Cell. Mol. Life Sci. 63 (2006) 796–819
1420-682X/06/080796-24
DOI 10.1007/s00018-005-5391-x


Cellular and Molecular Life Sciences

# Review

# Pre-messenger RNA processing and its regulation: a genomic perspective

**M. Soller**

Department of Biology and Volen Center for Complex Systems, MS 008, Brandeis University, Waltham, Massachusetts 02454 (USA), Fax: +1 781 736 3107, e-mail: msoller@brandeis.edu

Received 22 August 2005; received after revision 14 November 2005; accepted 5 December 2005
Online First 7 February 2006

**Abstract.** The sequence of entire genomes from higher eukaryotes revealed that an average gene is very large due to the interruption of the coding sequence with large non-coding introns. Introns are co-transcriptionally removed by splicing with great accuracy and fidelity, although contrary to our expectations, currently known signals required for pre-messenger RNA (mRNA) processing are very degenerate and redundant. Furthermore, the vast majority of genes are alternatively processed. A large number of proteins are therefore involved in generating specificity in pre-mRNA processing that requires a dedicated mechanisms to operate at genomic dimensions. In this review I will summarize recent progress in understanding how established principles of pre-mRNA processing extend to genomic dimensions and discuss emerging concepts in coupling of pre-mRNA processing with other nuclear events and nuclear organization.

**Keywords.** Alternative splicing, alternative polyadenylation, editing, capping, hnRNP, coupling of RNA processing to transcription, genomics.

## Introduction

Since the genomes of several higher eukaryotes have been sequenced, it has become clear that most genes are very large due to the interruption of the coding sequence with large portions of non-coding sequences (introns). Furthermore, the size of introns correlates with genome size, and both increase from lower to higher eukaryotes [1]. In addition, an increased size of introns has also been associated with a smaller population size of a species. An average gene in humans is transcribed under the control of transcription factors from about 30 kb of chromosomal DNA into a pre-messengerRNA (pre-mRNA). The pre-mRNA is then processed by splicing out introns to an mRNA of only about 3 kb. The mRNA is then transported to the cytoplasm, where it serves as template for the synthesis of a protein [2]. Despite the generally very short and degenerate nature of splicing signals, processing of pre-mRNAs occurs with amazing fidelity and accuracy independent of intron size. Central to the processing of pre-mRNAs are a large number of combinatorial interactions among regulatory factors. Of particular interest to our understanding of pre-mRNA processing is the elucidation of combinatorial codes to predict processing choices in situations of alternative pre-mRNA processing or to evaluate the impact of genetic polymorphisms.

The annotation of eukaryotic genomes by experimental determination from complementary DNA (cDNA) libraries and homology to other genomes has much improved since publication of the first draft sequences. One of the striking results of this effort has been that alternative pre-mRNA processing is a major theme in the expression of genes. In humans, for example, about 60–80% of genes are alternatively spliced in at least one exon, and at least 30% of genes have alternative polyA sites [3–5]. Since a mere number around 25,000 genes in humans has been below expectations with regard to organismal complexity compared with other eukaryotic genomes such as *Drosophila* (14,000 genes), *Caenorhabditis elegans* (19,000 genes) or *Arabidopsis* (~25,000), al-

ternative RNA processing has gained attention as a mechanism to generate molecular diversity and contribute to morphological complexity (for references on gene numbers see [1]). Notably, the number of alternatively spliced genes increases from 10% in *C. elegans* and plants to 25–40% in *Drosophila* and 60–80% in humans, and thus correlates with organismal complexity [3–9]. Previously, gene duplication, particularly the duplication of transcription factor genes in association with their regulatory potential to act in a combinatorial manner, has been proposed to provide the major contribution to an increase in organismal complexity [10]. Gene duplication and alternative RNA processing, however, show an inverse correlation in evolutionarily conserved gene families, suggesting that both mechanisms are evolutionarily interchangeable strategies to genic proliferation [11]. A link between the elaboration of transcriptional control and alternative RNA processing is also suggested by the fact that transcription and pre-mRNA processing are highly coupled processes (see below, [12, 13]). Consequently, the type of promoter and its associated transcription factors and co-regulators have been shown to determine alternative pre-mRNA splicing choices [14, 15].

Among the different pre-mRNA processing possibilities, alternative splicing is the most prevalent mechanism to generate proteomic diversity while alternative polyA site choices can influence proteomic diversity through the use of different terminal exons, and regulatory potential through 3′UTR sequences, respectively (Fig. 1). RNA editing is found in a minority of transcripts. To name only a few astounding examples of extensively alternatively spliced genes, the *Down syndrome cell adhesion molecule* gene (*Dscam)* from *Drosophila* or the *Neurexin* and *CD44* genes in humans can produce as many as about 38,000, 3000 and 1000 different splice forms, respectively (Fig. 2) [16–18]. An interesting feature of the *Dscam* gene, a cell adhesion molecule involved in axon guidance, are three regions in the extracellular domain where a single exon from a number of possible choices is spliced in a mutual exclusive fashion, thereby generating a highly variable region (Fig. 2c). Cloning of *Dscam* cDNAs from *Drosophila* revealed that most isoforms are made, while for *CD44* alternative splicing in a cell line with metastatic growth, 95% of alternative splice forms belong to the 15 most frequently found isoforms [16, 17]. Functional relevance of *Dscam* diversity is further indicated by its expression in the immune system and by the homophilic interaction of extracellular domains illustrating the potential of alternative splicing to contribute to the complex wiring of the nervous system [19, 20]. Another example of an extensively regulated gene is the *lola* transcription factor from *Drosophila* (Fig. 2d). Here, choice of a different terminal exon regulates the DNA binding properties of this transcription factor involved in wiring of the nervous system [21]. The importance of alternative splicing in the nervous system is further pointed out by the much higher number of affected genes in this tissue compared with other human tissues [22, 23]. Alternative splicing is not restricted to functional groups of proteins, although some cellular functions such as apoptosis seem to be highly regulated by alternative splicing [22, 24, 25].

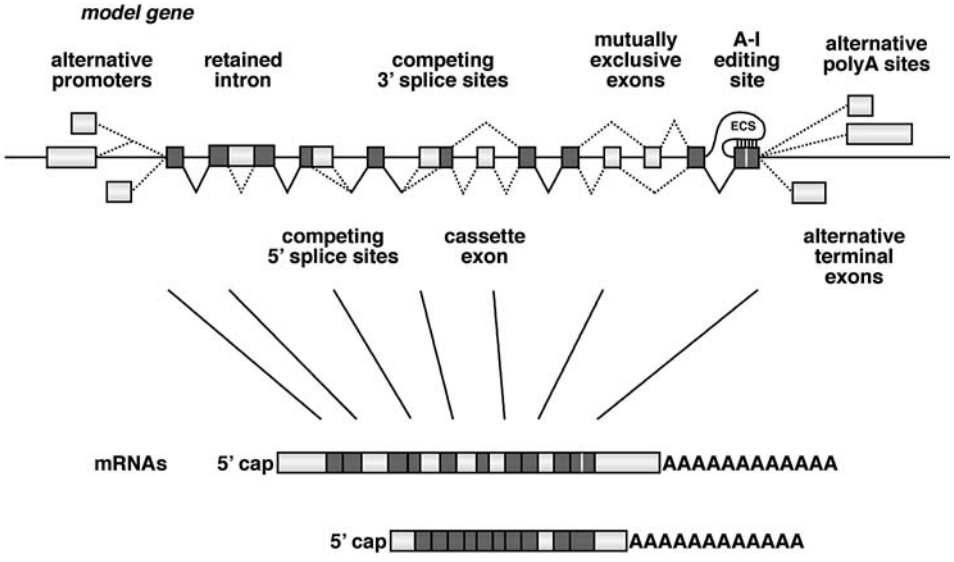


**Figure 1.** Alternative pre-mRNA processing types illustrated with an artificial model gene (top) transcribed and processed into possible mRNA isoforms (bottom). Exons are shown as boxes and introns or intergenic regions as lines. Gene regions with alternative pre-mRNA processing choices are illustrated in white and connected with dashed lines, while constitutive parts are depicted in black and connected with solid lines. For adenosine to inosine editing (A to I) an editing site complementary sequence (ECS) located in an intron pairs with the edited site in the exon.

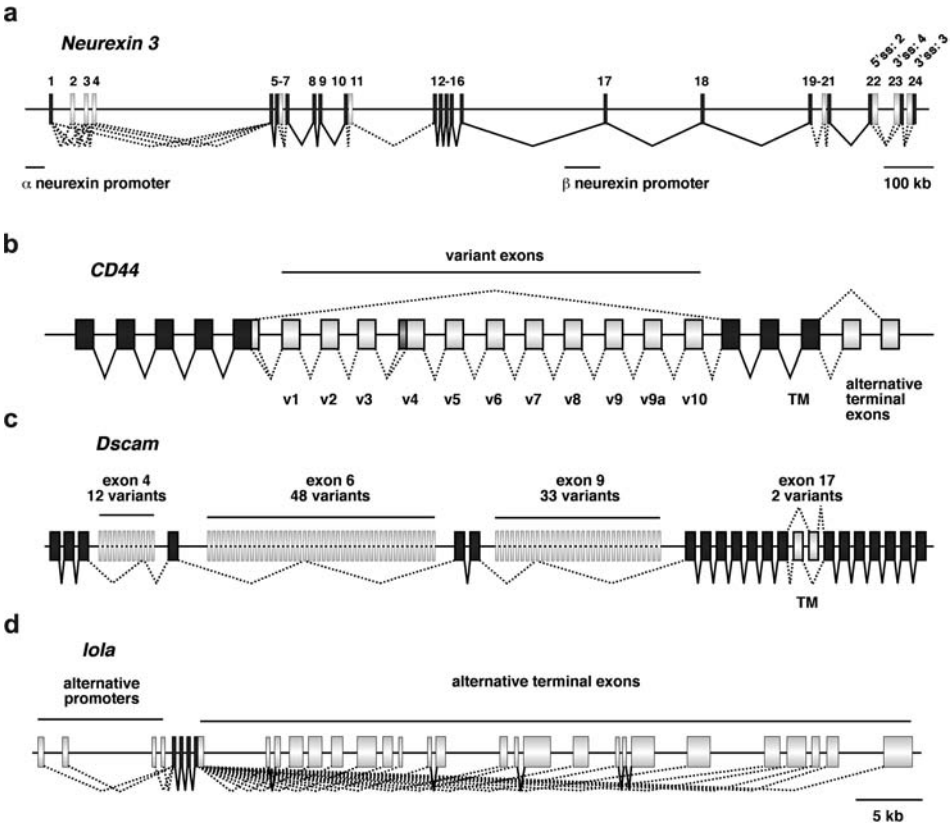


**Figure 2.** Examples of genes where alternative splicing generates extraordinary molecular diversity. Exons are shown as boxes and introns or intergenic regions as lines. Gene regions with alternative pre-mRNA processing choices are illustrated in white and connected with dashed lines while constitutive parts are depicted in black and connected with solid lines. (*a*) The human *Neurexin 3* gene coding for a neuronal cell adhesion molecule spans over 1600 kb of genomic DNA and encodes for 24 exons (modified form [18]). In addition to two transcriptional start sites, the 5′ and 3′ regions of the gene contain variable exons (exons 2–4) or splice sites (exons 22–24). The *neurexin 3* gene's largest constitutively spliced intron (292 kb) is between exons 16 and 17, and the largest alternatively spliced intron (347 kb) is between exon 1 and 5. (*b*) The human CD44 gene contains an extended variable region in the extracellular domain and two alternative terminal exons located intracellularly (modified from [313]). Although many of the possible isoforms are made, different cell types preferentially express a specific combination of variable exons [313, 314]. TM, transmembrane domain. (*c*) The *Drosophila Down syndrome cell adhesion molecule* (*Dscam*) gene contains three variable extracellular regions where one exon is included from a number of possible choices, and a region with two mutually exclusive exons in the transmembrane domain (TM, modified from [16]). (*d*) The *Drosophila longitudinal lacking* (*lola*) transcription factor gene spans over 60 kb, contains four alternative promoters and 19 different terminal exons encoding zinc finger DNA binding domains (modified from [222]). The constant region (black) encodes a BTB dimerization domain.

