a and SF3b complexes and additional U2-specific and -related proteins. The U2 snRNA, like U1 snRNA, forms four stem-loops in the Sm-snRNA complex, of which SL3 and SL4 are bound by U2 snRNP A′ (U2-A′) and B″ (U2-B″) proteins. SL2a is contacted by the SF3b complex. The branchpoint-interacting stem-loop (BSL) is located between SL2a and SL1, and is clamped between SF3B1's HEAT domains. SF3A3 contacts the base of the BSL. The Sm site is located between SL2b and SL3 [71]. The SF3a complex consists of splicing factor 3A subunits 1, 2 and 3 (SF3A1, SF3A2 and SF3A3). SF3A1 facilitates the interaction between the U1 and U2 snRNPs by interacting with the U1 snRNA through its ubiquitin-like (UBL) domain [72]. SF3A1, but also U2-related Calcium Homeostasis Endoplasmic Reticulum Protein plays an additional role in the recruitment of U2 snRNP towards the pre-mRNA through their interactions with branchpoint-bridging protein (BBP)/Splicing Factor 1 (SF1) [73]. The SF3a complex bridges the Sm-snRNA complex with the SF3b complex [71]. The SF3b complex consists of RNA splicing factor 3B subunits 1, 2, 3, 4, 5 and 6 (SF3B1-6) and PHD finger-like domain-

containing protein 5A (PHF5A). In this complex, SF3B6 is positioned in such a way that it can bind to the branchpoint sequence (BPS), facilitating BPS recognition by the U2 snRNP. SF3B1 adopts a closed conformation surrounding SF3B6, and serves as a platform for BPS binding together with PHF5A [71][74]. SF3B1 also appears to play a role in guiding the U2 snRNP towards the pre-mRNA, as this mRNA splicing factor was shown to interact with chromatin at nucleosomes located at exons to be spliced [75].

While U1 and U2 snRNPs assemble individually before being recruited in the spliceosome to participate in the RNA splicing reaction, U4, U5 and U6 preform a tri-snRNP complex in two steps, where U4/U6 first assemble as di-snRNP before the U5 snRNP attaches. U5 snRNA contains one large stem-loop (SL1) and a smaller SL2. The Sm site is located between these loops [76] at the 3′ end [4]. In the mRNA splicing process, SL1 is important for basepairing with the 5′ exon in the pre-mRNA [77]. In yeast, the U5 Sm core serves as a protein-binding platform for U5 specific proteins Prp8/PRPF8 and Snu114/EFTUD2 that associate either through direct interaction with U5 snRNA or with the Sm ring, respectively [4]. The U4 and U6 snRNAs are different from U1, U2 and U5, as these are duplexed within the U4/U6 di-snRNP and U4/U6.U5 tri-snRNP. The U4 snRNA comprises three SLs. The Sm site is located at the 3′ end [78] and is flanked by SL2 above the flat face of the Sm ring, and by SL3 below the tapered side of the Sm ring. The α-helix that makes up the long N-terminus of SmD2 interacts with SL2 and its lysine-rich L4 loop between β3 and β4 interacts with the backbone of SL3 of the U4 snRNA. Moreover, SL2 interacts with SmB and SmG, and SL3 interacts with all Sm proteins except SmG and SmD3 [38]. On either side of SL2, stem 1 and 2 are basepaired with the U6 snRNA [4][78]. The U4/U6.U5 tri-snRNP is cone-shaped, with the U5 snRNP core located at the tip and the U4/U6 di-snRNP in the broader top part, with the (L)Sm heptamers located at the outer corners of the cone. The following proteins are involved in the U4/U6.U5 snRNP assembly. Small Nuclear Ribonucleoprotein 13 (SNU13) binds to a stem-loop in the U4 snRNA duplexed with the U6 snRNA. Next, pre-mRNA Processing Factors (PRPF) 31, 3 and 4 are recruited, giving rise to the complete U4/U6 di-snRNP [79]. Prior to U5 snRNP assembly, PRPF8, EFTUD2 and SNRNP200 form an assembly intermediate with protein AAR2 homolog. The actual assembly of the U5 snRNP is supported by heat shock protein 90 and R2TP complex (consisting of RuvB-like 1, RuvB-like 2, PIH1 domain-containing protein 1 and Homeobox-containing protein 1) and zinc finger HIT domain-containing protein 2 [80][81].

The U6 snRNA is the exception from all other snRNAs in the sense that it is not bound by Sm, but by LSm proteins. It does so at its 3′ end, where it is uridylated by Terminal Uridylyl Transferase 1 [82]. The LSm ring can only recognize U6 snRNA (and not the other snRNAs) because this is the only snRNA that contains the 3′-terminal U tract [83][84]. In yeast, U6 snRNA's 3′ end reaches into the ring structure, but does not stick through it as observed for the other snRNAs, thereby only interacting with one side of the ring [85]. The authors speculate that this leaves RNA-binding domains on the other side of the ring accessible to facilitate interactions between the U4 and U6 snRNAs [85]. Indeed, the LSm proteins were shown to facilitate the formation of the U4/U6 duplex [83]. LSm proteins share homology with Sm proteins [83], also form the Sm fold consisting of a short N-terminal α helix and five anti-parallel β strands [86] and similar to the Sm ring also assemble in a stepwise manner. LSm6-LSm5-LSm7 resembles SmF-SmE-SmG but at least in yeast forms a hexameric LSm657-657 intermediate, which subsequently incorporates LSm2-LSm3 (resembling SmD1-SmD2) and finally LSm4-LSm8 (resembling SmD3-SmB/B'), to form the nuclear LSm2–8 complex that is incorporated in the U6 snRNP [86]. Similar to SmD3 in the Sm ring, LSm4 is symmetrically dimethylated which enables interaction with the SMN complex [87]. Comparable to the Sm ring, for each LSm protein, its β4 strand interacts with the β5 strand of the neighboring LSm protein, and the LSm ring is stabilized by hydrophobic interactions through N-terminal α helices [86].

2.1.4. Dynamic Composition of the Spliceosome: Assembly on the pre-mRNA Substrate

Throughout the RNA splicing cycle, different spliceosome intermediates are formed, termed the E (early), A (pre-spliceosome), B (pre-catalytic), B^{act} (activated), B* (catalytically activated; for the first transesterification reaction), C (catalytic), C* (catalytically activated; for the second transesterification reaction) and P (post-splicing) complexes (Figure 3). These intermediates correspond to specific phases of the splicing process, and consist of varying compositions of snRNPs and splicing factors that are described in more detail below. Hence, during the splicing reaction, snRNPs and splicing factors are recruited, rearranged and released in a sequential manner, making the spliceosome a highly dynamic and fluid structure. Many papers have been published over the past decade regarding the yeast and human spliceosome intermediates, describing their structural properties and protein and RNA components (reviewed by [88][89][90]), the major findings of which are summarized here. Notably, the genomic architecture in lower eukaryotes such as yeast is different from that in higher eukaryotes such as mammals. The former usually have relatively long exons and short introns; the latter often short exons and sometimes very long introns. Most fundamental studies into the biology of the spliceosome were done in yeast, or using recombinant transcripts with short introns. Therefore, the general description of spliceosome assembly below primarily applies to pre-mRNAs with short introns, known as the intron definition model. The steps in the

process that are probably different for transcripts with long introns, according to the postulated exon definition model [91], are mentioned separately.

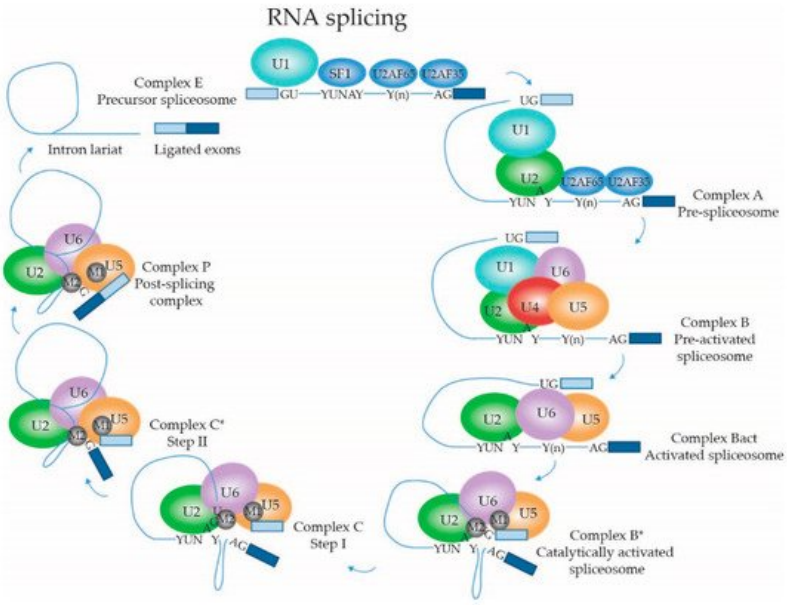

Figure 3. Pre-mRNA splicing reaction performed by the spliceosome. The dynamic composition of the spliceosome, with it different intermediate complexes, is illustrated. Details are given in the main text. Light and dark blue boxes, exons; line, intron; GU, 5′ splice site; AG, 3′ splice site; YUNAY, branchpoint sequence; Y(n), polypyrimidine tract; M1 and M2, Mg^{2+} metal ions at the catalytic site.

The early complex E is the first intermediate that can be discerned in the pre-mRNA splicing process. As mentioned above, the intronic sequence that is to be spliced out contains the highly conserved dinucleotides GT and AG at the 5′ss and 3′ss, respectively, that are recognized by the spliceosome. Moreover, the BPS and polypyrimidine tract (PPT) in the intron play crucial roles in the recruitment of splicing factors. The complex E intermediate is formed when U1 snRNP binds to the 5′ss through basepairing with the 5′ end of its U1 snRNA. SNRNP70, together with SmD3, coordinates SNRPC to support the base-pairing interaction between the 5′-end of the U1 snRNA and 5′ss on the pre-mRNA substrate through its zinc-finger domain [65][69][92]. The recruitment appears to be mediated by RNA polymerase II while it is synthesizing the pre-mRNA; and dependent on the presence of members of the SR family of RNA splicing enhancer proteins [93]. In yeast, recognition of the 5′ss was shown to be supported by U1C/SNRPC, Luc7/LUC7L2, Nam8/TIA-1 [94] and Prp39/PRPF39 [95]. SF1 binds to the BPS [94] and U2 snRNP auxiliary factor 65 kDa subunit (U2AF65) and 35 kDa subunit (U2AF35) are recruited to the PPT and intronic 3′ ss, respectively, of the target pre-mRNA [96][97]. Subsequently, on short introns U2 snRNP is recruited through interacting with U1 snRNP and SF1, replacing SF1 at the BPS. The association of U2 snRNP is further stabilized by U2AF65 [73]. On long introns, U2 snRNP is also recruited to SF1 and U2AF65 near the 3′ss and associates with U1 snRNP, snRNP but positions the U1 snRNP to the downstream 5′ss of the next intron [91]. ATP-dependent RNA helicase DDX46 is required for the transition from complex E to the pre-spliceosome A, and facilitates conformational changes within the U2 and the interaction between the U2 and U1 snRNPs [71]. DDX46 remodels the U2 snRNA, allowing its BSL to bind to the BPS in the intron in an ATP-dependent manner, where the adenosine in the YUNAY consensus sequence is excluded, which is important for later catalysis in the splicing reaction. In yeast, Prp39 anchors U2 snRNP to U1 snRNP by acting as a bridge between the U1C protein and U2 small nuclear ribonucleoprotein A′ (U2A′). In humans, an interaction between PRPF39 and SNRPC is also observed, but is not crucial for complex A formation [98]. This is in line with the exon definition model, where the recruited U1 snRNP and U2 snRNP are to participate in splicing of different introns on either side of the exon. For splicing of transcripts with long introns, neighboring exons must be juxtaposed, existing U1 snRNP-U2 snRNP interactions across exons need to be broken; and new contacts spanning introns need to be established. This transition is still poorly understood, but the process is inhibited by hnRNPI. In the presence of hnRNPI, spliceosome assembly with U1 and U2 snRNPs recruited around exons stalls in an A-like complex [99], showing that the transition occurs prior to U4/U6.U5 tri-snRNP recruitment. Recently, a model for early spliceosome assembly was proposed that unifies the intron definition and exon definition models [94]. Based on cryo-EM analysis of in vitro assembled complexes E and A it was concluded that the same structure can be formed across either an intron or an exon. Structural constraints of complexes formed across short exons make it difficult for the U4/U6.U5 tri-snRNP to subsequently join the spliceosome. This is postulated to be a main trigger for remodeling U1 snRNP-U2 snRNP interactions into an intron-spanning complex, allowing further spliceosome assembly [94].