Much has been learned about principles operating in constitutive pre-mRNA processing. Extending our knowledge to genomic dimensions, however, has been difficult due to large intron sizes and degenerate regulatory sequences. Deducing regulatory mechanisms of alternative splicing and polyadenylation has therefore mainly relied on experimental determination. In addition, since genomic dimensions mostly do not fit our test tubes to analyze alternative pre-mRNA processing, many simplifications in assay systems are made as a compromise to study a particular problem of regulated RNA processing. The availability of the sequence of entire genomes for bioinformatic approaches together with the powerful genetic systems of yeast, *Drosophila* and mice as well as RNAi (RNA interference) applications to generate hypomorphic conditions now provide the tools to extend our understanding of pre-mRNA processing and its regulation to genomic levels. This is of particular relevance as 15% of inherited diseases in humans result from mutations in the proximity of splice junctions, but considering all aspects of pre-mRNA processing, this number is underestimated [26–29]. In addition, splicing is also modulated by age [30].

In recent years, several strategies have proven successful to correct defects in pre-mRNA processing in simplified assay systems [31]. Currently, the most promising examples for therapeutical approaches are those where reconstitution from a null condition to residual gene activity provides large improvements to life quality; as, for example, in some cases of $\beta$-thalassemias, where skipping of a stop codon-containing exon results in a functional $\beta$-globin protein (for review see [31]). In many instances, how-

ever, the delicate balance of different isoforms is important, as shown in the regulated inclusion of exon 10 in the *tau* gene, whose mis-regulation has been linked to frontotemporal dementia and Parkinsonism (FTDP) [32]. In stress induced neural hypersensitivity, a short temporal switch in acetylcholinesterase alternative splicing results in changes in dendritic physiology lasting weeks [33]. Thus, the development and test of strategies for therapeutic use requires model systems that operate at genomic and organismal levels, particularly with regard to long-term effects on brain function and the regulation of behavioral output.

In the following section, I will give a brief introduction to the individual steps of pre-mRNA processing required for mRNA maturation into an export competent ribonucleoprotein particle (RNP) and the mechanisms that operate to recognize the highly degenerate processing signals that are dispersed at genomic levels by large introns. As alternative pre-mRNA processing is a major theme in gene expression, I am going to summarize recent advances in our current understanding of the regulatory mechanism and how they apply to genomic dimensions. In recent years, mechanisms that deal with splicing of large introns have come to our attention, and I will discuss those in light of additional complications for the processing of pre-mRNAs that arise at genomic levels. Although individual pre-mRNA processing reactions have been viewed as separate reactions for many years, coupling to transcription has become a major focus of recent research, and concepts have emerged regarding how organization into processing units can both add new regulatory potential and reduce sequence complexity. Finally, I will point out some of the future directions that will advance our understanding of the mechanistic aspects of pre-mRNA processing and how this in turn can be applied to improve human health.

## Pre-mRNA processing reactions: capping, editing, splicing, 3′end processing

The first step in the processing of a pre-mRNA is capping. After synthesizing 20–30 nucleotides, RNA Polymerase II (RNA Pol II) pauses, and in a three-step reaction a $N^7$-methyl GMP is added in an unusual 5′-5′ triphosphate linkage [34, 35]. The cap is then bound by the nuclear cap binding complex consisting of a 20- and 80-kDa protein [36]. Nucleotides adjacent to the cap structure are to various degrees also ribose methylated [37]. These 2′-O-methylations are carried out by enzymatic activities different from methylation of internal adenosine residues, which form $N^6$-methyladenosine before splicing, but no unambiguous nuclear function has yet been attributed to this type of modification [38]. The cap structure is important for protecting the RNA from 5′-3′ exonucleases, stimulating splicing of the first intron and 3′end processing, and for enhancing translation [36, 39–42].

Until recently RNA editing was known to occur only in a handful of transcripts and was largely underestimated owing to the difficulty in detecting editing sites, as only a single nucleotide is changed [43, 44]. Editing is either achieved by deamination or, as in mitochondria of trypanosomes, a lower eukaryote by insertion of uridines into the pre-mRNA with the help of a guide RNA [45]. Cytidine to uridine deamination (C to U) occurs on single-stranded RNA as substrate, and involves the editing activity APOBEC-1 (apoB editing catalytic subunit 1) and at least one auxiliary factor from the alternatively spliced ACF gene (APOBEC-1 complementation factor, ACF65 and ACF64) that recognizes a specific RNA sequence termed mooring sequence [46]. Only two genes are known, apolipoprotein B and neurofibromin, that are C to U edited by APOBEC-1, and mice devoid of APOBEC-1 are viable. In humans, a family of APOBEC-1 related proteins are present, but they have primarily been associated with DNA editing in immune cells important for class switch recombination and hypermutation of immunoglobulins, and of retroviral genomes [47]. Adenosine to inosine deamination (A to I) is catalyzed by ADAR's (adenosine deaminase acting on RNA) and seems to be particularly prominent in the nervous system based on phenotypic analysis [48]. The editing site usually consists of an imperfect duplex RNA formed by base pairing between the editing site and an editing site complementary sequence (ECS, Fig. 1) that can be thousands of nucleotides away [49]. Since auto-regulatory editing of a single nucleotide in ADAR2 transcripts changes its splicing pattern, and since the ECS is mostly present in introns, editing is thought to occur before splicing [43, 50]. A to I editing has been largely underestimated as recent bioinformatic approaches together with experimental validation showed A to I editing in *Alu* elements, widespread repetitive elements that comprise at least 10% of primate genomes, but are absent in others [51–53]. The repetitive *Alu* elements are about 300 nt long and form extensive double-stranded regions likely with a widespread impact on further pre-mRNA processing. In addition, a large number of *Alu* elements have also become alternatively spliced exons [54].

The removal of introns by splicing occurs in two trans-esterification steps resulting in a spliced RNA and a lariat of the intron. Splicing requires four loosely defined sequence elements, which are the 5′ splice site (consensus in mammals: AG/GURAGU), the branchpoint (YNYURAC), a variable stretch of pyrimidines termed polypyrimidine tract and the 3′ splice site (YAG/N; / denotes the exon/intron boundary). During the first step, the 2′-hydroxyl group of the branchpoint adenosine, which is up to 100 nt upstream of the 3′ splice site, attacks the phosphodiester linkage of

the 5′ splice site, leaving a 3′-hydroxyl group at the end of the first exon. In the second step, this 3′-hydroxyl group attacks the phosphodiester linkage at the 3′ splice site resulting in joining of the two exons and release of the lariat intron. Splicing involves the stepwise assembly of five (U1, U2, U4, U5 and U6) small ribonucleoprotein particles (snRNPs) and a large number of proteins onto the pre-mRNA to form a large complex called the spliceosome [55]. The first step in spliceosome assembly is the recognition of the 5′ splice site by U1 snRNP, binding of SF1/mBBP and of U2 auxiliary factor (U2AF35 and U2AF65) to the branchpoint and to the polypyrimidine tract and the AG of the 3′ splice site, respectively. Subsequently, U2 snRNP binds to the branchpoint and then the U4/U6*U5 tri-snRNP joins to form the spliceosome that is catalytically active upon rearrangement and destabilization of U1 and U4 snRNPs [56, 57].

Rather surprisingly, a second type of intron exists, which is spliced by a second type of spliceosome present in higher eukaryotes (reviewed in [58]). Here, U11, U12, U4atac and U6atac snRNPs replace U1, U2, U4 and U6 snRNPs, respectively. U5 is shared between both spliceosomes, but several minor variant U5 snRNAs are present in *Drosophila* that potentially could be involved in U12 intron splicing [59]. Although historically termed AT-AC introns, deduced from the first known introns of this second type in relation to the change of the consensus of GT-AG of regular introns, systematic analysis of more introns did not reveal such a strict consensus for splice sites of this second type of spliceosome. Therefore, this type of intron is now named U12 intron and is determined by the more tightly constrained consensus sequence at the 5′ splice site and the lack of a polypyrimidine tract in front of the 3′ splice site. U12 introns in humans count for less than 1% of all introns.

At the 3′ end, all RNA Pol II transcripts are cleaved, and with the exception of histone RNAs and snRNAs, a polyA tail of about 200 adenosines is added. Consensus sequence elements for processing of polyadenylated mRNAs in animals consist of an AAUAAA hexamer sequence, a CA dinucleotide after which cleavage occurs and a U- or GU-rich downstream sequence element (DSE). For some polyA sites, a stimulatory upstream sequence element (USE) has also been defined that seems to be important for processing of weak polyA sites [60, 61]. Recognition of a 3′end processing site occurs in a highly cooperative fashion by the two multiprotein complexes, cleavage and polyadenylation specificity factor (CPSF), which binds to the AAUAAA hexamer, and cleavage stimulatory factor (CstF), which binds to the DSE. To direct cleavage, two additional multiprotein factors are recruited, cleavage factor I and II (CFI and II) in addition with polyA polymerase (PAP). Symplekin also stimulates 3′end processing and is thought to connect the multiprotein processing subunits [62]. After cleavage, PAP together with CPSF and stimulated by polyA binding protein adds the polyA tail [63, 64]. Histone transcripts are processed at the 3′end in a single cleavage reaction after a stem loop structure by a multiprotein complex containing U7 snRNP [65]. A long-standing mystery in 3′end processing field has been the nature of the endonuclease that directs cleavage. Recently, the CPSF-73 subunit has been implicated to be the long-sought endonuclease due to homology to known endonucleases and by its site-specific cross-linking to the cleavage site in both polyA and histone transcripts [66, 67].

## Combinatorial interactions define exons, introns and polyA sites

Pre-mRNA processing occurs with high fidelity and accuracy despite the high degeneracy of pre-mRNA processing signals. Therefore, models have been proposed that define processing sites on a basis of combinatorial interactions among regulatory factors and pre-mRNA processing machinery [68–71]. Key to the formulation of the ‘exon definition model’ (Fig. 3a) was the finding by Berget and co-workers that a downstream 5′ splice site can stimulate a weak upstream 3′ splice site [68]. Further support for this model comes from splicing stimulatory sequences termed exonic splicing enhancers (ESEs), which are preferentially bound by serine-arginine-rich proteins (SR proteins), that contain an RNA recognition motif (RRM) in addition to a arginine-serine-rich domain (RS domain, [72, 73]). ESE-bound SR proteins mediate cross-exon interactions and were also shown to contact the branchpoint [72, 74]. Experimental and bioinformatics approaches have revealed a number of consensus sequences for ESEs that were also shown to be bound by SR proteins and are most prominently found in constitutive exons [73, 75–78]. Regulatory elements that antagonize exon definition are termed intronic or exonic splicing silencers (ISSs or ESSs) and are bound by negative regulators of splicing such as hnRNP proteins (heterogeneous nuclear ribonucleoproteins) [79–81].