The pre-activated spliceosome or complex B is formed when the U4/U6.U5 tri-snRNP is recruited. The U5-specific PRPF8 with its N-terminus is able to interact with EFTUD2, DDX23 and the U5 snRNA, which interacts with the pre-mRNA substrate [100]. As was shown in yeast, Prp8/PRPF8′s C-terminus interacts in U5 with the N-terminal helicase domain of SNRNP200. SNRNP200′s C-terminus interacts with EFTUD2 and Ubiquitin Specific Peptidase 39 (USP39). Positioned at the interface of U4/U6 and U5 snRNPs, USP39 is crucial for the stability of the tri-snRNP [101]. Moreover, USP39 is postulated to keep the SNRNP200 RNA helicase positioned away from the U4/U6 duplex, preventing premature unwinding of the U4/U6 snRNAs and thereby of spliceosome catalytic activity in the pre-catalytic stage [78].

During the association of the tri-snRNP with the U2 snRNP, the U1 snRNP places its snRNA between the U4 snRNA and PRPF8, while the U1 SmE and SmG interact with U5-specific ATP-dependent RNA helicase DDX23 [100]. DDX23 unwinds the U1 snRNA:5′ss duplex, and is therefore required for B complex formation, as U6 snRNP replaces U1 snRNP at the 5′ss. Mutations in the DDX23 domain involved in ATP hydrolysis stall the spliceosome before complex B formation, in which U1 snRNP remains associated with the pre-mRNA and the tri-snRNP is not stably integrated yet [102]. The U4 and U6 snRNAs partly form a duplex within the tri-snRNP, rendering U6 snRNA in its inactive configuration. EFTUD2 is also involved in the recruitment of the NTC and NineTeen/Prp19 complex related (NTR) complexes. Recruitment of the NTC and NTR induce conformational changes within the snRNPs necessary for the formation of the active site for the splicing reaction, such as basepairing of the ACAGAGA box of the U6 snRNA with the 5′ss [103][104][105]. As was demonstrated in yeast, pairing of U6 snRNA with the 5′ss occurs prior to U4:U6 duplex unwinding within the embrace of PRPF8 and represents a checkpoint for proper complex B assembly [105].

2.1.5. Dynamic Composition of the Spliceosome: Activation and Catalytic Steps

The unwinding of the U4/U6 snRNAs represents a checkpoint for complex B activation, creating complex Bact. To achieve this, USP39 dissociates, which repositions SNRNP200 and induces conformational changes in SNRNP200 that prompt its helicase activity. During activation of catalytic activity, additional conformational changes occur in the tri-snRNP complex. DDX23 migrates from the outer side of the tri-snRNP towards the RNase H domain of PRPF8 in the center of the complex where the 5′ss basepairing is switched from U1 to U6 snRNA [78]. This results in the release of the U1 snRNP, thereby preventing steric clash of this snRNP with SNRNP200 [98]. This transition from U1 to U6 5′ss basepairing is further supported by the U4/U6.U5 tri-snRNP specific SNRNP27, as was demonstrated in *C. elegans* [106]. Upon the 5′ss transition to the U6 snRNA, SNRNP200 unwinds the U4:U6 snRNA duplex resulting in the dissociation of U4 snRNP from the spliceosome. This allows the 3′ end of the U6 snRNA to basepair with the 5′ end of the U2 snRNA (forming helix I); and also to form a highly conserved internal stem loop (ISL) within the U6 snRNA [107]. Meanwhile, PRPF8 undergoes rearrangements from an open to a closed conformation, as a pocket must be formed to harbor the newly formed U2:U6 duplex and the U5 snRNA SL1, which is necessary to form the active catalytic site [78]. Both the helix I and ISL are involved in the coordination of catalytic metal ions. Overall, SNRNP200, EFTUD2 and PRPF8 are essential for the transition from the pre-catalytic B complex to the activated B^{act} complex. In this intermediate, the active site is cradled by PRPF8, consisting of helix I of the U2:U6 duplex, ISL of U6 snRNA, five Mg^{2+} ions and SL1 of the U5 snRNA [4]. The SL1 of U5 snRNA is basepaired with the 5′ exon [77]. Moreover, within the active site, a triplex structure is formed by several nucleotides of the U6 snRNA [77][108][109].

ATP-dependent RNA helicase-like protein Prp2/DHX16 promotes the transition from the activated B^{act} complex to the catalytically active B* complex [110], through rearrangement of the U2 snRNP around the U2 snRNA:BPS duplex [77][109]. Moreover, several nucleotides of the U6 snRNA are involved in the coordination of Mg^{2+} metal ions via binding to their phosphate groups [77][111]. Of these, two are directly involved in catalysis of the splicing reaction, and the other three fulfill more structural roles [4]. The rearrangements around the U2 snRNA:BPS are supported by step I factors YJU2 [4][109][111][112][113] and CWC25 [4][113] and the presence of one of the two catalytic metal ions (M2) activates the 2′-OH BPS adenosine to perform step I of the splicing reaction; a nucleophilic attack on the phosphorous atom of the 5′ss G nucleotide in which the covalent bond between the 5′ exon and 5′ss is broken. A phosphodiester bond is formed between the BPS adenosine and the guanine of the 5′ss, resulting in an intron-3′exon lariat structure and a free 5′ exon, which remains anchored to loop I of U5 snRNA [4][114] and is stabilized by Prp8 [112]. This represents the complex C spliceosome intermediate [114][115].

Transition from step I complex C into the step II catalytically activated C* complex is facilitated by ATP-dependent RNA helicase DHX38, which triggers the release of step I factors [116] and a conformational change in the Prp8 -encapsulated active site, leading to the replacement of the lariat by the 3′ss at the active site [112]. The introduction of the 3′ss in the active site is stabilized by SLU7 [113][116]. During the second transesterification reaction, supported by Prp8/PRPF8, Prp17 and Prp18, the 3′-OH of the 5′ exon performs a nucleophilic attack on the phosphate of the 3′ exon [114]. This results in a covalent bond between the two exons and an intron lariat still bound by spliceosomal components: the post-splicing complex P.

The exon junction complex (EJC) is formed over the ligated exons and connects splicing to other downstream mRNA processes, such as export, translation and NMD [117][118]. The ligated exons are bound by U5 snRNA loop I and the 3′ end of the ligated exon pulled from the intron lariat and the spliceosome by the ATP-dependent RNA helicase Prp22/DHX8 while the intron lariat is released by the ATP-dependent RNA helicase Prp43/DHX15, giving rise to the intron lariat spliceosome intermediate and the spliced mRNA [3]. In a final step, Prp43/DHX15 releases the U2, U5 and U6 snRNPs and the NTC and NRC from the intron lariat, facilitating the recycling of these spliceosome components into the next splicing reaction.

2.2. Regulation of (Alternative) mRNA Splicing

Although for most human genes that undergo alternative splicing one mature transcript variant is usually dominant, representing at least 30% of the total transcripts, they express many splice variants simultaneously [119]. Exons that are included in all transcript variants are called constitutive exons; those that are present in only a subset of the transcripts cassette exons or alternative exons. The exclusion of these cassette exons represents one example of the most common form of AS—namely exon skipping. Other forms of AS are intron retention, selective incorporation of mutually exclusive exons and the usage of alternative 3′ or 5′ splice sites, resulting in exclusion of part of an exon or inclusion of part of an intron. The different types of AS events are illustrated in Figure 4a.

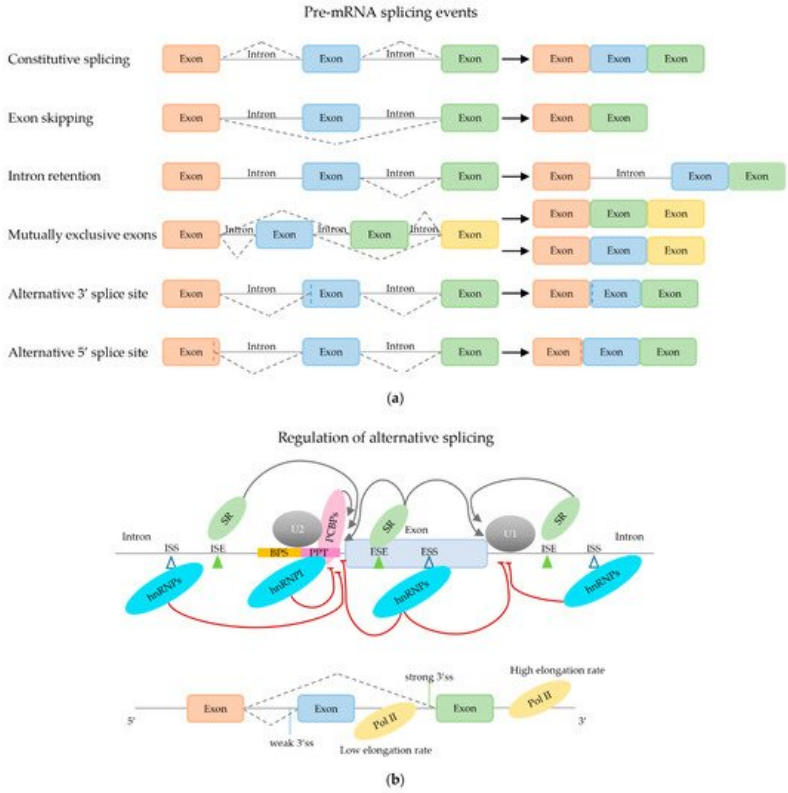

Figure 4. Pre-mRNA splicing events and regulation of alternative splicing. (**a**) Schematic representation of constitutive and common types of alternative splicing events. (**b**) Pre-mRNA splicing regulation. Upper figure: Alternative splicing is mainly regulated by trans-acting splicing factors (such as hnRNPs and SR proteins, which usually inhibit and promote splice site usage, respectively) binding to cis-acting splicing regulatory elements (ESE, ISE, ESS and ISS) on the pre-mRNA substrate. Lower figure: alternative splicing also depends on the elongation rate of RNA polymerase II (Pol II). A low elongation rate of RNA Pol II might provide more opportunities to weak 3′ss usage; while a high elongation rate of RNA Pol II might give priority to strong 3′ss usage. Open triangle: splicing silencer element; filled triangle: splicing enhancer element; arrow: promoting; –I: inhibiting.

AS is regulated through many different factors and intrinsic properties of the pre-mRNA sequence, reviewed extensively elsewhere [120][121][122] and illustrated in Figure 4b. Foremost, the splice sites themselves are involved in mRNA splicing regulation. They have limited sequence constraints. According to their compliance to the consensus splice site sequence, a subdivision can be made in “strong” and “weak” splice sites. However, the strength of a splice site is not only dependent on its sequence, but is also influenced by the gene context. This is illustrated in a study in which libraries of BRCA2, SMN1 and ELP1 minigenes harboring sets of randomized 5′ss were transfected into cells and their splicing products were analyzed by RT-PCR. The authors found that the randomized 5′ss brought about very similar splicing patterns (as demonstrated by the percent spliced-in values) within the same gene, but very different splicing patterns in the three genes [15].

The strength of the splice site determines the efficiency of recognition by the snRNPs and other splice factors and competition between splice sites leads to AS. Strong splice sites usually delineate constitutive exons. Splice site usage is influenced by cis-acting RNA sequence elements, including intronic splicing enhancers (ISEs), intronic splicing silencers (ISSs), exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs). These cis-acting regions can be bound by trans-acting regulatory proteins. The most important trans-acting factors in mRNA splicing are heterogeneous nuclear ribonucleoproteins (hnRNPs) and serine and arginine-rich (SR) proteins. SR proteins are generally regarded as splicing enhancers and predominantly bind to ESEs, whereas hnRNPs appear to inhibit splicing, through preventing assembly of the spliceosome at splice sites. Moreover, SR proteins indirectly promote splicing by impeding hnRNP-mediated repressive effects. The main cis-acting elements and trans-acting proteins are depicted in Figure 4b. However, opposing roles for these proteins have also been described and seem to depend on the sequence of the cis-acting element, on the position with regards to the target site in the pre-mRNA [123], and whether the surrounding exons are constitutive or not [124][125].

Obviously, mutations in the splice site sequences, which are associated with disease but rare in lung cancer (see Section 3), affect their recognition by the mRNA splicing machinery and thus AS. In addition, mutations elsewhere on the pre-mRNA transcript may introduce cryptic splice sites that compete with the canonical sites. Here, in particular the role of Alu retrotransposons is worth mentioning. These most abundant transposable elements in the human genome are present in most primary transcripts [126]. Alu elements contain cryptic splice sites and when they are inserted in an intron in the antisense orientation, their poly(A) tract can be recognized as PPT sequence, promoting recruitment of the spliceosome to the cryptic splice site [126]. If multiple Alu elements are integrated in a long intron they can together delineate a cryptic exon. When these cryptic sites are used by the mRNA splicing machinery, this creates a new exon. This event known as exonization contributes to evolutionary complexity. In addition, a systematic analysis of Alu elements integrated in introns near exons with rather weak splice sites showed that they can alter exon incorporation efficiencies [127].

2.2.1. Trans-Acting mRNA Splicing Factors

hnRNPs inhibit mRNA splicing through interfering with the core spliceosome. These factors accomplish this by binding to ISSs or ESSs and are thought to sterically hinder the interaction of the pre-mRNA with core splice factors. Inhibition of the spliceosomal machinery typically leads to exon exclusion. There are 32 hnRNP proteins known to date (according to KEGG, amiGO and Reactome database) that can bind distinct sequences in the pre-mRNA and were shown to play a role in the splicing of different genes. For example, hnRNPL recognizes CA-rich RNA sequences. For the CD44 gene, splicing of the variant exon 15 (or V10) was enhanced or abolished by the removal or addition of these repeats, respectively. Inhibition of splicing was explained through inhibition by hnRNPL of U2AF65 binding to the PPT [128]. Another example is hnRNPC, which was found to compete with U2AF35 for binding to uridine-rich 3′ss. This prevents aberrant inclusion of cryptic exons, amongst others for the apoptotic regulator BAX mRNA [97]. Interestingly, hnRNPC was also shown to protect the transcriptome against exonization by competing with U2AF65 for binding at cryptic splice sites created by integrated Alu elements [97]. Another RBP, hnRNPI, was speculated to act in a similar manner as hnRNPC [97]. The hnRNP-related Poly(C)-Binding Proteins (PCBPs) promote inclusion of cassette exons by binding to cytosine-rich PPTs upstream of exons [129]. Interestingly, while hnRNPs are primarily mRNA splicing inhibitors, hnRNPH can either promote or repress mRNA splicing when binding to different elements on pre-mRNAs. hnRNPH binds to poly-guanine sequences containing GGG triplets (so-called G-runs). Intronic G-runs nearby 5′ splice sites function as ISE motif, whereas G-runs in exons are commonly ESS motifs. Hence, recruitment of hnRNPH to a G-run ISE motif increases the strength of the neighboring 5′splice site thus promoting splicing [130], whereas hnRNPH inhibits splicing when it binds to a G-run ESS [131].

The SR protein family consists of 19 members (according to KEGG, amiGO and Reactome databases) and these proteins contain one or two N-terminal RNA recognition motifs (RRMs) and one arginine-serine (RS) domain consisting of at least 50 amino acids with successive RS or SR dipeptides at the C-terminus. The RRM binds certain 4–8 nucleotide consensus sequences in the target pre-mRNA while the RS domain is responsible for protein–protein interactions. With other proteins in the spliceosome. For instance, SRSF1 recognizes 5′ss through its RRM domain [132] and directly interacts with SNRNP70, recruiting the U1 snRNP to the 5′ splice site [133], where the U1 snRNP further stabilizes the interaction between SRSF1 and the pre-mRNA [132]. SR proteins bind to pre-mRNA with low affinity and specificity, which contributes to the highly dynamic nature of the spliceosome [134]. SR proteins are subject to post-translational modifications, such as phosphorylation by SR protein kinases (SRPKs) and CDC-like kinases (CLKs); and dephosphorylation by serine/threonine-protein phosphatases. Hypophosphorylation of SRSF1 results in intramolecular interactions within SRSF1 that prevent its binding to SNRNP70, decreasing the recruitment of U1 snRNP to the pre-mRNA and thereby reducing mRNA splicing. Conversely, hyperphosphorylation of SRSF1 results in the formation of a splicing-promoting complex consisting of the ESE, SRSF1 and U1 snRNP [133]. Interestingly, upon phosphorylation by CLK1, SRSF1 was

found to remain bound to CLK1. SRSF1 was released from CLK1 upon SRPK1 interacting with CLK1, thereby becoming active and able to recruit U1 snRNP to the pre-mRNA. Thus, SRPK1 and CLK1 activities were both needed to allow mRNA splicing [135].

Apart from SR and hnRNP proteins, several other proteins in the spliceosome were also shown to affect AS. For example, the complex B-specific proteins SMU1 and protein RED are involved in the splicing of short introns (i.e., introns in which the distance between the 5′ss and the BPS is relatively short). Silencing of these proteins resulted in an increased retention of this type of introns, but also in skipping of cassette exons and alternative 5′ and 3′ ss selection [136]. In another study, it was found that knockdown of SMU1 and protein RED show overlapping effects on AS [137]. For another complex B-specific protein, microfibrillar-associated protein 1, knockdown led to similar effects on intron splicing as knocking down SMU1 and protein RED, but to a different subset of introns [136]; and the KHDRBS1 protein was found to interact with U1-A protein, thereby promoting the recruitment of the U1 snRNP at 5′ss [138]. Nuclear cyclophilins are additional examples of proteins other than SR or hnRNP proteins that can regulate RNA splicing, specifically through interfering with spliceosome assembly [139]. In addition to these mRNA splicing enhancers and inhibitors that are widely expressed in many tissues, there are also highly tissue-specific trans-acting factors that are expressed almost exclusively in e.g., neurons or muscle cells. For the purpose of this review, these mRNA splicing factors are disregarded.

2.2.2. Effect of Secondary mRNA Structure

The secondary structure of the pre-mRNA substrate can also affect (alternative) splicing (reviewed in [140]). For example, stem-loop hairpin structures caused by intramolecular basepairing alter the local accessibility for trans-acting proteins. If a cis-acting enhancer or silencer motif is present in the stem of a hairpin structure, it is generally inaccessible for protein binding and thus dysfunctional for mRNA splicing regulation. Conversely, ESEs that are located immediately downstream of a hairpin structure usually exhibit strong enhancer activity. Hence, mutations in the pre-mRNA sequence that change its secondary structure by increasing or decreasing the stability of hairpins may affect splice factor binding and thus AS. In addition, mutations near splice sites may create hairpins that sequester the splice site sequence, thereby inhibiting recruitment of the spliceosome. In contrast, certain RBPs bind specifically to hairpins. For example, MBNL1 was shown to bind a hairpin sequence that contains a binding site for U2AF65 in its loop portion. MBNL1 and U2AF65 compete for binding, where MBNL1 inhibits U2AF65 binding and thus U2 snRNP recruitment when the intron adopts a hairpin structure, whereas U2AF65 binds to allow splicing when the sequence is in its single-strand fashion [141]. Another secondary RNA structure that affects splicing efficiency is the so-called G-quadruplex structure. This can be formed by a conserved sequence motif comprising at least four tracts of GG dinucleotides, folding into a helix consisting of stacked planar structures that are held together through Hoogsteen hydrogen bonding. In pull-down assays with G-quadruplex-forming RNA oligonucleotides the spliceosome proteins U2AF65, SRSF1, SRSF9, hnRNPF, hnRNPH and hnRNPU were identified [142]. The effects of G-quadruplex sequences on AS probably depend on their position within the RNA sequence and on which trans-acting protein they bind. At least for hnRNPF and hnRNPH there is suggestive evidence that their binding to G-quadruplex structures changes mRNA splicing [143][144]. Finally, RNA duplex structures formed by intramolecular interaction between sequence motifs located at sometimes very large distance in introns were found to determine AS. An example of this is the alternative exon incorporation in the FGFR2 gene causing different isoforms expressed in different cell types. The formation of the duplex structure was concluded to function solely to juxtaposition otherwise distant cis-acting elements [145].

2.2.3. Effect of mRNA Elongation Rate

As mRNA splicing occurs co-transcriptionally, the rate by which RNA polymerase II elongates the pre-mRNA also influences the mRNA splicing process (Figure 4b). Intuitively, one could argue that if transcription rate is high, producing a high concentration of pre-mRNA substrate for the splicing reaction, components of the mRNA splicing machinery could become limiting, resulting in less efficient splicing. Indeed, studies in yeast revealed that pre-mRNAs compete for the limited supply of splicing machinery components [146]. However, Ding and Elowitz observed the opposite in mammalian cells [147]. By measuring constitutive mRNA splicing efficiencies at transcription active sites in individual cells, they found that splicing efficiency increased with increasing levels of transcription. Since the efficiency of splicing regulates the nuclear export of mature mRNAs, they hypothesized that this effect amplifies the expression of more strongly transcribed genes and reduces expression of low-level transcribed genes, thus contributing to gene expression modulation. Conversely, mRNA splicing also influences RNA elongation. Alexander et al. observed in yeast that RNA polymerase II paused periodically around the 3′ end of introns [148]. This coincided with—and was dependent on—the recruitment of U2 and U5 snRNPs. RNA polymerase II that stalled near the 3′splice site was hyperphosphorylated in its carboxy-terminal domain known to recruit RNA processing factors. The authors propose that the mRNA splicing machinery controls a transcription checkpoint that is activated during formation of the step II catalytically activated spliceosome and released upon completion of the second step of the RNA splicing reaction. Although there are many candidates, the RNA splicing

factor responsible for activating the proposed transcription checkpoint and its possible association with RNA polymerase II has not yet been ascertained.

The RNA elongation rate might also affect the fidelity of the splicing reaction and splice site preference. Regulation of transcription initiation and elongation is reviewed elsewhere [149]. Critical factors controlling productive elongation by RNA polymerase II include nucleosome positioning and histone modification. These epigenetic processes influence the speed at which RNA polymerase II can move forward along the DNA template. How this may affect mRNA splicing is already extensively discussed by others [150]. During transcription, for any set of alternative splice sites, the upstream located splice site is created sooner and thus available for recognition by the RNA splicing machinery earlier, giving it a selection advantage over more downstream splice sites. Thus, in theory, the slower the elongation rate, the longer the time difference in creation of the competing splice sites; and thus the larger the preference for upstream splice site selection. However, the kinetic coupling of pre-mRNA elongation and splicing processes appears more complicated. A genome-wide analysis of AS at different elongation rates revealed that, while the elongation rate clearly affected both constitutive and alternative mRNA splicing, elongation-rate dependent AS events were not always consistent with this reasoning [151]. Exons that were included more frequently at slow elongation rates usually had weaker splice sites than those that were included more often at high elongation rates. However, when the elongation rate was experimentally increased or reduced, AS of many genes was affected in the same way, rather than having opposing effects, suggesting that a proper balance of RNA splice variants requires an optimal RNA transcription rate [151]. It also suggests that mRNA splicing patterns could change if the transcription rate is changed in response to physiological stimuli or disease processes. This, obviously, is relevant in cancer, where transcription elongation rates of many genes are altered by oncogene expression [152][153].