Since 5′ and 3′ splice sites come into proximity during spliceosome assembly through a number of RNA-protein and protein-protein interactions, the opposite ‘intron definition model’ is also proposed (Fig. 3b, [68]). This model most likely applies for short introns. Interestingly, short introns are also centered around an ‘ideal’ length of about 60 nt in *Drosophila*, suggesting that this is the space occupied by the spliceosome [82]. In the case of bigger introns, this model has been proposed when splice sites are brought into proximity through extensive RNA folding or through multimerization of hnRNP proteins (see below).

To define the first and last exon of a transcript, the exon definition model has been extended. The cap structure has been shown to interact with the first 5′ splice site and stim-

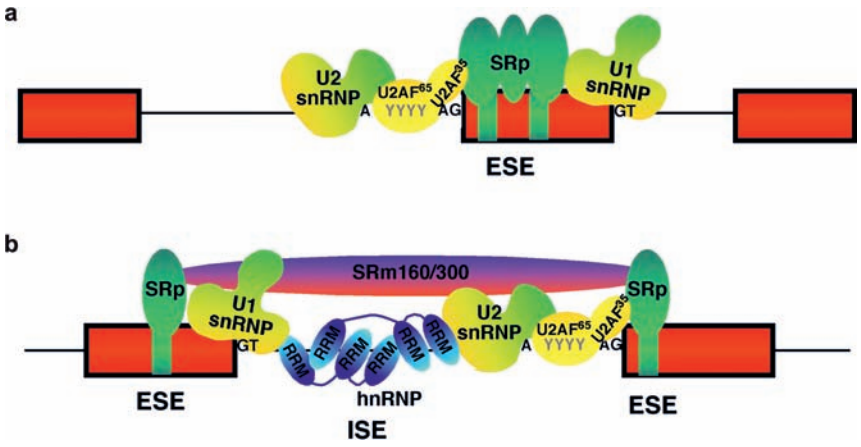


**Figure 3.** Splice site localization in pre-mRNA via exon definition (*a*) or intron definition (*b*) models. (*a*) Splice site recognition in the exon definition model is enhanced across the exon by exon bound proteins, e.g. SR proteins (SRp, green) bound to splicing enhancers (ESE), that interact with spliceosomal components (yellow) to stabilize their interaction with the pre-mRNA. Exons are shown as red boxes, introns as solid lines and ESEs as green boxes. (*b*) In the intron definition model, cross intron interactions, e.g. mediated by hnRNP proteins (blue) bound to intronic splicing enhancers (ISE), or mediated by SRm160 or SRm300 proteins (purple), that interact with U1 snRNP, U2snRNP (yellow) and SR proteins (green), stabilize spliceosomal components.

ulates splicing [36, 40]. At the other end of a transcript, the 3′ splice site from the last exon is important for cross-talk between splicing and 3′end formation to define the last exon [83–87]. Here, interaction of the 3′ splice site through U2AF 65 with polyA polymerase has been found to stimulate 3′end processing [88, 89]. In addition, components involved in the recognition of 5′ splice sites, such as U1A have also been found to stimulate 3′end processing by binding to a USE and interaction with the CPSF 160-kDa subunit [90]. The role of U1 snRNP, or its components U1A and U1 70K alone in defining 3′end processing sites, however, is ambiguous as they can also inhibit 3′end processing from either USEs or DSEs [91, 92].

## Regulators of alternative pre-mRNA processing

The most prominent regulators of alternative splicing comprise a large group of RNA binding proteins. RNA binding proteins have a modular composition consisting of one or more RNA binding domains and mostly at least one additional auxiliary domain [93–96]. The RRM is the most prevalent RNA binding domain and one of the most frequent protein domains in eukaryotes [93, 97, 98]. The ~80 amino acid RRM contains two highly conserved short motifs, RNP1 and RNP2, and folds into a characteristic structure of a four-stranded antiparallel $\beta$-sheet and two $\alpha$-helices that binds 4–7 nt [98, 99]. In addition to the RRM, the KH, RGG and double-stranded RNA binding domain are other common RNA binding domains, but a number of other, less frequently found RNA binding domains are also known [93, 100, 101]. Binding specificities for many RNA binding proteins have been studied in detail with short sequences *in vitro* [93, 96, 102–105]. Using such short binding motifs in bioinformatics approaches has resulted in identification of regulated genes in some instances (e.g. [106, 107]). In most cases, however, predicting sequence specific binding in a larger context based on short binding motifs is difficult due to sequence redundancy and therefore relies mainly on experimental determination of binding sites (e.g. [105, 108–111]). A key role in achieving binding specificity *in vivo* has been attributed to either multiple RNA binding domains in a single protein or to protein-protein interactions among RNA binding proteins, or to both. Although many RNA binding proteins engage in homo- and heterophilic interactions and are present in large RNPs *in vivo* (e.g. [93, 112]), elucidating combinatorial codes that explain binding specificity in a complex cellular environment is challenging due to difficulties in recapitulating the assembly of macromolecular complexes. In the case of the gene-specific RNA binding protein ELAV, a defined multimeric complex assembles on its target sequence from the *Drosophila ewg* gene *in vitro* and confines sequence specificity by binding to multiple short and spaced binding motifs. This array of evolutionarily conserved binding motifs is also required for splicing regulation in *Drosophila* neurons [113]. Multiple ISI and ESI elements are also important for hnRNP A1-mediated regulation of intron 3 alternative splicing from human immunodeficiency virus (HIV) tat transcripts [114–116]. Current models of this scenario propose that multimerization of hnRNP A1 in the context of RNA secondary structure is key to hnRNP A1-mediated repression [115, 116]. Regulators of alternative pre-mRNA processing also often form alternative complexes with components of processing machinery (eg. [117–121]).

A further class of proteins involved in RNA processing are RNA helicases, mostly containing a DEAD or DEXH motif [122]. These proteins have also been termed 'unwindases' as they can alter RNA structure by hydrolysing ATP, and some of them are essential for conformational changes in the spliceosome during splicing [56]. The action of RNA helicases is not restricted to remodeling double-stranded regions in RNA alone, and they can interupt RNA-protein interactions independent of duplex unwinding [123]. Distinct RNA helicases are involved in a number of constitutive cellular processes such as splicing, processing of ribosomal RNA or snoRNAs (small nucleolar RNAs), but have also been shown regulatory roles in development and have been implicated in alternative splicing [124, 125]. How RNA helicases recognize specific target RNAs, however, is not clear.

An additional class of proteins involved in pre-mRNA processing do not bind RNA directly, but associate with RNA binding proteins or subunits of processing machineries as scaffolding factors. Examples of such proteins are PRP31p, which is essential for the association of pre-spliceosomes with U4/U6*U5 snRNPs, SRm300, an SR protein that lacks an RNA binding moiety and stimulates splicing or the 3′end processing factor symplekin that connects the multiprotein processing subunits involved in cleavage and polyadenylation [62, 126–128]. A prominent coupling function has also been attributed to the C-terminal domain of RNA Pol II (CTD) as it connects pre-mRNA processing with transcription (see below, [12, 129]).

## Combinatorial interactions regulate alternative pre-mRNA processing

Central to the regulation of alternative splicing is the recognition of splice sites by the splicing machinery. A common theme regarding the recognition of splice sites includes 'weak' splice sites or splicing enhancers that are less well recognized since they diverge from the consensus sequence. When a constitutive splice site is put in a competitive context with other splice sites, often little is needed to switch a particular splice site from a constitutive to an alternative one. Numerous examples of this scenario have been described leading to alternative 5′ or 3′ splice site usage or to skipping of an exon (Fig. 1) [130, 131] or acquisition of new exons from repetitive *Alu* elements [54, 132]. Of further importance in the regulation of alternative splicing is the interplay of positive (ESEs and ISEs) and negative (ESSs and ISSs) regulatory elements, and differential concentrations of antagonistic factors have been found to be important for alternative splice site choice [133]. Based on results from a number of *in vitro* systems, SR proteins have mostly been associated with binding to ESEs and stimulate splicing, while hnRNP proteins were mostly found to be negative regulators by binding to ISSs [115, 134–137]. For hnRNP I (or polypyrimidine tract binding protein, PTB), a prominent role has been found to bind to ESSs and antagonize exon definition [138, 139] and 3′end processing [140]. As indicated by the mostly non-overlapping binding sites of antagonistic factors, this view of a set of positive and negative splicing regulators seems to be too simplistic. Consequently, SR proteins have also been shown to act as negative splicing regulators [141]. In addition, a global analysis of alternative splicing using mutants for either SR proteins or hnRNP proteins also supports both positive and negative functions for either of the two classes [111].

Since sequences of whole genomes have become available, a number of bioinformatic approaches in combination with experimental validation have addressed the distribution of regulatory sequences with respect to alternative splicing (e.g. [106, 107, 142, 143]). In addition, sequence comparison of multiple closely related species allows for the identification of conserved elements in alternatively processed genes (e.g. [113, 144, 145]). Recently, as more features of alternatively spliced exons become evident, machine learning techniques are being applied for alternative splicing predictions [146, 147] and promise progress in defining combinatorial codes that operate in alternative pre-mRNA processing.

A common theme in regulating alternative splicing involves the presence of tissue-specific factors in one cell type, but not in another, and numerous examples have been described for antagonistic situations where these factors either enhance or repress use of a particular splice site [71, 130, 148–150]. CELF (CUG-BP and ETR3-like factors), Fox and raver1 family proteins, for example, relieve the repressive activity of PTB, resulting in the inclusion of a regulated exon in a number of transcripts in muscles or neurons [151–157]. Similar situations have also been described for neuron-specific NOVA proteins [158, 159].

Examples of complete activation or repression of splice sites (activator and repressor models) for tissue-specific factors are most well known from the *Drosophila* sex determination pathway [130, 148, 149, 160]. Here, the female-specific RNA binding protein Sex-lethal (Sxl) autoregulates splicing of its own transcript and splicing of the RNA binding protein *transformer* (*tra*). In both cases, binding of Sxl results in blockage of the male-specific splicing mode. In the *Sxl* gene, a male-specific exon is skipped in females, and in the *tra* gene, Sxl binding blocks the use of a 3′ splice site, resulting in expression of Sxl and Tra proteins in females, but not in males. Sxl is also required for dosage compensation in females by blocking splicing of an intron in *male-specific-lethal-2* (*msl-2*), and also translation of *msl-2*, to prevent upregulation of transcription on the X chromosome as occurs in

males [161]. Tra is then required together with constitutive factors Tra-2 and RBP1, among others, to activate a weak 3′ splice site in the *doublesex* (*dsx*) gene that leads to the inclusion of a female-specific exon and results in female differentiation in the fly. Regulation of *dsx* by *tra* illustrates an example of an activator model. An additional branch in the sex-determination pathway required for establishing sex-specific courtship behavior of the male fly is also regulated at the level of splicing. In the fruitless (*fru*) gene, Tra together with Tra-2 activates a 5′ splice site in females, resulting in a switch to a *fru* spliceform that does not translate into a protein. Strikingly, the Fru protein from the male spliceform is both necessary and sufficient, if expressed in females, to induce male courtship behavior [162–164].