2.2.4. Effect of Chromatin Structure

Apart from epigenetic processes affecting mRNA splicing through the kinetic coupling of transcription and splicing discussed above, there is also evidence that chromatin organization and histone modifications have more direct effects on mRNA splicing, by contributing to exon definition and splice site choice (reviewed in [154][155]). Nucleosomes are found enriched at exon-intron junctions suggesting that they play a role in exon definition [156][157]. Moreover, in alternatively spliced genes they are more highly enriched around included exons than around excluded ones. Together, this strongly argues for a function of nucleosome positioning in regulating splicing [156]. Also, certain histone modification marks are found enriched at exons, more than would be expected as a consequence of nucleosome distribution [157]. For some histone modifications a mechanism by which they modulate mRNA splicing was revealed. For example, H3K4me3 was shown to bind the U2 snRNA via the chromatin remodeling protein CHD1; and this promoted recruitment of the U2 snRNP to the pre-mRNA branch site [158]. Another example is the recruitment of hnRNPI via MRG15 binding to H3K36me3 [159]. Thus, histone modifications appear to influence AS by recruiting mRNA splicing factors to the pre-mRNA substrate via chromatin-binding adapter proteins.

2.3. Other RNA Processing Activities in the Supraspliceosome

In addition to the mRNA splicing machinery, the supraspliceosome harbors associated enzymes with 5′-RNA end capping, 3′-RNA end cleavage, RNA polyadenylation and RNA base editing activities (Figure 5). The co-transcriptional conjunction of these processes in close proximity in the supraspliceosome makes many of these processes interdependent. In this section, we discuss these processes and their effects on mRNA splicing.

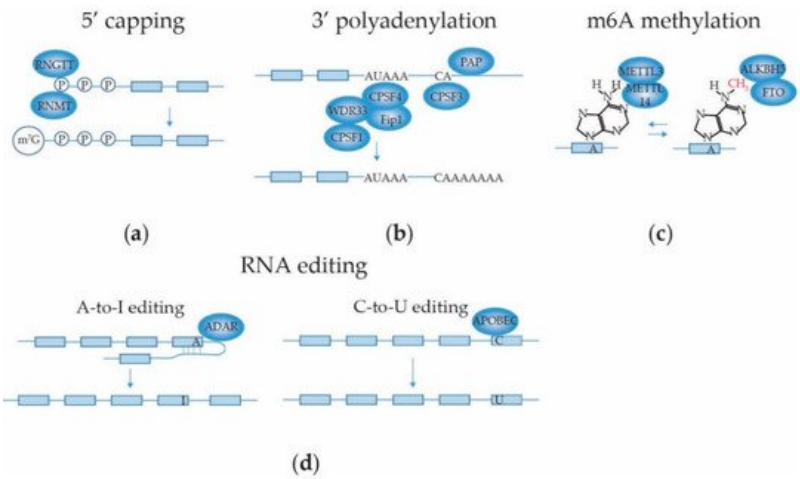

Figure 5. Illustration of other (pre-)mRNA processing activities performed in the supraspliceosome simultaneous with pre-mRNA splicing. (**a**) 5′ capping, (**b**) 3′ polyadenylation, (**c**) m6A methylation, (**d**) A-to-I and C-to-U RNA base editing. Details

are given in the main text.

2.3.1. mRNA 5′ Capping

RNA 5′ capping (Figure 5a) involves the attachment of an N7-methylated guanosine (m7G) to the first nucleotide of the pre-mRNA, which protects it from 5′ to 3′ exonuclease cleavage, supports the initiation of translation and recruits the RNA splicing machinery as well as the factors required for 3′ polyadenylation and nuclear export. The 5′ capping takes place in three steps, in humans catalyzed by RNA Guanylyltransferase and 5′-Phosphatase (RNGTT) and RNA guanine N7-methyltransferase (RNMT) (reviewed by [160]). Additionally, the +1 nucleotide can be methylated at the 2′O of its ribose by mRNA Cap 2′-O-methyltransferase. The cap structure is bound by the cap binding complex (CBC) in the nucleus. The CBC proteins CBP20 and CBP80 were found in the supraspliceosome [10]. The CBC recruits spliceosomes and factors involved in 3′ end processing, RNA export and NMD to the mRNA. After guiding the transport of the processed mRNA into the cytoplasm, the CBC is replaced by the eukaryotic translation initiation factor 4E (eIF4E) to start mRNA translation [160].

2.3.2. mRNA 3′ End Cleavage and Polyadenylation

The mechanism and structural details of the 3′RNA end cleavage and polyadenylation machinery that adds a 3′ polyadenosine (poly(A)) tail to most mRNAs to safeguard mRNA stability and regulate translation are reviewed by Kumar et al. [161] and is illustrated in Figure 5b. Co-transcriptionally, the AAUAAA hexanucleotide within the polyadenylation signal (PAS) on the pre-mRNA is recognized by a protein complex composed of the cleavage and polyadenylation specificity factor 1 (CPSF1), CPSF4, WD repeat domain 33 (WDR33) and pre-mRNA 3′-end-processing factor FIP1. The CPSF proteins associate with the carboxy-terminal domain of RNA polymerase II to regulate their activities on the mRNA. This leads to the cleavage of the pre-mRNA, 10–30 nucleotides downstream of the PAS, by CPSF3. Finally, Poly(A) polymerase (PAP) interacts with FIP1 and adds the poly(A) tail to the cleaved (pre-)mRNA. 3′-end processing complexes involved in the cleavage of the pre-mRNA downstream of the PAS, i.e., NUDT21, CPSF6, CPSF7 [24][162], CPSF1, 2, 3, 4 and FIP1 [24][162], CSTF 1, 2 and 3 and symplekin [10][24][162] and polyadenylate-binding protein 2 [24][162] were all shown to be present in the supraspliceosome. Most of the mRNA splicing reactions on the pre-mRNA are completed before 3′RNA end cleavage and polyadenylation occur, but this may be completed for some introns on the polyadenylated RNA [12]. Co-transcriptionally, U1 snRNP not only plays a crucial role in the mRNA splicing reaction by defining the 5′ss of introns, but also protects the pre-mRNA from premature cleavage and polyadenylation at cryptic PASs in introns; a process coined telescripting [163][164]. Telescripting requires specific base pairing of the U1 snRNA to the pre-mRNA. Presumably, the binding of U1 snRNP to the intron physically competes with binding of the cleavage/polyadenylation machinery at nearby sites. Since 5′ splice sites could be at too large a distance of cryptic PASs in long introns, cryptic 5′ splice sites may primarily serve to protect the pre-mRNA from premature termination by recruiting U1 snRNP [164]. Modulation of PAS suppression by U1 snRNP also affected alternative splicing patterns [163].

2.3.3. mRNA Internal Adenosine Methylation

The most prevalent internal mRNA modification in eukaryotes is methylation of adenosine residues at their N6-position (depicted as m^6A) in certain consensus sequence motifs that are enriched near stop codons and in long exons [165] (Figure 5c). The m^6A modification is important for gene expression regulation and is reversible. It is installed co-transcriptionally on the mRNA by methyltransferases, most notably the nuclear complex consisting of METTL3/METTL14 heterodimers with accessory proteins; and can be removed by the demethylases FTO and ALKBH5. Dysregulated expression of these m^6A modification enzymes as well as of proteins that bind to m^6A sites is associated with cancer [166]. While many effects of m^6A modification on gene expression involve nuclear export and cytosolic processes such as mRNA stability and translation efficiency, which for the purpose of this review are disregarded, m^6A modification also affects mRNA splicing. The modification changes the mRNA structure, increasing the accessibility for binding of mRNA splicing silencers hnRNPC and hnRNPG [167][168]. Consequently, knockdown of METTL3 or METTL14, which reduces m^6A modification, induced AS with similar patterns as knockdown of hnRNPs. Conversely, depletion of demethylases ALKBH5 or FTO increased m^6A levels and changed mRNA splicing patterns [169][170]. ALKBH5 expression was shown to promote nuclear localization of SRPK1 and this was associated with increased levels of phosphorylated SRSF1 [170], which activates this SR protein to recruit U1 snRNP [135]. Depletion of FTO promoted recruitment of SRSF2 and increased exon inclusion, presumably by increasing methylation of adenosines in ESEs [169]. Furthermore, the nuclear m^6A-binding protein YTDHC1 was shown to bind several members of the SR family of proteins [171]. When YTDHC1 bound to an m^6A motif in an exon, this promoted binding of SRSF3 to the mRNA and exon inclusion. At the same time, YTDHC1 inhibited binding of SRSF10, which would otherwise promote exon skipping. Hence, m^6A modification clearly favored inclusion of the exon. Together, these observations suggest that structural changes in the pre-mRNA imposed by methylation contribute to creating cis-acting signals for regulating mRNA splicing.

2.3.4. mRNA Base Editing

RNA editing consists of adenosine-to-inosine (A-to-I) substitution, cytosine-to-uridine (C-to-U) substitution, U insertion/deletion, C or guanosine (G) insertion, 2′-O-methyl ribose nucleotide modification, and U-to-pseudo U conversion, all resulting in an altered RNA sequence compared to the genetic code on the DNA template [172]. Of these, A-to-I and C-to-U editing (Figure 5d) are most thoroughly described. A-to-I editing is the most prevalent form of RNA editing in mammalian cells. The reaction is catalyzed by C6-adenosine deaminases acting on RNA (ADARs). Three mammalian ADARs are known to date (ADAR1-3), of which ADAR1 and 2 possess catalytic activity, propagated through their C-terminal deaminase catalytic domains. Site-specificity for editing in exons is dictated by cis-acting complementary repeats in the intronic sequence (the editing site complementary sequence; ECS) adjacent to the exon containing the adenosine that is to be edited. An imperfect duplex structure is formed between the intronic sequence and spanning the exonic editing site, which is essential for recognition by ADARs through their double-stranded RNA (dsRNA) binding motifs [173]. C-to-U editing is performed by the cytidine deaminase apolipoprotein B mRNA editing enzyme (APOBEC) family. APOBEC1 binds to single-stranded RNA through its cytidine deaminase domain. Editing activity of APOBEC1 is dependent on a cis-acting 11 nucleotide sequence adjacent to the editing site and guidance by RNA-binding protein cofactor A1CF towards the editing site [174]. At least the RNA editing enzymes ADAR1 and ADAR2 were found in the supraspliceosome, in association with Sm and SR proteins present in the spliceosomes, while retaining their A-to-I editing activity [11]. RNA base editing can influence simultaneous mRNA splicing and vice versa. While most A-to-I editing sites are situated in non-coding regions of the mRNA, such as intronic short Alu repeats and 3′-untranslated regions (3′-UTRs) [175], editing may change splicing factor recognition sites or cis-acting elements on the template mRNA, thereby changing the efficiency of splice site recognition and thus causing AS. The mRNA splicing machinery recognizes an inosine as guanine. Thus, A-to-I editing at an intronic AA dinucleotide can create an AI sequence that effectively mimics the conserved AG sequence found at 3′ splice sites [176] creating an alternative splice acceptor site, or, conversely, the strength of a genuine 3′ss can be dramatically reduced if it is converted from AG to IG [177]. On the other hand, mRNA splicing may remove the intronic ECS required to form the dsRNA editing template and thus inhibit RNA editing in an exon. In addition, mRNA splicing factors and base editor enzymes in the supraspliceosome could compete for binding to the mRNA template through steric hindrance. For example, it has been observed that binding of SRSF9 to a mouse minigene containing the ECS of the voltage-gated calcium channel CaV1.3 mRNA inhibited A-to-I editing by ADAR2 [178]. Conversely, binding of ADAR2 to a dsRNA template formed by intron 2 and exon 3 sequences of the RELL2 gene blocked access of splicing factor U2AF35 to the 3′ splice site, thereby reducing exon 3 inclusion [179]. This exon skipping event led to degradation of RELL2 transcripts through NMD, possibly contributing to tumorigenesis [179]. RNA splicing can also influence RNA editing via AS of ADAR transcripts. For instance, ADAR2 pre-mRNA carries a cassette exon in intron 7 that contains multiple stop codons through which transcripts with this exon are targeted for NMD. In tissues in which inclusion of the cassette exon is high, ADAR2 expression is decreased and ADAR2-mediated A-to-I editing is low [180].