In case of Tra-regulated sex-specific splicing of *dsx* and *fru*, multiple Tra binding sites are found adjacent to the regulated splice sites. In contrast, binding sites for many other tissue-specific RNA binding proteins have been found in places where regulatory mechanisms are not readily obvious, and in many cases multiple binding sites are distributed in the vicinity of processing signals [110, 165]. For the auto-regulatory skipping of the male-specific exon in *Sxl* transcripts, several Sxl binding sites flanking the regulated exon are important [166–168]. Although Sxl seems to interfere with recognition of the 3′splice site of the skipped exon here by regulating the second step of splicing in an *in vitro* system [169], the situation in the female fly seems to be more complicated and also involves Sxl-mediated interactions of U1 snRNP and U2AF [170]. In *Drosophila nrg* transcripts, where ELAV regulates skipping of the terminal exon, several ELAV binding sites are spread over the regulated intron and locate in the proximity of processing signals [171]. In the case of Nova-1 regulated splicing of a GlyR$\alpha$2 intron, the Nova-1 binding sites are clustered ~1 kb upstream of the regulated exon, and the presence of Nova-1 leads to inclusion of the regulated exon [158]. Interestingly, Nova-1 also auto-regulates inclusion of an exon in its own gene by binding to this exon, which is then skipped [159]. Hence, as illustrated by the role of Nova-1 in auto-regulation and neuron-specific splicing of GlyR$\alpha$2, it is not a factor per se that determines the splicing mode; rather, the output is determined by the context-dependent integration of exonic and intronic regulatory elements via bound trans-acting factors.

A further extension to the regulation of splicing has been the unexpected finding that expression levels of spliceosomal components attributed to constitutive functions vary during development in *Drosophila* [172]. Conversely, to maintain the delicate balance of positive and negative splicing signals, many RNA binding proteins auto-regulate their expression [159, 160, 173–179]. Consequently, overexpression of RNA binding proteins in *Drosophila* is mostly deleterious for the organism [180, 181]. Thus, tight control of expression or activity levels of the proteins involved in pre-mRNA processing is also required to compensate for the sequence degeneracy of binding sites.

A further mechanism to vary the information content of transcripts is the regulated use of 3′end processing sites, and many genes have alternative 3′ exons, or alternative polyA sites in the 3′UTR [4, 182]. Although much less is known about alternative polyA site choice and the mechanisms that coordinate 3′end processing and splicing in alternative situations, similar principles as for alternative splicing apply. In immunoglobulin M (IgM) pre-mRNA processing, splicing of the last intron resulting in membrane-bound IgM is prevented by use of an intronic pA site to produce secreted IgM in later stages of B-cell development (reviewed in [183]). Here, a competitive situation in polyA site recognition mediated by increased activity and concentrations of CstF64 and counteracted by hnRNP F and U1A have been attributed to premature 3′end formation versus splicing [184–188]. In addition, U1A also binds upstream of the regulated polyA site and affects polyA tail length, thus further reducing levels of secreted IgM by reducing mRNA stability [189, 190]. Inhibition of polyA tail length by U1A occurs through a direct interaction with PAP, and this interaction is also important for U1A auto-regulation [176, 191]. In addition to U1A, another component of U1 snRNP, U1 70K, can also inhibit PAP [192] and U1 snRNP or U1A can also prevent cleavage if the binding site confers a DSE [92, 188]. A further mechanism to favor intronic 3′end processing of IgM transcripts has been attributed to an RNA Pol II pause site following the intronic polyA site [193]. Another type of regulating 3′end processing has been described for a polyA site in the last intron of the *Drosophila ewg* gene. This intronic polyA site is used in non-neuronal tissue, while in neurons splicing of the last intron is induced [117, 194]. In neurons, the tissue-specific RNA binding protein ELAV inhibits cleavage by binding to a DSE. In contrast to the situation in the IgM gene, however, ELAV does not interfere with polyA site recognition, but inhibits cleavage at later steps and as a consequence allows for splicing. Interestingly, binding of a partial polyA complex together with ELAV is further important for 3′splice site choice [117], indicating that deliberate combinatorial interactions operate to coordinate splicing and 3′end processing.

In addition to the numerous combinatorial interactions of RNA processing factors their regulatory potential can be further increased by post-translational modifications, for example, phosphorylation or methylation, to regulate properties involved in binding RNA or interacting with other proteins (for review see [195–197]). The most prominent phosphorylated RNA binding proteins are SR proteins [72]. Phosphorylation is required for the activity of SR proteins and can result in the regulation of alternative splicing [198]. Dephosphorylation of the SR protein

SRp38 has also been linked to the shutdown of all splicing upon heat shock [199]. In recent years the number of post-translational modifications identified in RNA binding proteins has increased, but only in a few examples have these modifications been connected to cellular signaling pathways or to the regulation of pre-mRNA processing. One of the most-studied examples is the regulated inclusion of exon v5 of CD44 by the ras-Erk pathway [200, 201]. Here, phosphorylation of SAM68 regulates binding to target RNA and inclusion of exon v5 in CD44 [202]. A broader involvement of the ras-Erk pathway in the regulation of RNA processing is further indicated, as it also phosphorylates hnRNP K [203]. In other instances, sequence elements but no trans-acting factors have been identified that regulate alternative splicing through cellular signaling as in slowpoke $Ca^{2+}$ and voltage-gated $K^+$ channels expressed in neurons. Here, inclusion of the STREX exon is inhibited by CaMK IV upon stress hormone-induced depolarization that increases intracellular $Ca^{2+}$ levels [204]. Interestingly, the STREX exon enhances the signaling strength of the channel, and thus exclusion of the STREX exon provides an adaptive feedback regulation to increased stimulation.

Genetic polymorphisms are a further factor that can affect the regulation of pre-mRNA processing [28]. The potential impact of genetic polymorphisms is illustrated by the allele-specific alternative splicing in the human growth hormone receptor gene [205]. Here exon 3 is skipped depending on the allele present.

## Splicing of large introns: regulatory mechanisms at genomic dimensions

A common feature of genes in higher eukaryotes is the presence of very large introns, often extending over tens of kilobases. In the human *neurexin 3* gene, which spans 1600 kb, the largest constitutively and alternatively spliced introns are 292 kb (between exon 16 and 17) and 347 kb (between exon 1 and 5), respectively (Fig. 2a) [18], and an intron in a *Drosophila* fertility gene on the Y chromosome has a size of 3000 kb [206]. Frequently hidden in such large introns are very short alternatively spliced exons (cassette exons, Fig. 1) such as the 12 nt exon 4 in the *neurexin 3* gene (Fig. 2a). Furthermore, large introns can also harbor entire additional genes (7% in *Drosophila*), or large genes overlap with ones that are transcribed in the opposite direction (15% in *Drosophila* [207]).

One mechanism to facilitate splicing of large introns involves an extension of the intron definition model. As RNA is single stranded, it folds into elaborate secondary and tertiary structures. Superimposing structural constraints on intron sequences can therefore result in looping out of intronic sequences and bring splice sites into proximity (Fig. 4a). An example for the importance of RNA folding for correct pre-mRNA processing is found in *para* sodium channel transcripts. In a temperature-sensitive allele of a DEADbox-helicase ($nap^{ts}$) in *Drosophila*, pre-mRNAs do not correctly fold, as detected by the absence of A-I editing at a specific site, and misfolding causes a splicing catastrophe in *para* transcripts [208]. Using a complementary RNA to sequences flanking both sides of an exon has also been shown to induce exon skipping of an in-frame exon in a gene-therapy approach to correct mutations in the $\beta$-globin gene [209]. In a model for splicing in the mutually exclusive exon 6 cluster in the *Dscam* gene (Fig. 2c), long-range interactions together with splicing repression have been proposed key to selection of a single exon from the 48 variants [144]. Here, pairing between a docking site close to the 5′ splice site in intron 6 and a complementary sequence in front of every exon in the intron 6 cluster brings the 3′ splice site of a single exon into the proximity of the 5′ splice site. Splicing of the proximal intron is then proposed to relieve the repressed state of the selected exon 6 variant and allow for splicing of the distal intron to join to the constant exon 7.

Another way to bring splice sites into proximity by looping out intronic sequence is through protein-protein interactions. Of particular relevance for this scenario are a number of abundant hnRNP proteins that co-transcriptionally cover native transcripts, associate into multimers [93, 94, 96, 210] and can bring splice sites into proximity, according to the intron definition model (Fig. 4b). As this scenario can also apply to alternative splicing, such a loop-out model has been proposed for auto-regulatory exon skipping in *hnRNP A1* transcripts when hnRNP A1 concentrations increase [174]. A similar model has been described for the regulated skipping of a neuronal exon in *cSrc* splicing, where PTB binds to sites flanking the regulated exon and loops it out. In neurons, a neuronal form of PTB, nPTB, has been implicated in disrupting PTB multimeric interactions that loop out the neuronal exon [119, 211, 212]. Skipping of the regulated exon, however, also involves PTB interference with intron definition (Fig. 3b) by preventing the assembly of U2AF to the constitutive 3′ splice site to form a pre-spliceosomal complex with U1 snRNP from the regulated exon [120].

Besides looping-out of intronic sequences, splicing of large introns can occur by three additional mechanisms: recursive, intra- and trans-splicing. In the *Drosophila Ubx* gene, two cassette exons of 51 nt that are also alternatively spliced are located in a total of 73.5 kb of intronic sequence (individual intron sizes are 7.4, 15.6 and 51.5 [213]). Splicing of this large intron involves recursive splicing: after splicing of an intron, the 5′ splice site is regenerated and is used again (Fig. 4c). Bioinformatic, phylogenetic and experimental evidence suggests that recursive splicing is a general mechanism in *Drosophila* to

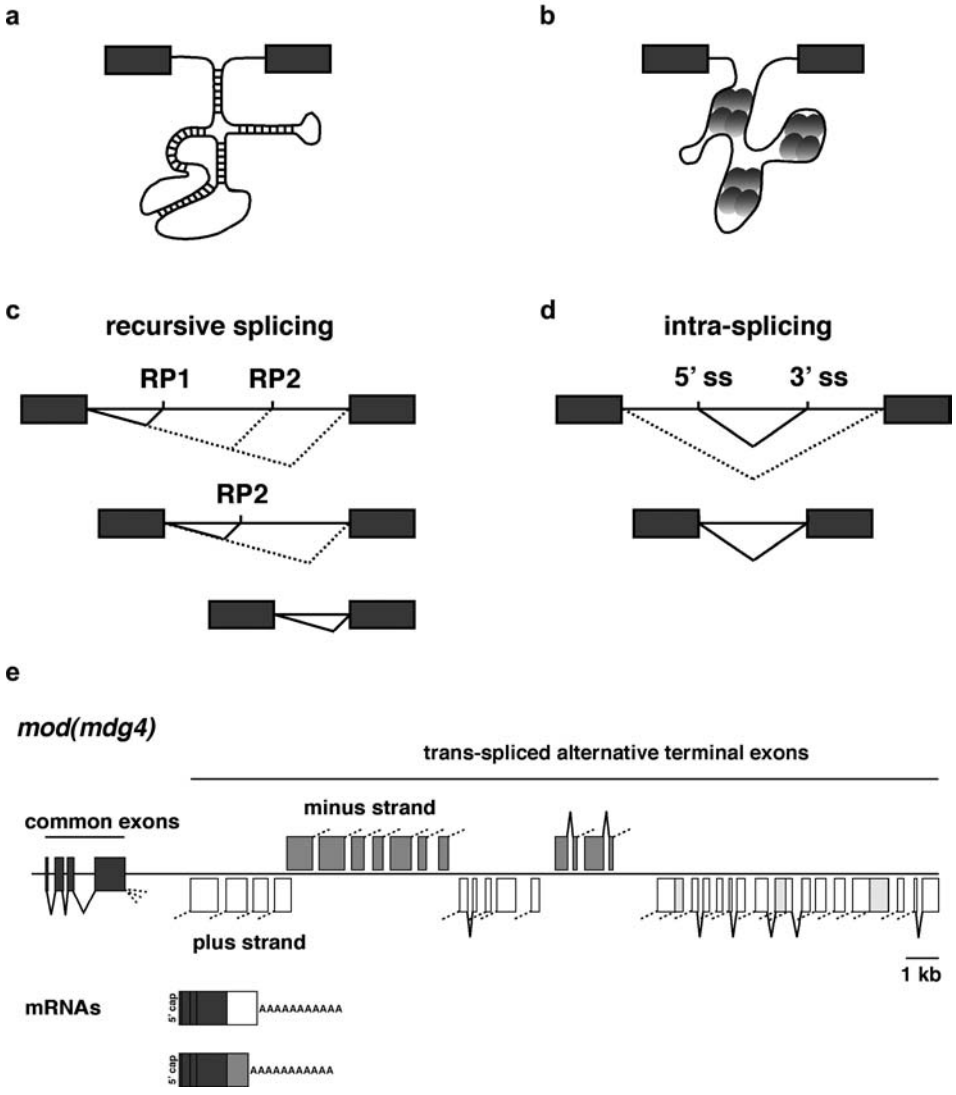


**Figure 4.** Mechanisms involved in splicing of large introns. RNA structure (*a*) or multimerization of hnRNP proteins (*b*) can support the intron definition model by looping out intronic sequence. Recursive splicing (*c*) can sequentially shorten introns as they are synthesized by RNA Pol II through splice sites that are regenerated after splicing (RP, ratcheting point). In intra-splicing (*d*) a nested intron is first spliced before flanking splice sites are used. Trans-splicing (*e*) of two separately transcribed RNAs from a constant (black) or variable region (white and grey) generates *mod(mdg4)* mRNA isoforms in *Drosophila* (bottom). White exons are encoded on the same DNA strand as the constant region and grey exons are encoded by the opposite DNA strand (modified from [219]).

splice large introns of 10 kb and more, and also includes the use of non-exonic splice sites [214].