2.3.5. Intronic Pri-miRNAs

Intronic pri-miRNAs and key microprocessor proteins Drosha and DGCR8 were found associated with the supraspliceosome [181]. This suggested that intronic pri-miRNAs are processed in the supraspliceosome. Indeed, for the miR-106b-25 cluster that is located in intron 13 of the MCM7 pre-mRNA and harbors miRNAs 106b, 93 and 25, processing from pri-miRNA to pre-miRNA in the supraspliceosome was demonstrated [181]. Interestingly, intronic miRNA processing and mRNA splicing appeared interdependent. Usage of an alternative 3′ splice acceptor site in MCM7 intron 13 added the miR-25 sequences to an extended exon 14, rather than splicing this section out and subjecting it to the microprocessor complex. Hence, this AS event resulted in lower miR-25 levels. Specific inhibition of this AS event using antisense morpholinos, favoring inclusion of the miR-25 pre-miRNA hairpin sequences in the intron, increased miR-25 expression. General inhibition of mRNA splicing using the SF3b1 inhibitor spliceostatin A reduced alternative splice site usage more than constitutive splice site usage, thereby also increasing miRNA expression. Conversely, inhibition of miRNA processing through silencing of Drosha increased AS, possibly because the alternative splice site near the hairpin in the intron is more accessible to the splicing machinery in the absence of Drosha [181]. Together, these observations support the notion that intronic pri-miRNAs are processed in the supraspliceosome in conjunction with the mRNA splicing process before being handed over to the nuclear export machinery. A comprehensive analysis of miRNAs in the supraspliceosome revealed that while most were intronic miRNAs, a substantial fraction was of intergenic origin [182]. In addition, the majority were mature miRNAs that are apparently shuttled from the cytoplasm to the nucleus; and most of these were not detected in precursor format [182]. Many miRNAs were found enriched in the supraspliceosome, suggesting selective association with the supraspliceosome. Hence, their presence in the supraspliceosome is not a simple reflection of their production in the supraspliceosome during the mRNA splicing process, but suggests an undiscovered function for miRNAs in this complex. Compelling evidence was obtained to suggest that at least some of these miRNAs regulate gene expression and perhaps splicing [182].

2.3.6. mRNA Transport

Aside from proteins with direct roles in RNA processing and editing, factors involved in mRNA transport (TRanscription-EXport (TREX) complex) and surveillance and localization (EJC) were found to be associated with the supraspliceosome [162]. Interestingly, also nuclear matrix and filament proteins were detected, as well as proteins of the nuclear pore complex. It was speculated that these proteins might play a role in guiding the mRNA from the splicing machinery to the nuclear pores.

Recently, more striking examples of the interaction between mRNA splicing and other mRNA processing steps were described. U1 snRNP was found to be involved in the nuclear retention of long noncoding RNAs (lncRNAs), by tethering these molecules to chromatin. This occurs co-transcriptionally and is dependent on RNA polymerase II. Interestingly, inhibition of U2 snRNP through treatment with the splicing inhibitor E7107 displayed similar, albeit more subtle effects on chromatin retention of lncRNAs as compared to interfering with U1 snRNP. In contrast, inhibition of the U4/U6.U5 tri-snRNP by the drug isoginkgetin did not have any effect [183]. In another study, U1 proteins SNRNP70 and SNRPA, but also SmD2 were shown to be involved in nuclear retention of spliced lncRNAs [184]. Hence, U1 appears to display multiple functions to ensure transcriptome integrity distinct from mRNA splicing. For the SF3b complex, a role in mRNA export was also described. This is, however, independent of U2 snRNP and is achieved through interacting with the TREX complex [185]. Finally, the U4/U6.U5 tri-snRNP has been implicated in maintaining proper chromatin cohesion and enabling mitotic progression through its interaction with cohesion [186].

Together, the supraspliceosome offers a platform for coordinated mRNA processing, splicing, editing, surveillance and transport. This machine controls all nuclear steps necessary to obtain mature mRNA that can be translated to proteins in the cytoplasm.

References

1. Berget, S.M.; Moore, C.; Sharp, P.A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. USA 1977, 74, 3171–3175.
2. Chow, L.T.; Gelinas, R.E.; Broker, T.R.; Roberts, R.J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell 1977, 12, 1–8.
3. Bai, R.; Yan, C.; Wan, R.; Lei, J.; Shi, Y. Structure of the Post-catalytic Spliceosome from Saccharomyces cerevisiae. C ell 2017, 171, 1589–1598.e8.
4. Wan, R.; Yan, C.; Bai, R.; Huang, G.; Shi, Y. Structure of a yeast catalytic step I spliceosome at 3.4 A resolution. Scienc e 2016, 353, 895–904.
5. Yan, C.; Hang, J.; Wan, R.; Huang, M.; Wong, C.C.; Shi, Y. Structure of a yeast spliceosome at 3.6-angstrom resolutio n. Science 2015, 349, 1182–1191.
6. Zhan, X.; Yan, C.; Zhang, X.; Lei, J.; Shi, Y. Structure of a human catalytic step I spliceosome. Science 2018, 359, 537–545.
7. Bentley, D.L. Rules of engagement: Co-transcriptional recruitment of pre-mRNA processing factors. Curr. Opin. Cell Bio l. 2005, 17, 251–256.
8. Azubel, M.; Habib, N.; Sperling, R.; Sperling, J. Native spliceosomes assemble with pre-mRNA to form supraspliceoso mes. J. Mol. Biol. 2006, 356, 955–966.
9. Muller, S.; Wolpensinger, B.; Angenitzki, M.; Engel, A.; Sperling, J.; Sperling, R. A supraspliceosome model for large nu clear ribonucleoprotein particles based on mass determinations by scanning transmission electron microscopy. J. Mol. Biol. 1998, 283, 383–394.
10. Raitskin, O.; Angenitzki, M.; Sperling, J.; Sperling, R. Large nuclear RNP particles—The nuclear pre-mRNA processing machine. J. Struct. Biol. 2002, 140, 123–130.
11. Raitskin, O.; Cho, D.S.; Sperling, J.; Nishikura, K.; Sperling, R. RNA editing activity is associated with splicing factors in lnRNP particles: The nuclear pre-mRNA processing machinery. Proc. Natl. Acad. Sci. USA 2001, 98, 6571–6576.
12. Girard, C.; Will, C.L.; Peng, J.; Makarov, E.M.; Kastner, B.; Lemm, I.; Urlaub, H.; Hartmuth, K.; Luhrmann, R. Post-trans criptional spliceosomes are retained in nuclear speckles until splicing completion. Nat. Commun. 2012, 3, 994.
13. Shefer, K.; Sperling, J.; Sperling, R. The Supraspliceosome—A Multi-Task Machine for Regulated Pre-mRNA Processin g in the Cell Nucleus. Comput. Struct. Biotechnol. J. 2014, 11, 113–122.
14. Roca, X.; Krainer, A.R.; Eperon, I.C. Pick one, but be quick: 5′ splice sites and the problems of too many choices. Gene s Dev. 2013, 27, 129–144.

15. Wong, M.S.; Kinney, J.B.; Krainer, A.R. Quantitative Activity Profile and Context Dependence of All Human 5′ Splice Sites. Mol. Cell 2018, 71, 1012–1026.e3.

16. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtukova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative isoform regulation in human tissue transcriptomes. Nature 2008, 456, 470–476.

17. Tung, K.F.; Pan, C.Y.; Chen, C.H.; Lin, W.C. Top-ranked expressed gene transcripts of human protein-coding genes investigated with GTEx dataset. Sci. Rep. 2020, 10, 16245.

18. Fiszbein, A.; Kornblihtt, A.R. Alternative splicing switches: Important players in cell differentiation. Bioessays 2017, 39.

19. Maniatis, T.; Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 2002, 418, 236–243.

20. Noh, S.J.; Lee, K.; Paik, H.; Hur, C.G. TISA: Tissue-specific alternative splicing in human and mouse genes. DNA Res. 2006, 13, 229–243.

21. Singh, R.K.; Cooper, T.A. Pre-mRNA splicing in disease and therapeutics. Trends Mol. Med. 2012, 18, 472–482.

22. Kahles, A.; Lehmann, K.V.; Toussaint, N.C.; Huser, M.; Stark, S.G.; Sachsenberg, T.; Stegle, O.; Kohlbacher, O.; Sander, C.; The Cancer Genome Atlas Research Network; et al. Comprehensive Analysis of Alternative Splicing Across Tumors from 8705 Patients. Cancer Cell 2018, 34, 211–224.e6.

23. Sperling, R.; Koster, A.J.; Melamed-Bessudo, C.; Rubinstein, A.; Angenitzki, M.; Berkovitch-Yellin, Z.; Sperling, J. Three-dimensional image reconstruction of large nuclear RNP (lnRNP) particles by automated electron tomography. J. Mol. Biol. 1997, 267, 570–583.

24. Kotzer-Nevo, H.; de Lima Alves, F.; Rappsilber, J.; Sperling, J.; Sperling, R. Supraspliceosomes at defined functional states portray the pre-assembled nature of the pre-mRNA processing machine in the cell nucleus. Int. J. Mol. Sci. 2014, 15, 11637–11664.

25. Cohen-Krausz, S.; Sperling, R.; Sperling, J. Exploring the architecture of the intact supraspliceosome using electron microscopy. J. Mol. Biol. 2007, 368, 319–327.

26. Frankenstein, Z.; Sperling, J.; Sperling, R.; Eisenstein, M. A unique spatial arrangement of the snRNPs within the native spliceosome emerges from in silico studies. Structure 2012, 20, 1097–1106.

27. Sharp, P.A.; Burge, C.B. Classification of introns: U2-type or U12-type. Cell 1997, 91, 875–879.

28. Fischer, U.; Englbrecht, C.; Chari, A. Biogenesis of spliceosomal small nuclear ribonucleoproteins. Wiley Interdiscip. Rev. RNA 2011, 2, 718–731.

29. Kiss, T. Biogenesis of small nuclear RNPs. J. Cell Sci. 2004, 117, 5949–5951.

30. Matera, A.G.; Wang, Z. A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. 2014, 15, 108–121.

31. Coady, T.H.; Lorson, C.L. SMN in spinal muscular atrophy and snRNP biogenesis. Wiley Interdiscip. Rev. RNA 2011, 2, 546–564.

32. Galganski, L.; Urbanek, M.O.; Krzyzosiak, W.J. Nuclear speckles: Molecular organization, biological function and role in disease. Nucleic Acids Res. 2017, 45, 10350–10368.

33. Smith, K.P.; Hall, L.L.; Lawrence, J.B. Nuclear hubs built on RNAs and clustered organization of the genome. Curr. Opin. Cell Biol. 2020, 64, 67–76.

34. Ghaemi, Z.; Peterson, J.R.; Gruebele, M.; Luthey-Schulten, Z. An in-silico human cell model reveals the influence of spatial organization on RNA splicing. PLoS Comput. Biol. 2020, 16, e1007717.

35. Veretnik, S.; Wills, C.; Youkharibache, P.; Valas, R.E.; Bourne, P.E. Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome. PLoS Comput. Biol. 2009, 5, e1000315.

36. Lee, M.S.; Lin, Y.S.; Deng, Y.F.; Hsu, W.T.; Shen, C.C.; Cheng, Y.H.; Huang, Y.T.; Li, C. Modulation of alternative splicing by expression of small nuclear ribonucleoprotein polypeptide N. FEBS J. 2014, 281, 5194–5207.

37. Hermann, H.; Fabrizio, P.; Raker, V.A.; Foulaki, K.; Hornig, H.; Brahms, H.; Luhrmann, R. snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. EMBO J. 1995, 14, 2076–2088.

38. Li, J.; Leung, A.K.; Kondo, Y.; Oubridge, C.; Nagai, K. Re-refinement of the spliceosomal U4 snRNP core-domain structure. Acta Crystallogr. D Struct. Biol. 2016, 72, 131–146.

39. Raker, V.A.; Hartmuth, K.; Kastner, B.; Luhrmann, R. Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. Mol. Cell. Biol. 1999, 19, 6554–6565.

40. Neuenkirchen, N.; Englbrecht, C.; Ohmer, J.; Ziegenhals, T.; Chari, A.; Fischer, U. Reconstitution of the human U snRN P assembly machinery reveals stepwise Sm protein organization. EMBO J. 2015, 34, 1925–1941.

41. Prusty, A.B.; Meduri, R.; Prusty, B.K.; Vanselow, J.; Schlosser, A.; Fischer, U. Impaired spliceosomal UsnRNP assembly leads to Sm mRNA down-regulation and Sm protein degradation. J. Cell Biol. 2017, 216, 2391–2407.