As suggested by bioinformatic analysis, but not yet experimentally proven, the size of large introns might also be reduced by intra-splicing (Fig. 4d) [215]: splicing of one or more introns within an intron. Intra-splicing might also occur in nested genes that are transcribed in the same direction, as it would exclude ambiguity of splice sites between the two transcription units. In contrast to recursive splicing, where splice sites can be used as they appear on native transcripts, intra-splicing involves two 5′ splice sites that potentially could compete. Intra-splicing might therefore involve a protecting mechanism for the first 5′ splice site to reserve it for the last splice. Alternatively, intra-splicing could lead to the generation of functional splice sites and sequential removal of nested introns, as was demonstrated in an artificial system in yeast [216].

A third possibility to overcome processing of large introns is trans-splicing of exons that are transcribed separately. Trans-splicing is used for all transcripts in trypanosomes and in about 15% of the genes in *C. elegans* to add a 5′ leader sequence, which is transcribed from a separate gene, to the open reading frame encoding part of a pre-mRNA [217, 218]. In *Drosophila*, trans-splicing has been demonstrated as a mechanism to generate molecular diversity in the *mod(mdg4)* gene (Fig. 4e) and in the *lola* gene (Fig. 2d). In the *mod(mdg4)* gene, a total of 31 terminal exons are present in the C-terminal part of this tran-

scription factor that encode zinc finger DNA binding domains, and intriguingly, some are transcribed in the opposite direction relative to the constant 5′ part of the gene that encodes a BTB dimerization domain (Fig. 4e, [219, 220]). Trans-splicing of exons encoded on both strands has been shown by transgene experiments [220, 221]. Transgenes containing a series of variable exons inserted at different chromosomal locations can rescue *mod(mdg4)* mutants, and expression of protein is detected in these flies. A further example of functionally relevant trans-splicing has been found in the *lola* gene, another BTB zinc finger transcription factor, where intragenic complementation of two otherwise lethal mutations was observed [222]. Since intragenic complementation has been reported in many other large genes in *Drosophila*, trans-splicing might be more common than anticipated. In vertebrates, trans-splicing has been detected in several genes, but might not be functionally relevant as it occurs at very low levels and no protein has been detected for these isoforms, and may therefore rather comprise 'splicing noise' [148, 223]. A reason why trans-splicing might be more favored in *Drosophila* compared with vertebrates could be explained in the pairing of chromosomes in interphase nuclei [224] and the presence of multiple chromosomal copies (polyteny), giving the potential that several transcripts are available for splicing at the same time. Nevertheless, trans-splicing has proven successful in gene therapy approaches by expressing a corrected version of an exon that pairs with the native transcript through an artificially introduced complementary sequence in the intron, thus competing with the endogenous exon for splicing [31]. Potentially, trans-splicing between two pre-mRNAs could also occur from an intrachromosomal configuration to overcome transcription of large introns either by transcription from two separate promoters upstream of spliced exons or by the presence of transcripts from consecutive polymerases. Such a mechanism, however, likely requires spatial organization of transcription and pre-mRNA processing into an 'mRNA factory' to preserve transcription units (see below and Fig. 5).

Another complication observed in genes with large introns is the frequent presence of consensus AAUAAA polyA recognition sites. Premature 3′end processing in introns would generate a truncated protein and probably also lead to abortion of transcription by the torpedoing 5′-3′ exonuclease Xrn2 involved in transcriptional termination [225]. Intron sequences are generally AU-rich, and many of the AAUAAA sequences also have downstream U-rich sequences that could be recognized by CstF and lead to cooperative formation of a CPSF-CstF polyA recognition complex necessary to initiate 3′end processing. In addition, about 7% of *Drosophila* genes encode nested genes in introns that are transcribed in the same direction and ergo, contain functional cleavage sites [207]. Many of these nested genes are also polyadenylated genes. As aberrant proteins from such internally truncated transcripts are not recognized by the cell's quality control machinery according to the current understanding, presence of such truncated proteins would likely result in deleterious effects (see below). Thus, mechanisms must exist that distinguish intronic 3′end processing sites in internal introns from processing sites following the terminal exon. Possible roles in inhibiting intronic 3′end processing could be attributed to the CTD of RNA Pol II, as it has been shown to bind to AC-rich RNA and suppresses transcription-coupled 3′end formation [226]. Also, presence of a strong 5′ splice site can inhibit 3′end processing through U1 snRNP as described earlier, and in simplified systems the presence of an artificially inserted polyA site in an efficiently spliced intron is ignored [61]. In large introns, however, often several consensus AAUAAA polyA recognition sites are present, eluding the unique effect of a single 5′ splice site. Therefore, it is likely that in larger introns several regulatory backups may operate to prevent premature 3′end processing. In contrast to regulated polyA sites, however, 3′end processing at intronic sites must be completely inhibited, but no such rigorous mechanisms have been described yet.

A prominent class of intron encoded genes are snoRNAs that are mainly involved in post-transcriptional modification of rRNAs [227, 228]. In mammals, one class of snoRNA genes is preferentially located about 70 nt upstream of the 3′ splice site, and processing is coupled to late steps in splicing [229, 230]. In a second class of snoRNAs, an external stem-loop structure is required for processing, but processing is independent of splicing and exhibits no restriction on intronal position [230].

A further intriguing situation regarding the definition of a transcription unit is observed in the functionally unrelated *NUP62* and *IL41* genes, which have become connected through 'promoter sharing' [231]. The *NUP62* gene is ubiquitously expressed, while the following *IL41* gene, transcribed in the same direction, is only expressed in B cells. In neurons and testes, however, transcripts initiate at the *NUP62* promoter and splice out the *NUP62* open reading frame to express the *IL41* gene, likely representing the acquisition of an evolutionarily new function of *IL41* in these tissues through regulation of splicing and 3′end processing.

## Coupling of transcription with pre-mRNA processing

Initially, the ability to perform all pre-mRNA processing reactions individually *in vitro* as well as the efficient splicing of pre-mRNAs injected into *Xenopus* eggs has suggested that first, all pre-mRNA processing reactions occur relatively autonomously from each other *in vivo*, and second, that they occur post-transcriptionally. Several

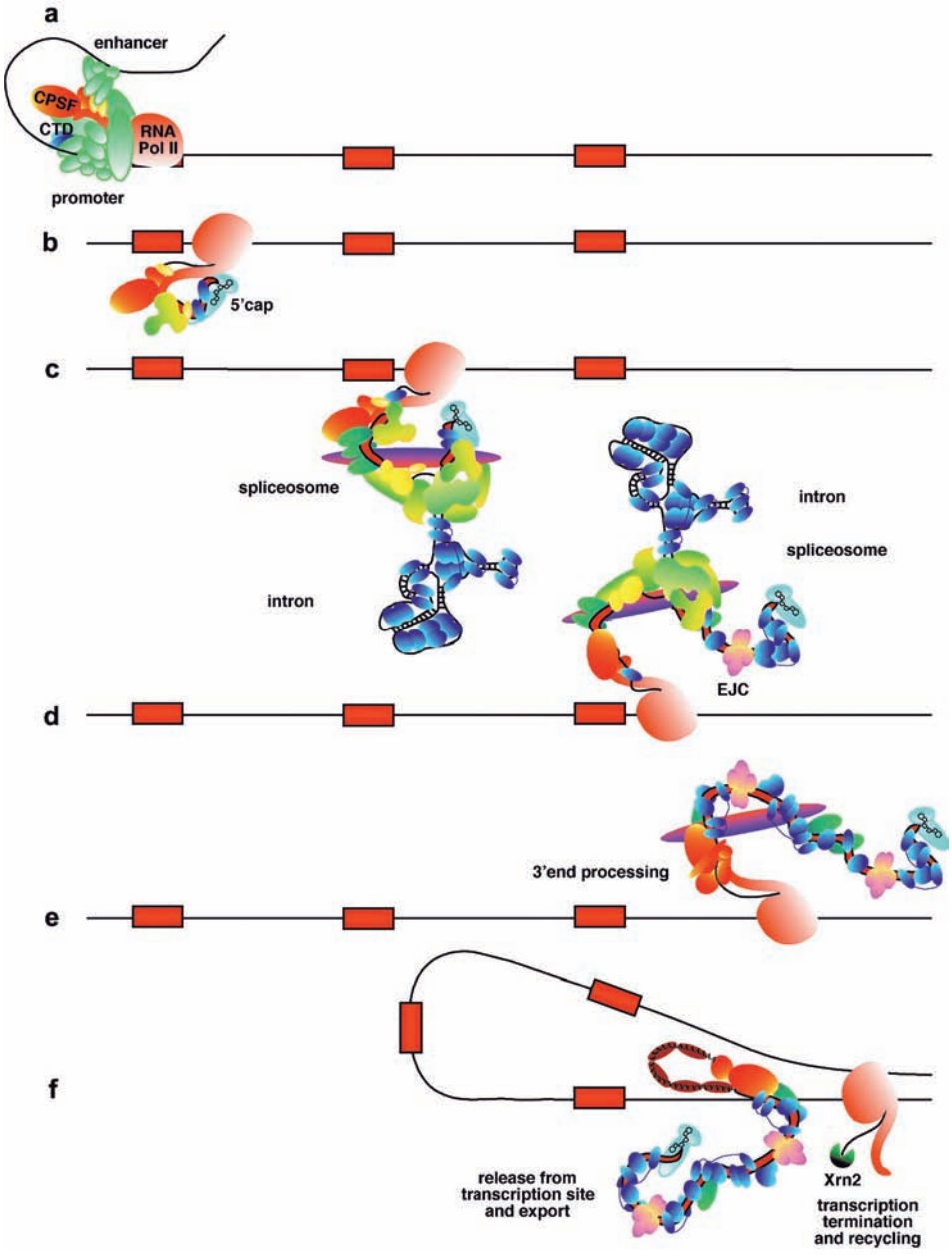


**Figure 5.** Depiction of a model for an 'mRNA factory' through the coupling of transcription with pre-mRNA processing during one cycle starting with the initiation of transcription at the promoter (*a*) and association with pre-mRNA processing factors (yellow), elongation of the synthesized pre-mRNA by RNA Pol II (brown, *b–e*) concomitant with splicing of introns (*c* and *d*) and deposition of the exon junction complex (EJC, pink, *d*) followed by 3′end processing (*e* and *f*), transcriptional termination involving the 5′-3′ exonuclease Xrn2 (dark green, *f*) and recycling of RNA Pol II and pre-mRNA processing factors as proposed by a gene loop model (*f* and *a*). The model has been drawn according to numerous published interactions (see text [129, 233, 237]) and also incorporates structural information from Wetterberg et al. [315]. Transcriptional regulators are shown in light green, the cap binding complex in light blue, hnRNP proteins in blue, SR proteins in green, 3′end processing factors in red and polyA binding protein in dark red.

recent findings, however, point out that pre-mRNA processing is highly coupled to transcription. One of the first indications was the observation that transcription of snRNA genes, which have a RNA Pol II promoter but are not polyadenylated, from promoters coding for polyadenylated genes results in 3′end processing defects [232]. Subsequently, 3′end processing factors were found to associate with promoters [233]: CPSF interacts with general transcription factor TFIID bound to the core promoter, and CstF interacts with the transcriptional coactivator PC4, suggesting that also remote regulatory sequences are involved in coupling of pre-mRNA processing with transcription (Fig. 5a). Besides these interactions, numerous other factors with roles in pre-mRNA

processing have also been found to interact with promoter-associated factors [15, 234]. A major integrating function is attributed to a unique feature of RNA Pol II, the CTD, to orchestrate pre-mRNA processing through its interaction with many RNA processing factors and recruitment to sites of transcription [235, 236]. Sequential appearance of processing signals when nascent transcripts are synthesized by RNA Pol II for co-transcriptional processing also adds restriction on the choice of processing sites and counteracts sequence degeneracy. The coupling of transcription with pre-mRNA processing has also led to the model of 'RNA-factories' (Fig. 5) [13, 237], which might even be gene-specific due to the promoter-specific effects on pre-mRNA processing (see below and [15, 234]).

The CTD of RNA Pol II is composed of 52 imperfect tandem repeats of the heptad consensus sequence YSPTSPS and is subject to differential regulatory phosphorylation during transcription. Ser5 is highly phosphorylated when transcription is initiated, while during elongation Ser2 is phosphorylated [237]. The efficiency of all major pre-mRNA processing reactions (capping, editing, splicing and 3′end processing) is enhanced by the CTD *in vivo* (reviewed in [12, 129, 238]). As a consequence, shortening of the CTD, or transcription with other polymerases that lack a CTD (RNA Pol I or III, or T7 RNA Pol) results in severe processing defects [232, 234]. Intriguingly, different portions of the CTD were also found to support different pre-mRNA processing reactions [239, 240]. The N-terminus supports only capping, while the C-terminus supports all reactions, and this activity is attributed to 10 additional amino acids at the C-terminus of the CTD. This portion of the CTD might integrate several functions and also confers stability to RNA Pol II [241]. Central to the coupling function of RNA Pol II is the dynamic association of RNA processing factors with the CTD [12, 129]. In particular, phosphorylation of Ser2 that promotes elongation has been implicated to regulate the association with processing factors [242]. A further level of control in RNA Pol II CTD function has been indicated by the involvement of Pin1 in regulating the isomerization state of prolyl residues that affects the structure of the CTD and determines the degree of CTD phosphorylation [243]. The CTD also enhances the efficiency of *in vitro* splicing and polyadenylation reactions [244, 245]. In *in vitro* splicing assays, the effect of the CTD is strongest if the splicing substrate contains two introns [246]. Hence, the CTD enhances exon definition, most likely by stabilizing combinatorial interactions of various factors bound to the exon to facilitate formation of the spliceosome.

Current models suggest that RNA processing factors are loaded onto the RNA Pol II CTD and deposited on native transcripts as they are synthesized. These models also imply that specific promoters might differentially affect alternative pre-mRNA processing by interaction with different factors. Indeed, promoter dependence has been demonstrated for alternative splicing of the EDI exon in fibronectin transcripts [14, 247]. Here, the promoter determines processivity of RNA Pol II. A fast polymerase leads to skipping of the EDI exon, while a slow polymerase allows for inclusion of the regulated exon, presumably by giving more time for splice site recognition. This mechanism also involves a polar effect. Inclusion of distal alternative exons coordinately stimulates inclusion of proximal alternative exons, but not vice versa [248]. The degree of exon skipping inversely correlates with the strength of the 3′ splice site of the skipped exon. RNA Pol II processivity as measured by exon skipping is further regulated by its phosphorylation state and chromatin structure. Similar results regarding Pol II phosphorylation and chromatin structure have also been obtained in a yeast system [249]. Recently, intragenic DNA methylation was shown to reduce RNA Pol II processivity through a closer chromatin structure [250], indicating that alternative splicing might also be subject to epigenetic regulation.

The effect of promoters on alternative splicing is mediated through associated factors. Several co-regulators of nuclear hormone receptors showed different effects on alternative splicing of steroid hormone-regulated genes, but had no effect on alternative splicing of other genes [251, 252]. In case of the co-regulator TRBP/ASC-2, which is recruited to regulatory regions on the DNA by steroid hormone-activated nuclear hormone receptors, the effects on alternative splicing are mediated by interaction with the co-activator CoAA, an hnRNP-like protein [253]. The functions in transcriptional regulation and alternative splicing could also be attributed to different parts of the U2AF65-related protein CAPER$\beta$ [254]. In addition, transcriptional activators were shown to affect the efficiency of constitutive splicing in a promoter and RNA Pol II CTD-dependent manner [255].

Transcription of a pre-mRNA involves co-transcriptional packaging with RNA binding proteins. Packaging occurs mostly with the ubiquitously present hnRNP-type proteins and plays an important role in the regulation of pre-mRNA processing and formation of an export-competent mRNP particle [93, 210, 256–258]. Co-transcriptional packaging of nascent transcripts with the THO/TREX export factor complex in yeast or with the SR protein ASF/SF2 in mammalian cells is also important to prevent hyperrecombination of transcribed genes [259, 260]. Packaging defects result in heteroduplex formation of the nascent transcript with the DNA template (R-loop) behind the elongating RNA Pol II. R-loop formation stimulates recombination and, as a result of decreased elongation rates, causes pre-mRNA processing defects such as pre-mature termination [261, 262].

Coupled to transcription are mechanisms that control the quality of transcripts. Premature stop codons (PTCs) in mammals or failure to form an export-competent RNP

due to defective 3′ ends in yeast can lead to an accumulation of transcripts at the site of transcription which are degraded by the exosome in yeast [263, 264]. Currently a controversial issue is if and how these PTCs are recognized in the nucleus and whether this involves nuclear translation (for pro and contra arguments see refs. in [265]). PTC-containing transcripts can be degraded in a process called nonsense mediated decay (NMD), and this involves a pioneer round of translation in mammals, where mRNAs are still bound by the nuclear cap binding complex and nuclear polyA binding protein [266]. This pioneer round of translation removes a protein complex termed exon junction complex (EJC) which has been deposited after splicing 20–24 nt upstream of a splice junction [265, 267]. If a stop codon is located 50–55 nt upstream of an EJC, then NMD can be triggered in mammals. Many alternatively spliced exons introduce PTCs and might be subject to degradation [268]. This mechanism is in fact used for auto-regulation of PTB, TIA-1 and TIAR expression [178, 179]. Not all PTCs, however, trigger NMD, as transcript levels of apoB mRNAs in which a PTC is introduced by editing remain unchanged [269]. The nuclear events leading to co-transcriptional formation of an export-competent mRNP particle can also determine the fate of mRNPs in the cytoplasm [270]. Splicing-dependent deposition of an EJC complex, together with 3′UTR sequences has been shown to be important for localization of *oskar* RNA to the posterior pole of *Drosophila* oocytes, suggesting that splicing can determine the architecture of the *oskar* mRNP [271].

Transcription is coupled to 3′end processing by polyA factors that interact with promoters and with the CTD of RNA Pol II [233]. Although a functional polyA signal is a prerequisite for transcriptional termination, 3′end processing and transcriptional termination are not coincident events, and several models of how transcriptional termination is realized exist [63]. Unlike in prokaryotes, where transcription is terminated shortly after the protein coding sequence either by specific sequences forming a stem loop structure (anti-terminator model) or by the exonucleolytic Rho terminator protein (torpedo model) [272], the situation in eukaryotes is less clear as termination occurs at several sites extending over several kilobase pairs. Clearly, mechanisms must exist that keep RNA Pol II transcribing through large introns, but then stop transcription at the end of a gene, as transcription into neighboring genes is likely deleterious. While the presence of 5′ introns has a stimulatory effect on transcription [273], functional polyA sites result in the opposite effect [63]. PolyA site recognition has been associated with a reduction in elongation rates of RNA Pol II as a consequence of downloading polyA factors from the CTD and with a change in the phosphorylation state of the CTD [274–278]. Pause sites following polyA sites have also been associated with termination [279], and a pause site following an intronic polyA site has been attributed to alternative regulation of the IgM processing in favor of an upstream polyA site [193]. A particular role in transcriptional termination at pause sites has been revealed for Pcf11, a subunit of cleavage factor II, in triggering transcript release from Pol II through an interaction with the CTD [280]. Pause sites per se, however, are not sufficient to trigger termination as seen in *Drosophila* heat shock genes [281]. Chromatin remodeling has also been implicated in transcriptional termination by reducing RNA Pol II processivity through a closer chromatin structure [282].

Recent support of the 'torpedo' model has come from the identification of a conserved 5′-3′ exonuclease that uses the cleavage site as entry point to catch up with RNA Pol II through its high processivty to displace it [225, 283]. Intriguingly, a phenomenon termed co-transcriptional cleavage (CoTC) has been described in the intergenic region of the $\beta$-globin gene [284]. This region where CoTC occurs has been shown to encode a ribozyme activity that leads to self-cleavage of nascent transcripts and thus generates the entry point for the torpedoing exonuclease [285]. Hence, given this recent support of the 'torpedo' model, cleavage at intronic sites likely is deleterious and must be efficiently inhibited (see above).

An increase in regulatory potential at genomic levels coupled to transcription arises in regions with overlapping genes. This situation is encountered in 22% of *Drosophila* genes, and in about 10 and 7% of human and *Arabidopsis* genes, respectively [207, 286, 287]. In *Arabidopsis*, a large portion of overlapping genes are co-expressed, and despite that overlapping genes could form double-stranded RNA subject to degradation by the RNAi pathway (reviewed in [288]), transcripts from these genes are not underrepresented. In contrast, alternative splicing and polyadenylation are increased in regions where introns or 3′UTRs are overlapping, respectively [287]. Evidence for direct involvement of anti-sense transcripts in the regulation of alternative splicing is indicated in the thyroid hormone receptor ErbA$\alpha$ [289]. Given that large portions of the human and fly genomes are expressed at low levels into non-coding and anti-sense RNAs holds great potential for the regulation of alternative pre-mRNA processing [8, 290].