42. Deng, X.; Lu, T.; Wang, L.; Gu, L.; Sun, J.; Kong, X.; Liu, C.; Cao, X. Recruitment of the NineTeen Complex to the activ ated spliceosome requires AtPRMT5. Proc. Natl. Acad. Sci. USA 2016, 113, 5447–5452.

43. Gonsalvez, G.B.; Tian, L.; Ospina, J.K.; Boisvert, F.M.; Lamond, A.I.; Matera, A.G. Two distinct arginine methyltransfera ses are required for biogenesis of Sm-class ribonucleoproteins. J. Cell Biol. 2007, 178, 733–740.

44. Grimm, C.; Chari, A.; Pelz, J.P.; Kuper, J.; Kisker, C.; Diederichs, K.; Stark, H.; Schindelin, H.; Fischer, U. Structural bas is of assembly chaperone- mediated snRNP formation. Mol. Cell 2013, 49, 692–703.

45. Yi, H.; Mu, L.; Shen, C.; Kong, X.; Wang, Y.; Hou, Y.; Zhang, R. Negative cooperativity between Gemin2 and RNA provi des insights into RNA selection and the SMN complex's release in snRNP assembly. Nucleic Acids Res. 2020, 48, 895 –911.

46. Ma, Y.; Dostie, J.; Dreyfuss, G.; Van Duyne, G.D. The Gemin6-Gemin7 heterodimer from the survival of motor neurons complex has an Sm protein-like structure. Structure 2005, 13, 883–892.

47. Burghes, A.H.; Beattie, C.E. Spinal muscular atrophy: Why do low levels of survival motor neuron protein make motor n eurons sick? Nat. Rev. Neurosci. 2009, 10, 597–609.

48. Chen, T.; Zhang, B.; Ziegenhals, T.; Prusty, A.B.; Frohler, S.; Grimm, C.; Hu, Y.; Schaefke, B.; Fang, L.; Zhang, M.; et a l. A missense mutation in SNRPE linked to non-syndromal microcephaly interferes with U snRNP assembly and pre-m RNA splicing. PLoS Genet. 2019, 15, e1008460.

49. Jin, W.; Wang, Y.; Liu, C.P.; Yang, N.; Jin, M.; Cong, Y.; Wang, M.; Xu, R.M. Structural basis for snRNA recognition by t he double-WD40 repeat domain of Gemin5. Genes Dev. 2016, 30, 2391–2403.

50. Tang, X.; Bharath, S.R.; Piao, S.; Tan, V.Q.; Bowler, M.W.; Song, H. Structural basis for specific recognition of pre-snR NA by Gemin5. Cell Res. 2016, 26, 1353–1356.

51. Xu, C.; Ishikawa, H.; Izumikawa, K.; Li, L.; He, H.; Nobe, Y.; Yamauchi, Y.; Shahjee, H.M.; Wu, X.H.; Yu, Y.T.; et al. Stru ctural insights into Gemin5-guided selection of pre-snRNAs for snRNP assembly. Genes Dev. 2016, 30, 2376–2390.

52. Yong, J.; Kasim, M.; Bachorik, J.L.; Wan, L.; Dreyfuss, G. Gemin5 delivers snRNA precursors to the SMN complex for s nRNP biogenesis. Mol. Cell 2010, 38, 551–562.

53. So, B.R.; Wan, L.; Zhang, Z.; Li, P.; Babiash, E.; Duan, J.; Younis, I.; Dreyfuss, G. A U1 snRNP-specific assembly path way reveals the SMN complex as a versatile hub for RNP exchange. Nat. Struct. Mol. Biol. 2016, 23, 225–230.

54. Weber, G.; Trowitzsch, S.; Kastner, B.; Luhrmann, R.; Wahl, M.C. Functional organization of the Sm core in the crystal structure of human U1 snRNP. EMBO J. 2010, 29, 4172–4184.

55. Lund, E.; Dahlberg, J.E. Cyclic 2′,3′-phosphates and nontemplated nucleotides at the 3′ end of spliceosomal U6 small n uclear RNA's. Science 1992, 255, 327–330.

56. Sontheimer, E.J.; Steitz, J.A. Three novel functional variants of human U5 small nuclear RNA. Mol. Cell Biol. 1992, 12, 734–746.

57. Urlaub, H.; Raker, V.A.; Kostka, S.; Luhrmann, R. Sm protein-Sm site RNA interactions within the inner ring of the splic eosomal snRNP core structure. EMBO J. 2001, 20, 187–196.

58. Schwer, B.; Kruchten, J.; Shuman, S. Structure-function analysis and genetic interactions of the SmG, SmE, and SmF subunits of the yeast Sm protein ring. RNA 2016, 22, 1320–1328.

59. Mouaikel, J.; Narayanan, U.; Verheggen, C.; Matera, A.G.; Bertrand, E.; Tazi, J.; Bordonne, R. Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. EMBO Rep. 2003, 4, 616–622.

60. Huber, J.; Dickmanns, A.; Luhrmann, R. The importin-beta binding domain of snurportin1 is responsible for the Ran- an d energy-independent nuclear import of spliceosomal U snRNPs in vitro. J. Cell Biol. 2002, 156, 467–479.

61. Roithova, A.; Klimesova, K.; Panek, J.; Will, C.L.; Luhrmann, R.; Stanek, D.; Girard, C. The Sm-core mediates the reten tion of partially-assembled spliceosomal snRNPs in Cajal bodies until their full maturation. Nucleic Acids Res. 2018, 46, 3774–3790.

62. Aoyama, T.; Yamashita, S.; Tomita, K. Mechanistic insights into m6A modification of U6 snRNA by human METTL16. N ucleic Acids Res. 2020, 48, 5157–5168.

63. Goh, Y.T.; Koh, C.W.Q.; Sim, D.Y.; Roca, X.; Goh, W.S.S. METTL4 catalyzes m6Am methylation in U2 snRNA to regula te pre-mRNA splicing. Nucleic Acids Res. 2020, 48, 9250–9261.

64. Mauer, J.; Sindelar, M.; Despic, V.; Guez, T.; Hawley, B.R.; Vasseur, J.J.; Rentmeister, A.; Gross, S.S.; Pellizzoni, L.; De bart, F.; et al. FTO controls reversible m(6)Am RNA methylation during snRNA biogenesis. Nat. Chem. Biol. 2019, 15, 3 40–347.

65. Pomeranz Krummel, D.A.; Oubridge, C.; Leung, A.K.; Li, J.; Nagai, K. Crystal structure of human spliceosomal U1 snR NP at 5.5 A resolution. Nature 2009, 458, 475–480.

66. Sun, S.; Ling, S.C.; Qiu, J.; Albuquerque, C.P.; Zhou, Y.; Tokunaga, S.; Li, H.; Qiu, H.; Bui, A.; Yeo, G.W.; et al. ALS-cau sative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. Nat. Co mmun. 2015, 6, 6171.

67. Li, X.; Liu, S.; Jiang, J.; Zhang, L.; Espinosa, S.; Hill, R.C.; Hansen, K.C.; Zhou, Z.H.; Zhao, R. CryoEM structure of Sac charomyces cerevisiae U1 snRNP offers insight into alternative splicing. Nat. Commun. 2017, 8, 1035.

68. Nelissen, R.L.; Will, C.L.; van Venrooij, W.J.; Luhrmann, R. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. EMBO J. 1994, 13, 4113–4125.

69. Kondo, Y.; Oubridge, C.; van Roon, A.M.; Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleopr otein particle, reveals the mechanism of 5′ splice site recognition. Elife 2015, 4.

70. Rosel-Hillgartner, T.D.; Hung, L.H.; Khrameeva, E.; Le Querrec, P.; Gelfand, M.S.; Bindereif, A. A novel intra-U1 snRNP cross-regulation mechanism: Alternative splicing switch links U1C and U1-70K expression. PLoS Genet. 2013, 9, e100 3856.

71. Zhang, Z.; Will, C.L.; Bertram, K.; Dybkov, O.; Hartmuth, K.; Agafonov, D.E.; Hofele, R.; Urlaub, H.; Kastner, B.; Luhrma nn, R.; et al. Molecular architecture of the human 17S U2 snRNP. Nature 2020, 583, 310–313.

72. Martelly, W.; Fellows, B.; Senior, K.; Marlowe, T.; Sharma, S. Identification of a noncanonical RNA binding domain in th e U2 snRNP protein SF3A1. RNA 2019, 25, 1509–1521.

73. Crisci, A.; Raleff, F.; Bagdiul, I.; Raabe, M.; Urlaub, H.; Rain, J.C.; Kramer, A. Mammalian splicing factor SF1 interacts with SURP domains of U2 snRNP-associated proteins. Nucleic Acids Res. 2015, 43, 10456–10473.

74. Cretu, C.; Schmitzova, J.; Ponce-Salvatierra, A.; Dybkov, O.; De Laurentiis, E.I.; Sharma, K.; Will, C.L.; Urlaub, H.; Luhr mann, R.; Pena, V. Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. Mol. Cell 2016, 64, 307–319.

75. Kfir, N.; Lev-Maor, G.; Glaich, O.; Alajem, A.; Datta, A.; Sze, S.K.; Meshorer, E.; Ast, G. SF3B1 association with chroma tin determines splicing outcomes. Cell Rep. 2015, 11, 618–629.

76. Sander, B.; Golas, M.M.; Makarov, E.M.; Brahms, H.; Kastner, B.; Luhrmann, R.; Stark, H. Organization of core spliceo somal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomic roscopy. Mol. Cell 2006, 24, 267–278.

77. Haselbach, D.; Komarov, I.; Agafonov, D.E.; Hartmuth, K.; Graf, B.; Dybkov, O.; Urlaub, H.; Kastner, B.; Luhrmann, R.; Stark, H. Structure and Conformational Dynamics of the Human Spliceosomal B(act) Complex. Cell 2018, 172, 454–46 4.e11.

78. Agafonov, D.E.; Kastner, B.; Dybkov, O.; Hofele, R.V.; Liu, W.T.; Urlaub, H.; Luhrmann, R.; Stark, H. Molecular architect ure of the human U4/U6.U5 tri-snRNP. Science 2016, 351, 1416–1420.

79. Hardin, J.W.; Warnasooriya, C.; Kondo, Y.; Nagai, K.; Rueda, D. Assembly and dynamics of the U4/U6 di-snRNP by sin gle-molecule FRET. Nucleic Acids Res. 2015, 43, 10963–10974.

80. Cloutier, P.; Poitras, C.; Durand, M.; Hekmat, O.; Fiola-Masson, E.; Bouchard, A.; Faubert, D.; Chabot, B.; Coulombe, B. R2TP/Prefoldin-like component RUVBL1/RUVBL2 directly interacts with ZNHIT2 to regulate assembly of U5 small n uclear ribonucleoprotein. Nat. Commun. 2017, 8, 15615.

81. Malinova, A.; Cvackova, Z.; Mateju, D.; Horejsi, Z.; Abeza, C.; Vandermoere, F.; Bertrand, E.; Stanek, D.; Verheggen, C. Assembly of the U5 snRNP component PRPF8 is controlled by the HSP90/R2TP chaperones. J. Cell Biol. 2017, 21 6, 1579–1596.

82. Yamashita, S.; Takagi, Y.; Nagaike, T.; Tomita, K. Crystal structures of U6 snRNA-specific terminal uridylyltransferase. Nat. Commun. 2017, 8, 15788.

83. Achsel, T.; Brahms, H.; Kastner, B.; Bachi, A.; Wilm, M.; Luhrmann, R. A doughnut-shaped heteromer of human Sm-like proteins binds to the 3′-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. EMBO J. 1999, 18, 5789 –5802.

84. Zhou, L.; Hang, J.; Zhou, Y.; Wan, R.; Lu, G.; Yin, P.; Yan, C.; Shi, Y. Crystal structures of the Lsm complex bound to th e 3′ end sequence of U6 small nuclear RNA. Nature 2014, 506, 116–120.

85. Montemayor, E.J.; Didychuk, A.L.; Yake, A.D.; Sidhu, G.K.; Brow, D.A.; Butcher, S.E. Architecture of the U6 snRNP rev eals specific recognition of 3′-end processed U6 snRNA. Nat. Commun. 2018, 9, 1749.