## Defining processing units: nuclear compartments, alternative promoters, gene loops

Despite the high sequence degeneracy of pre-mRNA processing signals, pre-mRNA processing occurs with high fidelity and accuracy in a complex cellular environment. Spatial organization of processing units would greatly reduce sequence complexity and the number of processing choices. Several distinct compartments have been de-

scribed in the nucleus such as the nucleolus, which is the site of transcription and processing of ribosomal RNAs, the nuclear 'speckles', which are the assembly/storage site for spliceosomal components and the Cajal body, a proposed site for snRNP assembly [291–293]. Chromosomes are also organized into distinct nuclear territories and many co-expressed genes reside in neighborhoods or are clustered such as genes for ribosomal RNAs (rRNAs), or transfer RNAs (tRNAs) [294]. The highly dynamic organization of the nucleus in living cells resulting, for example, in only very short interactions of transcription factors with sites of transcription and the lack of a static nuclear skeleton, however, have indicated that much of the interactions in the nucleus are stochastic. The presence of compartments in a highly dynamic environment, which can also be generated artificially as introducing a plasmid with ribosomal genes results in the formation of additional nucleoli, has led to the suggestion that the nucleus may be a self-organizing entity [295–297]. Despite the highly dynamic organization of the nucleus, essential functions to orchestrate spatial and temporal control of gene expression have also been attributed to nuclear matrix and scaffolding factors such as SAF-A/hnRNP U, SAF-B, SATB1 and ARBP [298]. Attachment of chromatin to the nuclear matrix, for example by SAF-B shown to stimulate transcription-dependent splicing, might thus provide a framework for organizing spatially restricted gene expression factories [13, 237, 299].

The strong connection between pre-mRNA processing and gene promoters also suggests that promoters might define units that can determine alternative processing of a pre-mRNA. Hence, differential loading of pre-mRNA processing factors depending on the promoter would also greatly reduce the sequence complexity pre-mRNA processing factors could potentially encounter. Although only a handful of examples of nuclear hormone receptor-regulated alternative pre-mRNA processing are currently known (see above [14, 15]), the high number of these types of transcriptional regulators present in the human genome and the frequent presence of alternative promoters offers a large potential for this type of coordinated regulation of pre-mRNA processing. Furthermore, this sort of regulation would also contribute to the generation of a 'nuclear history' of transcripts relevant for the diversification of fates in the cytoplasm.

A further possibility to define processing units is the use of the second type of intron. U12-type introns are spliced about sixfold slower, and the slow speed might be important in rate-limiting steps [300]. Functional relevance for U12-type introns is indicated by their conserved position in homologous genes or presence in paralogous genes [58]. In another instance found in the Drosophila *prospero* gene, a minor class U12 intron flanks a major class U2 intron, and splicing occurs by one or the other pathway, resulting in a five amino acid difference in the functionally important homeodomain [301]. How different spliceosomes are recruited, however, is currently unknown.

An interesting connection between the beginning and the end of a gene has been revealed by the association of 3′end processing factors with promoters [233]. A contrast to the idea that 3′end processing factors are loaded on RNA Pol II at the promoter and then deposited at the end of the gene is provided by results from ChIP analysis (chromatin immuno-precipitation) with 3′end processing factors. In several analyzed yeast genes, 3′end processing factors only localized to the terminator region [302]. An alternative explanation to the association of 3′end processing factors with the promoter could also be the juxtaposition of the terminator with the promoter region to form a 'gene loop' as demonstrated for yeast genes [303]. Similar loop structures have also been shown for enhancer-promoter interactions in the $\beta$-globin locus [304–306]. Given the interaction of CstF with enhancer-associated PC4 [233], all control regions of a gene might engage in physical interactions to define a processing unit and enhance recycling of processing factors (Fig. 5a, e). Although these interactions could be mediated by specific, yet to be identified sequence elements, chromatin structure could also contribute to this phenomenon. Furthermore, long-range interactions mediated by chromatin structure have been demonstrated through their effect on expression levels of genes in *trans* in the *Drosophila brown* locus [224]. Strikingly, an RNA-mediated mechanism in the nucleus was recently shown to locally modify chromatin structure (reviewed in [307]), and non-coding anti-sense transcripts might be important for defining transcription units [8, 290].

## Perspectives

The availability of whole genome sequences has very much facilitated the analysis of gene expression, starting with transcription of the DNA template under the control of transcription factors, followed by the processing of nascent transcripts into an mRNA and, after transport to the cytoplasm, translation into a protein. As indicated by the diverse organization of genes, numerous possibilities exist to regulate the expression of a gene into a protein at the level of RNA processing, ranging from changes in the primary sequence to spatiotemporal control of expression. Particularly overwhelming are the numerous examples for generating proteomic diversity through alternative pre-mRNA processing. Although our understanding of the principles that operate in pre-mRNA processing and its regulation has much increased in recent years, the degeneracy of regulatory elements suggests that highly combinatorial mechanisms must operate, as illustrated by the huge machineries involved in executing pre-mRNA

processing reactions to provide the great fidelity and accuracy. Deciphering the combinatorial codes that operate in localizing processing sites will certainly be one of the challenges in the coming years. The availability of whole genome sequences now allows for global analysis using experimental and bioinformatics approaches to RNA processing and its regulation, and will no doubt contribute to this goal.

The huge effort a cell puts into processing large genes into a mRNA involving many different proteins makes this process also vulnerable to genetic mutations that cause human disease [26–29]. In addition, as alternative splicing is particularly widespread in the brain, genetic polymorphisms or misregulation of feedback-enforced alternative splicing through neuronal activity might be a source of many psychiatric conditions and mood disorders [23, 33, 204]. Of particular interest regarding therapeutic approaches to interfere with these conditions are RNA binding proteins as regulators of alternative RNA processing. RNA binding proteins have specific sets of target genes and participate in distinct functional pathways as illustrated for the targets regulated by Nova-2 that are specifically enriched in synapses [308] and required for a novel form of synaptic plasticity, the cellular correlate of learning and memory [309]. Hence, the combinatorial interactions of factors involved in alternative RNA processing provide the setup to post-transcriptionally coordinate the expression of functionally related genes and also provide unique signatures for the regulation of a particular event [310]. Knowing the combinatorial codes therefore promises a high degree of specificity to therapeutically interfere with small molecule approaches in alternative pre-mRNA processing.

Further potential for pharmacological applications is also indicated by the numerous tissue-specific alternative spliced isoforms that can be targeted by specific drugs. An example of such applications is illustrated by non-steroidal anti-inflammatory cyclooxygenase inhibitors (e.g. aspirin, ibuprofen) which affect prostaglandine synthesis. Here, isoform-specific drugs such as acetaminophen that inhibit a brain-specific alternative splice isoform of the *COX1* gene are of interest, as general inhibition of COX results in upper gastrointestinal complications [31].

Recent discovery of microRNAs and their impact on the regulation of gene expression in the cytoplasm (reviewed in [288]), as well as transcription of much of the human genome into mostly non-coding RNAs [8, 290], indicates unexpected regulatory potential at the level of RNA processing. Furthermore, large portions of non-protein-coding parts of a *Drosophila* genome show selective adaptations arguing against previous assumptions that non-protein-coding regions are just 'junk' DNA [311]. Although the origin of life had been attributed to an 'RNA world', research from recent decades mainly focused on the impact of proteins. Clearly, protein function can be readily revealed with easily introduced point mutations that change an amino acid essential for protein function. RNA or DNA regulatory elements, however, are mostly immune to this type of mutational analysis. Hence, our view about the regulation of gene expression and pre-mRNA processing important for a functional genome has been very much biased [312].

*Acknowledgements.* I thank Irmgard Haussmann for comments and suggestions, and Sima Misra for a list of the nested genes in *Drosophila*. I am indebted to Kalpana White for her support. As this review intends to give a broad overview of all aspects of nuclear pre-mRNA processing and tries to integrate established and emerging concepts of genomic dimensions, I apologize to all colleagues for the bias in citations of original articles towards reviews of particular fields.

1 Lynch M. and Conery J. S. (2003) The origins of genome complexity. Science **302:** 1401–1404
2 Orphanides G. and Reinberg D. (2002) A unified theory of gene expression. Cell **108:** 439–451
3 Modrek B. and Lee C. (2002) A genomic view of alternative splicing. Nat. Genet. **30:** 13–19
4 Beaudoing E. and Gautheret D. (2001) Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST data. Genome Res. **11:** 1520–1526
5 Kampa D., Cheng J., Kapranov P., Yamanaka M., Brubaker S. and Cawley S. (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res. **14:** 331–342
6 Iida K., Seki M., Sakurai T., Satou M., Akiyama K., Toyoda T. et al. (2004) Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences. Nucleic Acids Res. **32:** 5096–5103
7 Reboul J., Vaglio P., Rual J. F., Lamesch P., Martinez M., Armstrong C. M. et al. (2003) C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. Nat. Genet. **34:** 35–41
8 Stolc V., Gauhar Z., Mason C., Halasz G., van Batenburg M. F., Rifkin S. A. et al. (2004) A gene expression map for the euchromatic genome of Drosophila melanogaster. Science **306:** 655–660
9 Lee B. T., Tan T. W. and Ranganathan S. (2004) DEDB: a database of Drosophila melanogaster exons in splicing graph form. BMC Bioinformatics **5:** 189
10 Levine M. and Tjian R. (2003) Transcription regulation and animal diversity. Nature **424:** 147–151
11 Kopelman N. M., Lancet D. and Yanai I. (2005) Alternative splicing and gene duplication are inversely correlated evolutionarily mechanisms. Nat. Genet. **37:** 588–589
12 Bentley D. (2002) The mRNA assembly line: transcription and processing machines in the same factory. Curr. Opin. Cell Biol. **14:** 336–342
13 Maniatis T. and Reed R. (2002) An extensive network of coupling among gene expression machines. Nature **416:** 499–506
14 Kornblihtt A. R. (2005) Promoter usage and alternative splicing. Curr. Opin. Cell Biol. **17:** 262–268
15 Auboeuf D., Dowhan D. H., Dutertre M., Martin N., Berget S. M. and O'Malley B. W. (2005) A subset of nuclear receptor coregulators act as coupling proteins during synthesis and maturation of RNA transcripts. Mol. Cell. Biol. **25:** 5307–5316
16 Schmucker D., Clemens J. C., Shu H., Worby C. A., Xiao J., Muda M. et al. (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell **101:** 671–684