86. Mund, M.; Neu, A.; Ullmann, J.; Neu, U.; Sprangers, R. Structure of the LSm657 complex: An assembly intermediate of the LSm1-7 and LSm2-8 rings. J. Mol. Biol. 2011, 414, 165–176.

87. Brahms, H.; Meheus, L.; de Brabandere, V.; Fischer, U.; Luhrmann, R. Symmetrical dimethylation of arginine residues i n spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. RNA 2001, 7, 1531–1542.

88. Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat. Rev. Mol. Cell Biol. 2017, 18, 655–670.

89. Wan, R.; Bai, R.; Shi, Y. Molecular choreography of pre-mRNA splicing by the spliceosome. Curr. Opin. Struct. Biol. 201 9, 59, 124–133.

90. Will, C.L.; Luhrmann, R. Spliceosome structure and function. Cold Spring Harb. Perspect Biol. 2011, 3.

91. Robberson, B.L.; Cote, G.J.; Berget, S.M. Exon definition may facilitate splice site selection in RNAs with multiple exon s. Mol. Cell Biol. 1990, 10, 84–94.

92. Schwer, B.; Shuman, S. Structure-function analysis of the Yhc1 subunit of yeast U1 snRNP and genetic interactions of Yhc1 with Mud2, Nam8, Mud1, Tgs1, U1 snRNA, SmD3 and Prp28. Nucleic Acids Res. 2014, 42, 4697–4711.

93. Das, R.; Yu, J.; Zhang, Z.; Gygi, M.P.; Krainer, A.R.; Gygi, S.P.; Reed, R. SR proteins function in coupling RNAP II trans cription to pre-mRNA splicing. Mol. Cell 2007, 26, 867–881.

94. Li, X.; Liu, S.; Zhang, L.; Issaian, A.; Hill, R.C.; Espinosa, S.; Shi, S.; Cui, Y.; Kappel, K.; Das, R.; et al. A unified mecha nism for intron and exon definition and back-splicing. Nature 2019, 573, 375–380.

95. Lockhart, S.R.; Rymond, B.C. Commitment of yeast pre-mRNA to the splicing pathway requires a novel U1 small nucle ar ribonucleoprotein polypeptide, Prp39p. Mol. Cell. Biol. 1994, 14, 3623–3633.

96. Briese, M.; Haberman, N.; Sibley, C.R.; Faraway, R.; Elser, A.S.; Chakrabarti, A.M.; Wang, Z.; Konig, J.; Perera, D.; Wic kramasinghe, V.O.; et al. A systems view of spliceosomal assembly and branchpoints with iCLIP. Nat. Struct. Mol. Biol. 2019, 26, 930–940.

97. Zarnack, K.; Konig, J.; Tajnik, M.; Martincorena, I.; Eustermann, S.; Stevant, I.; Reyes, A.; Anders, S.; Luscombe, N.M.; Ule, J. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu eleme nts. Cell 2013, 152, 453–466.

98. Plaschka, C.; Lin, P.C.; Charenton, C.; Nagai, K. Prespliceosome structure provides insights into spliceosome assembl y and regulation. Nature 2018, 559, 419–422.

99. Sharma, S.; Kohlstaedt, L.A.; Damianov, A.; Rio, D.C.; Black, D.L. Polypyrimidine tract binding protein controls the tran sition from exon definition to an intron defined spliceosome. Nat. Struct. Mol. Biol. 2008, 15, 183–191.

100. Charenton, C.; Wilkinson, M.E.; Nagai, K. Mechanism of 5′ splice site transfer for human spliceosome activation. Scien ce 2019, 364, 362–367.

101. Huang, Y.H.; Chung, C.S.; Kao, D.I.; Kao, T.C.; Cheng, S.C. Sad1 counteracts Brr2-mediated dissociation of U4/U6.U5 in tri-snRNP homeostasis. Mol. Cell. Biol. 2014, 34, 210–220.

102. Boesler, C.; Rigo, N.; Anokhina, M.M.; Tauchert, M.J.; Agafonov, D.E.; Kastner, B.; Urlaub, H.; Ficner, R.; Will, C.L.; Luh rmann, R. A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. Nat. Commun. 2016, 7, 11997.

103. de Almeida, R.A.; O'Keefe, R.T. The NineTeen Complex (NTC) and NTC-associated proteins as targets for spliceosom al ATPase action during pre-mRNA splicing. RNA Biol. 2015, 12, 109–114.

104. Hogg, R.; McGrail, J.C.; O'Keefe, R.T. The function of the NineTeen Complex (NTC) in regulating spliceosome conform ations and fidelity during pre-mRNA splicing. Biochem. Soc. Trans. 2010, 38, 1110–1115.

105. Nguyen, T.H.; Galej, W.P.; Bai, X.C.; Savva, C.G.; Newman, A.J.; Scheres, S.H.; Nagai, K. The architecture of the splic eosomal U4/U6.U5 tri-snRNP. Nature 2015, 523, 47–52.

106. Zahler, A.M.; Rogel, L.E.; Glover, M.L.; Yitiz, S.; Ragle, J.M.; Katzman, S. SNRP-27, the C. elegans homolog of the tri-s nRNP 27K protein, has a role in 5′ splice site positioning in the spliceosome. RNA 2018, 24, 1314–1325.

107. Townsend, C.; Leelaram, M.N.; Agafonov, D.E.; Dybkov, O.; Will, C.L.; Bertram, K.; Urlaub, H.; Kastner, B.; Stark, H.; L uhrmann, R. Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. Science

2020, 370.

108. Hang, J.; Wan, R.; Yan, C.; Shi, Y. Structural basis of pre-mRNA splicing. Science 2015, 349, 1191–1198.
109. Rauhut, R.; Fabrizio, P.; Dybkov, O.; Hartmuth, K.; Pena, V.; Chari, A.; Kumar, V.; Lee, C.T.; Urlaub, H.; Kastner, B.; et a l. Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science 2016, 353, 1399–1405.
110. Sun, C.; Rigo, N.; Fabrizio, P.; Kastner, B.; Luhrmann, R. A protein map of the yeast activated spliceosome as obtained by electron microscopy. RNA 2016, 22, 1427–1440.
111. Bao, P.; Boon, K.L.; Will, C.L.; Hartmuth, K.; Luhrmann, R. Multiple RNA-RNA tertiary interactions are dispensable for f ormation of a functional U2/U6 RNA catalytic core in the spliceosome. Nucleic Acids Res. 2018, 46, 12126–12138.
112. Galej, W.P.; Wilkinson, M.E.; Fica, S.M.; Oubridge, C.; Newman, A.J.; Nagai, K. Cryo-EM structure of the spliceosome i mmediately after branching. Nature 2016, 537, 197–201.
113. Zhang, X.; Yan, C.; Zhan, X.; Li, L.; Lei, J.; Shi, Y. Structure of the human activated spliceosome in three conformationa l states. Cell Res. 2018, 28, 307–322.
114. Yan, C.; Wan, R.; Bai, R.; Huang, G.; Shi, Y. Structure of a yeast step II catalytically activated spliceosome. Science 20 17, 355, 149–155.
115. Wan, R.; Bai, R.; Yan, C.; Lei, J.; Shi, Y. Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mechan ism of Branching. Cell 2019, 177, 339–351.e13.
116. Chung, C.S.; Tseng, C.K.; Lai, Y.H.; Wang, H.F.; Newman, A.J.; Cheng, S.C. Dynamic protein-RNA interactions in medi ating splicing catalysis. Nucleic Acids Res. 2019, 47, 899–910.
117. Gehring, N.H.; Lamprinaki, S.; Hentze, M.W.; Kulozik, A.E. The hierarchy of exon-junction complex assembly by the spl iceosome explains key features of mammalian nonsense-mediated mRNA decay. PLoS Biol. 2009, 7, e1000120.
118. Schlautmann, L.P.; Gehring, N.H. A Day in the Life of the Exon Junction Complex. Biomolecules 2020, 10, 866.
119. Djebali, S.; Davis, C.A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; et al. Landscape of transcription in human cells. Nature 2012, 489, 101–108.
120. Baralle, F.E.; Giudice, J. Alternative splicing as a regulator of development and tissue identity. Nat. Rev. Mol. Cell Biol. 2017, 18, 437–451.
121. Kornblihtt, A.R.; Schor, I.E.; Allo, M.; Dujardin, G.; Petrillo, E.; Munoz, M.J. Alternative splicing: A pivotal step between e ukaryotic transcription and translation. Nat. Rev. Mol. Cell Biol. 2013, 14, 153–165.
122. Ule, J.; Blencowe, B.J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. Mol. Cell 201 9, 76, 329–345.
123. Shen, M.; Mattox, W. Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. Nucleic Acids Res. 2012, 40, 428–437.
124. Bradley, T.; Cook, M.E.; Blanchette, M. SR proteins control a complex network of RNA-processing events. RNA 2015, 2 1, 75–92.
125. Pandit, S.; Zhou, Y.; Shiue, L.; Coutinho-Mansfield, G.; Li, H.; Qiu, J.; Huang, J.; Yeo, G.W.; Ares, M., Jr.; Fu, X.D. Gen ome-wide analysis reveals SR protein cooperation and competition in regulated splicing. Mol. Cell 2013, 50, 223–235.
126. Deininger, P. Alu elements: Know the SINEs. Genome Biol. 2011, 12, 236.
127. Payer, L.M.; Steranka, J.P.; Ardeljan, D.; Walker, J.; Fitzgerald, K.C.; Calabresi, P.A.; Cooper, T.A.; Burns, K.H. Alu inse rtion variants alter mRNA splicing. Nucleic Acids Res. 2019, 47, 421–431.
128. Loh, T.J.; Cho, S.; Moon, H.; Jang, H.N.; Williams, D.R.; Jung, D.W.; Kim, I.C.; Ghigna, C.; Biamonti, G.; Zheng, X.; et a l. hnRNP L inhibits CD44 V10 exon splicing through interacting with its upstream intron. Biochim. Biophys. Acta 2015, 1 849, 743–750.
129. Ji, X.; Park, J.W.; Bahrami-Samani, E.; Lin, L.; Duncan-Lewis, C.; Pherribo, G.; Xing, Y.; Liebhaber, S.A. alphaCP bindi ng to a cytosine-rich subset of polypyrimidine tracts drives a novel pathway of cassette exon splicing in the mammalian transcriptome. Nucleic Acids Res. 2016, 44, 2283–2297.
130. Xiao, X.; Wang, Z.; Jang, M.; Nutiu, R.; Wang, E.T.; Burge, C.B. Splice site strength-dependent activity and genetic buff ering by poly-G runs. Nat. Struct Mol. Biol. 2009, 16, 1094–1100.
131. Chen, C.D.; Kobayashi, R.; Helfman, D.M. Binding of hnRNP H to an exonic splicing silencer is involved in the regulatio n of alternative splicing of the rat beta-tropomyosin gene. Genes Dev. 1999, 13, 593–606.
132. Jamison, S.F.; Pasman, Z.; Wang, J.; Will, C.; Luhrmann, R.; Manley, J.L.; Garcia-Blanco, M.A. U1 snRNP-ASF/SF2 int eraction and 5′ splice site recognition: Characterization of required elements. Nucleic Acids Res. 1995, 23, 3260–3267.

133. Cho, S.; Hoang, A.; Sinha, R.; Zhong, X.Y.; Fu, X.D.; Krainer, A.R.; Ghosh, G. Interaction between the RNA binding do mains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. Proc. Natl. Aca d. Sci. USA 2011, 108, 8233–8238.

134. Jobbins, A.M.; Reichenbach, L.F.; Lucas, C.M.; Hudson, A.J.; Burley, G.A.; Eperon, I.C. The mechanisms of a mammali an splicing enhancer. Nucleic Acids Res. 2018, 46, 2145–2158.

135. Aubol, B.E.; Wu, G.; Keshwani, M.M.; Movassat, M.; Fattet, L.; Hertel, K.J.; Fu, X.D.; Adams, J.A. Release of SR Protei ns from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. Mol. Cell 201 6, 63, 218–228.

136. Keiper, S.; Papasaikas, P.; Will, C.L.; Valcarcel, J.; Girard, C.; Luhrmann, R. Smu1 and RED are required for activation of spliceosomal B complexes assembled on short introns. Nat. Commun. 2019, 10, 3639.

137. Papasaikas, P.; Tejedor, J.R.; Vigevani, L.; Valcarcel, J. Functional splicing network reveals extensive regulatory potenti al of the core spliceosomal machinery. Mol. Cell 2015, 57, 7–22.

138. Subramania, S.; Gagne, L.M.; Campagne, S.; Fort, V.; O'Sullivan, J.; Mocaer, K.; Feldmuller, M.; Masson, J.Y.; Allain, F. H.T.; Hussein, S.M.; et al. SAM68 interaction with U1A modulates U1 snRNP recruitment and regulates mTor pre-mRN A splicing. Nucleic Acids Res. 2019, 47, 4181–4197.

139. Adams, B.M.; Coates, M.N.; Jackson, S.R.; Jurica, M.S.; Davis, T.L. Nuclear cyclophilins affect spliceosome assembly and function in vitro. Biochem. J. 2015, 469, 223–233.

140. Bartys, N.; Kierzek, R.; Lisowiec-Wachnicka, J. The regulation properties of RNA secondary structure in alternative spli cing. Biochim. Biophys. Acta Gene Regul. Mech. 2019, 1862, 194401.

141. Warf, M.B.; Diegel, J.V.; von Hippel, P.H.; Berglund, J.A. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. Proc. Natl. Acad. Sci. USA 2009, 106, 9203–9208.

142. von Hacht, A.; Seifert, O.; Menger, M.; Schutze, T.; Arora, A.; Konthur, Z.; Neubauer, P.; Wagner, A.; Weise, C.; Kurreck, J. Identification and characterization of RNA guanine-quadruplex binding proteins. Nucleic Acids Res. 2014, 42, 6630–6644.

143. Conlon, E.G.; Lu, L.; Sharma, A.; Yamazaki, T.; Tang, T.; Shneider, N.A.; Manley, J.L. The C9ORF72 GGGGCC expansi on forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. Elife 2016, 5.

144. Huang, H.; Zhang, J.; Harvey, S.E.; Hu, X.; Cheng, C. RNA G-quadruplex secondary structure promotes alternative spli cing via the RNA-binding protein hnRNPF. Genes Dev. 2017, 31, 2296–2309.

145. Baraniak, A.P.; Lasda, E.L.; Wagner, E.J.; Garcia-Blanco, M.A. A stem structure in fibroblast growth factor receptor 2 tra nscripts mediates cell-type-specific splicing by approximating intronic control elements. Mol. Cell. Biol. 2003, 23, 9327–9337.

146. Munding, E.M.; Shiue, L.; Katzman, S.; Donohue, J.P.; Ares, M., Jr. Competition between pre-mRNAs for the splicing m achinery drives global regulation of splicing. Mol. Cell 2013, 51, 338–348.

147. Ding, F.; Elowitz, M.B. Constitutive splicing and economies of scale in gene expression. Nat. Struct. Mol. Biol. 2019, 26, 424–432.

148. Alexander, R.D.; Innocente, S.A.; Barrass, J.D.; Beggs, J.D. Splicing-dependent RNA polymerase pausing in yeast. Mo l. Cell 2010, 40, 582–593.

149. Chen, F.X.; Smith, E.R.; Shilatifard, A. Born to run: Control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 2018, 19, 464–478.

150. Naftelberg, S.; Schor, I.E.; Ast, G.; Kornblihtt, A.R. Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu. Rev. Biochem 2015, 84, 165–198.

151. Fong, N.; Kim, H.; Zhou, Y.; Ji, X.; Qiu, J.; Saldi, T.; Diener, K.; Jones, K.; Fu, X.D.; Bentley, D.L. Pre-mRNA splicing is f acilitated by an optimal RNA polymerase II elongation rate. Genes Dev. 2014, 28, 2663–2676.

152. Baluapuri, A.; Hofstetter, J.; Dudvarski Stankovic, N.; Endres, T.; Bhandare, P.; Vos, S.M.; Adhikari, B.; Schwarz, J.D.; Narain, A.; Vogt, M.; et al. MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. Mol. Cell 2019, 74, 674–687.e11.

153. Lin, C.Y.; Loven, J.; Rahl, P.B.; Paranal, R.M.; Burge, C.B.; Bradner, J.E.; Lee, T.I.; Young, R.A. Transcriptional amplific ation in tumor cells with elevated c-Myc. Cell 2012, 151, 56–67.

154. Luco, R.F.; Allo, M.; Schor, I.E.; Kornblihtt, A.R.; Misteli, T. Epigenetics in alternative pre-mRNA splicing. Cell 2011, 144, 16–26.

155. Narayanan, S.P.; Singh, S.; Shukla, S. A saga of cancer epigenetics: Linking epigenetics to alternative splicing. Bioche m. J. 2017, 474, 885–896.

156. Schwartz, S.; Meshorer, E.; Ast, G. Chromatin organization marks exon-intron structure. Nat. Struct. Mol. Biol. 2009, 16, 990–995.

157. Spies, N.; Nielsen, C.B.; Padgett, R.A.; Burge, C.B. Biased chromatin signatures around polyadenylation sites and exons. Mol. Cell 2009, 36, 245–254.

158. Sims, R.J., 3rd; Millhouse, S.; Chen, C.F.; Lewis, B.A.; Erdjument-Bromage, H.; Tempst, P.; Manley, J.L.; Reinberg, D. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol. Cell 2007, 28, 665–676.

159. Luco, R.F.; Pan, Q.; Tominaga, K.; Blencowe, B.J.; Pereira-Smith, O.M.; Misteli, T. Regulation of alternative splicing by histone modifications. Science 2010, 327, 996–1000.

160. Ramanathan, A.; Robb, G.B.; Chan, S.H. mRNA capping: Biological functions and applications. Nucleic Acids Res. 2016, 44, 7511–7526.

161. Kumar, A.; Clerici, M.; Muckenfuss, L.M.; Passmore, L.A.; Jinek, M. Mechanistic insights into mRNA 3′-end processing. Curr. Opin. Struct. Biol. 2019, 59, 143–150.

162. Chen, Y.I.; Moore, R.E.; Ge, H.Y.; Young, M.K.; Lee, T.D.; Stevens, S.W. Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. Nucleic Acids Res. 2007, 35, 3928–3944.

163. Berg, M.G.; Singh, L.N.; Younis, I.; Liu, Q.; Pinto, A.M.; Kaida, D.; Zhang, Z.; Cho, S.; Sherrill-Mix, S.; Wan, L.; et al. U1 snRNP determines mRNA length and regulates isoform expression. Cell 2012, 150, 53–64.

164. Kaida, D.; Berg, M.G.; Younis, I.; Kasim, M.; Singh, L.N.; Wan, L.; Dreyfuss, G. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 2010, 468, 664–668.

165. He, P.C.; He, C. m(6) A RNA methylation: From mechanisms to therapeutic potential. EMBO J. 2021, 40, e105977.

166. Wang, T.; Kong, S.; Tao, M.; Ju, S. The potential role of RNA N6-methyladenosine in Cancer progression. Mol. Cancer 2020, 19, 88.

167. Liu, N.; Dai, Q.; Zheng, G.; He, C.; Parisien, M.; Pan, T. N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature 2015, 518, 560–564.

168. Liu, N.; Zhou, K.I.; Parisien, M.; Dai, Q.; Diatchenko, L.; Pan, T. N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. Nucleic Acids Res. 2017, 45, 6051–6063.

169. Zhao, X.; Yang, Y.; Sun, B.F.; Shi, Y.; Yang, X.; Xiao, W.; Hao, Y.J.; Ping, X.L.; Chen, Y.S.; Wang, W.J.; et al. FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. Cell Res. 2014, 24, 1403–1419.

170. Zheng, G.; Dahl, J.A.; Niu, Y.; Fedorcsak, P.; Huang, C.M.; Li, C.J.; Vagbo, C.B.; Shi, Y.; Wang, W.L.; Song, S.H.; et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell 2013, 49, 18–29.

171. Xiao, W.; Adhikari, S.; Dahal, U.; Chen, Y.S.; Hao, Y.J.; Sun, B.F.; Sun, H.Y.; Li, A.; Ping, X.L.; Lai, W.Y.; et al. Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol. Cell 2016, 61, 507–519.

172. Samuel, C.E. RNA Editing. In Reference Module in Biomedical Sciences; Caplan, M., Ed.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 1–6.

173. Phelps, K.J.; Tran, K.; Eifler, T.; Erickson, A.I.; Fisher, A.J.; Beal, P.A. Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. Nucleic Acids Res. 2015, 43, 1123–1132.

174. Severi, F.; Conticello, S.G. Flow-cytometric visualization of C>U mRNA editing reveals the dynamics of the process in live cells. RNA Biol. 2015, 12, 389–397.

175. Levanon, E.Y.; Eisenberg, E.; Yelin, R.; Nemzer, S.; Hallegger, M.; Shemesh, R.; Fligelman, Z.Y.; Shoshan, A.; Pollock, S.R.; Sztybel, D.; et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. 2004, 22, 1001–1005.

176. Rueter, S.M.; Dawson, T.R.; Emeson, R.B. Regulation of alternative splicing by RNA editing. Nature 1999, 399, 75–80.

177. Hsiao, Y.E.; Bahn, J.H.; Yang, Y.; Lin, X.; Tran, S.; Yang, E.W.; Quinones-Valdez, G.; Xiao, X. RNA editing in nascent RNA affects pre-mRNA splicing. Genome Res. 2018, 28, 812–823.

178. Huang, H.; Kapeli, K.; Jin, W.; Wong, Y.P.; Arumugam, T.V.; Koh, J.H.; Srimasorn, S.; Mallilankaraman, K.; Chua, J.J.E.; Yeo, G.W.; et al. Tissue-selective restriction of RNA editing of CaV1.3 by splicing factor SRSF9. Nucleic Acids Res. 2018, 46, 7323–7338.

179. Tang, S.J.; Shen, H.; An, O.; Hong, H.; Li, J.; Song, Y.; Han, J.; Tay, D.J.T.; Ng, V.H.E.; Bellido Molias, F.; et al. Cis- and trans-regulations of pre-mRNA splicing by RNA editing enzymes influence cancer development. Nat. Commun. 2020, 1

1, 799.

180. Agranat, L.; Sperling, J.; Sperling, R. A novel tissue-specific alternatively spliced form of the A-to-I RNA editing enzyme ADAR2. RNA Biol. 2010, 7, 253–262.
181. Agranat-Tamir, L.; Shomron, N.; Sperling, J.; Sperling, R. Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome. Nucleic Acids Res. 2014, 42, 4640–4651.
182. Mahlab-Aviv, S.; Boulos, A.; Peretz, A.R.; Eliyahu, T.; Carmel, L.; Sperling, R.; Linial, M. Small RNA sequences derived from pre-microRNAs in the supraspliceosome. Nucleic Acids Res. 2018, 46, 11014–11029.
183. Yin, Y.; Lu, J.Y.; Zhang, X.; Shao, W.; Xu, Y.; Li, P.; Hong, Y.; Cui, L.; Shan, G.; Tian, B.; et al. U1 snRNP regulates chromatin retention of noncoding RNAs. Nature 2020, 580, 147–150.
184. Azam, S.; Hou, S.; Zhu, B.; Wang, W.; Hao, T.; Bu, X.; Khan, M.; Lei, H. Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. RNA Biol. 2019, 16, 1001–1009.
185. Wang, K.; Yin, C.; Du, X.; Chen, S.; Wang, J.; Zhang, L.; Wang, L.; Yu, Y.; Chi, B.; Shi, M.; et al. A U2-snRNP-independent role of SF3b in promoting mRNA export. Proc. Natl. Acad. Sci. USA 2019, 116, 7837–7846.
186. Kim, J.S.; He, X.; Liu, J.; Duan, Z.; Kim, T.; Gerard, J.; Kim, B.; Pillai, M.M.; Lane, W.S.; Noble, W.S.; et al. Systematic proteomics of endogenous human cohesin reveals an interaction with diverse splicing factors and RNA-binding proteins required for mitotic progression. J. Biol. Chem. 2019, 294, 8760–8772.