17 Zhu J., Shendure J., Mitra R. D. and Church G. M. (2003) Single molecule profiling of alternative pre-mRNA splicing. Science **301:** 836–888
18 Tabuchi K. and Sudhof T. C. (2002) Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing. Genomics **79:** 849–859
19 Wojtowicz W. M., Flanagan J. J., Millard S. S., Zipursky S. L. and Clemens J. C. (2004) Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell **118:** 619–633
20 Watson F. L., Puttmann-Holgado R., Thomas F., Lamar D. L., Hughes M., Kondo M. et al. (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science **309:** 1874–1878
21 Goeke S., Greene E. A., Grant P. K., Gates M. A., Crowner D., Aigaki T. et al. (2003) Alternative splicing of lola generates 19 transcription factors controlling axon guidance in Drosophila. Nat. Neurosci. **6:** 917–924
22 Lee C. J. and Irizarry K. (2003) Alternative splicing in the nervous system: an emerging source of diversity and regulation. Biol. Psychiatry **54:** 771–776
23 Grabowski P. J. and Black D. L. (2001) Alternative RNA splicing in the nervous system. Prog. Neurobiol. **65:** 289–308
24 Stamm S., Ben-Ari S., Rafalska I., Tang Y., Zhang Z., Toiber D. et al. (2005) Function of alternative splicing. Gene **344:** 1–20
25 Wu J. Y., Tang H. and Havlioglu N. (2003) Alternative pre-mRNA splicing and regulation of programmed cell death. Prog. Mol. Subcell. Biol. **31:** 153–185
26 Faustino N. A. and Cooper T. A. (2003) Pre-mRNA splicing and human disease. Genes Dev. **17:** 419–437
27 Stenson P. D., Ball E. V., Mort M., Phillips A. D., Shiel J. A., Thomas N. S. et al. (2003) Human Gene Mutation Database (HGMD): 2003 update. Hum. Mutat. **21:** 577–581
28 Pagani F. and Baralle F. E. (2004) Genomic variants in exons and introns: identifying the splicing spoilers. Nat. Rev. Genet. **5:** 389–396
29 Stoilov P., Meshorer E., Gencheva M., Glick D., Soreq H. and Stamm S. (2002) Defects in pre-mRNA processing as causes of and predisposition to diseases. DNA Cell. Biol. **21:** 803–818
30 Meshorer E. and Soreq H. (2002) Pre-mRNA splicing modulations in senescence. Aging Cell **1:** 10–16
31 Garcia-Blanco M. A., Baraniak A. P. and Lasda E. L. (2004) Alternative splicing in disease and therapy. Nat. Biotechnol. **22:** 535–546
32 Dredge B. K., Polydorides A. D. and Darnell R. B. (2001) The splice of life: alternative splicing and neurological disease. Nat. Rev. Neurosci. **2:** 43–50
33 Meshorer E., Erb C., Gazit R., Pavlovsky L., Kaufer D., Friedman A. et al. (2002) Alternative splicing and neuritic mRNA translocation under long-term neuronal hypersensitivity. Science **295:** 508–512
34 McCracken S., Fong N., Rosonina E., Yankulov K., Brothers G., Siderovski D. et al. (1997) 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. **11:** 3306–3318
35 Cho E. J., Takagi T., Moore C. R. and Buratowski S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. **11:** 3319–3326
36 Izaurralde E., Lewis J., McGuigan C., Jankowska M., Darzynkiewicz E. and Mattaj I. W. (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell **78:** 657–668
37 Bisaillon M. and Lemay G. (1997) Viral and cellular enzymes involved in synthesis of mRNA cap structure. Virology **236:** 1–7
38 Bokar J. A., Shambaugh M. E., Polayes D., Matera A. G. and Rottman F. M. (1997) Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA **3:** 1233–1247
39 Flaherty S. M., Fortes P., Izaurralde E., Mattaj I. W. and Gilmartin G. M. (1997) Participation of the nuclear cap binding complex in pre-mRNA 3′ processing. Proc. Natl. Acad. Sci. USA **94:** 11893–11898
40 Lewis J. D., Izaurralde E., Jarmolowski A., McGuigan C. and Mattaj I. W. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5′ splice site. Genes Dev. **10:** 1683–1698
41 Fortes P., Inada T., Preiss T., Hentze M. W., Mattaj I. W. and Sachs A. B. (2000) The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. Mol. Cell **6:** 191–196
42 Parker R. and Song H. (2004) The enzymes and control of eukaryotic mRNA turnover. Nat. Struct. Mol. Biol. **11:** 121–127
43 Hoopengardner B., Bhalla T., Staber C. and Reenan R. (2003) Nervous system targets of RNA editing identified by comparative genomics. Science **301:** 832–836
44 Eisenberg E., Nemzer S., Kinar Y., Sorek R., Rechavi G. and Levanon E. Y. (2005) Is abundant A-to-I RNA editing primate-specific? Trends Genet. **21:** 77–81
45 Smith H. C., Gott J. M. and Hanson M. R. (1997) A guide to RNA editing. RNA **3:** 1105–1123
46 Wedekind J. E., Dance G. S., Sowden M. P. and Smith H. C. (2003) Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. Trends Genet. **19:** 207–216
47 Turelli P. and Trono D. (2005) Editing at the crossroad of innate and adaptive immunity. Science **307:** 1061–1065
48 Reenan R. A. (2001) The RNA world meets behavior: A→I pre-mRNA editing in animals. Trends Genet. **17:** 53–56
49 Bass B. L. (2002) RNA editing by adenosine deaminases that act on RNA. Annu. Rev. Biochem. **71:** 817–846
50 Rueter S. M., Dawson T. R. and Emeson R. B. (1999) Regulation of alternative splicing by RNA editing. Nature **399:** 75–80
51 Levanon E. Y., Eisenberg E., Yelin R., Nemzer S., Hallegger M., Shemesh R. et al. (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. **22:** 1001–1005
52 Blow M., Futreal P. A., Wooster R. and Stratton M. R. (2004) A survey of RNA editing in human brain. Genome Res. **14:** 2379–2387
53 Athanasiadis A., Rich A. and Maas S. (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. **2:** e391
54 Lev-Maor G., Sorek R., Shomron N. and Ast G. (2003) The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. Science **300:** 1288–1291
55 Jurica M. S. and Moore M. J. (2003) Pre-mRNA splicing: awash in a sea of proteins. Mol. Cell **12:** 5–14
56 Staley J. P. and Guthrie C. (1998) Mechanical devices of the spliceosome: motors clocks springs and things. Cell **92:** 315–326
57 Hastings M. L. and Krainer A. R. (2001) Pre-mRNA splicing in the new millennium. Curr. Opin. Cell Biol. **13:** 302–309
58 Patel A. A. and Steitz J. A. (2003) Splicing double: insights from the second spliceosome. Nat. Rev. Mol. Cell Biol. **4:** 960–970
59 Chen L., Lullo D. J., Ma E., Celniker S. E., Rio D. C. and Doudna J. A. (2005) Identification and analysis of U5 snRNA variants in Drosophila. RNA **11:** 1473–1477
60 Venkataraman K., Brown K. M. and Gilmartin G. M. (2005) Analysis of a noncanonical poly(*a*) site reveals a tripartite mechanism for vertebrate poly(*a*) site recognition. Genes Dev. **19:** 1315–1327

61 Zhao J., Hyman L. and Moore C. (1999) Formation of mRNA 3′ ends in eukaryotes: mechanism regulation and interrelationships with other steps in mRNA synthesis. Microbiol. Mol. Biol. Rev. **63:** 405–445

62 Takagaki Y. and Manley J. L. (2000) Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol. Cell. Biol. **20:** 1515–1525

63 Proudfoot N. (2004) New perspectives on connecting messenger RNA 3′ end formation to transcription. Curr. Opin. Cell Biol. **16:** 272–278

64 Wahle E. and Ruegsegger U. (1999) 3′-End processing of pre-mRNA in eukaryotes. FEMS Microbiol. Rev. **23:** 277–295

65 Marzluff W. F. (2005) Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. Curr. Opin. Cell Biol. **17:** 274–280

66 Ryan K., Calvo O. and Manley J. L. (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3′ processing endonuclease. RNA **10:** 565–573

67 Dominski Z., Yang X. C. and Marzluff W. F. (2005) The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. Cell **123:** 37–48

68 Berget S. M. (1995) Exon recognition in vertebrate splicing. J. Biol. Chem. **270:** 2411–2414

69 Graveley B. R., Hertel K. J. and Maniatis T. (1999) SR proteins are 'locators' of the RNA splicing machinery. Curr. Biol. **9:** R6–R7

70 Smith C. W. and Valcarcel J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem. Sci. **25:** 381–388

71 Singh R. and Valcarcel J. (2005) Building specificity with nonspecific RNA-binding proteins. Nat. Struct. Mol. Biol. **12:** 645–653

72 Graveley B. R. (2000) Sorting out the complexity of SR protein functions. RNA **6:** 1197–1211

73 Blencowe B. J. (2000) Exonic splicing enhancers: mechanism of action diversity and role in human genetic diseases. Trends Biochem. Sci. **25:** 106–110

74 Shen H., Kan J. L. and Green M. R. (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol. Cell **13:** 367–376

75 Wang Z., Rolish M. E., Yeo G., Tung V., Mawson M. and Burge C. B. (2004) Systematic identification and analysis of exonic splicing silencers. Cell **119:** 831–845

76 Zheng Z. M. (2004) Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. J. Biomed. Sci. **11:** 278–294

77 Fairbrother W. G., Yeh R. F., Sharp P. A. and Burge C. B. (2002) Predictive identification of exonic splicing enhancers in human genes. Science **297:** 1007–1013

78 Tacke R. and Manley J. L. (1999) Determinants of SR protein specificity. Curr. Opin. Cell Biol. **11:** 358–362

79 Amendt B. A., Si Z. H. and Stoltzfus C. M. (1995) Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. Mol. Cell. Biol. **15:** 4606–4615

80 Staffa A. and Cochrane A. (1995) Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. Mol. Cell. Biol. **15:** 4597–4605

81 Pozzoli U. and Sironi M. (2005) Silencers regulate both constitutive and alternative splicing events in mammals. Cell. Mol. Life Sci. **62:** 1579–1604

82 Mount S. M., Burks C., Hertz G., Stormo G. D., White O. and Fields C. (1992) Splicing signals in Drosophila: intron size information content and consensus sequences. Nucleic Acids Res. **20:** 4255–4262

83 Niwa M. and Berget S. M. (1991) Mutation of the AAUAAA polyadenylation signal depresses *in vitro* splicing of proximal but not distal introns. Genes Dev. **5:** 2086–2095

84 Niwa M., Rose S. D. and Berget S. M. (1990) *In vitro* polyadenylation is stimulated by the presence of an upstream intron. Genes Dev. **4:** 1552–1559

85 Niwa M. and Berget S. M. (1991) Polyadenylation precedes splicing *in vitro*. Gene Expr. **1:** 5–14

86 Nesic D., Cheng J. and Maquat L. E. (1993) Sequences within the last intron function in RNA 3′-end formation in cultured cells. Mol. Cell. Biol. **13:** 3359–3369

87 Dye M. J. and Proudfoot N. J. (1999) Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. Mol. Cell **3:** 371–378

88 Vagner S., Vagner C. and Mattaj I. W. (2000) The carboxyl terminus of vertebrate poly(*a*) polymerase interacts with U2AF 65 to couple 3′-end processing and splicing. Genes Dev. **14:** 403–413

89 Millevoi S., Geraghty F., Idowu B., Tam J. L., Antoniou M. and Vagner S. (2002) A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3′-end processing. EMBO Rep. **3:** 869–874

90 Lutz C. S., Murthy K. G., Schek N., O'Connor J. P., Manley J. L. and Alwine J. C. (1996) Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency *in vitro*. Genes Dev. **10:** 325–337

91 Ashe M. P., Furger, A. and Proudfoot, N. J. (2000) Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. RNA **6:** 170–177

92 Vagner S., Ruegsegger U., Gunderson S. I., Keller W. and Mattaj I. W. (2000) Position-dependent inhibition of the cleavage step of pre-mRNA 3′-end processing by U1 snRNP. RNA **6:** 178–188

93 Dreyfuss G., Matunis M. J., Pinol-Roma S. and Burd C. G. (1993) hnRNP proteins and the biogenesis of mRNA. Annu. Rev. Biochem. **62:** 289–321

94 Krecic A. M. and Swanson M. S. (1999) hnRNP complexes: composition structure and function. Curr. Opin. Cell Biol. **11:** 363–371

95 McAfee J., Huang M., Soltaninassab S., Rech J., Iyengar S. and LeStourgeon, W. (1997) The packaging of pre-mRNA. In: Eukaryotic mRNA Processing, pp. 68–102, Krainer A. (ed.), IRL Press at Oxford University Press, Oxford

96 Weighardt F., Biamonti G. and Riva S. (1996) The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. Bioessays **18:** 747–756

97 Anantharaman V., Koonin E. V. and Aravind L. (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res. **30:** 1427–1464

98 Burd C. G. and Dreyfuss G. (1994) Conserved structures and diversity of functions of RNA binding proteins. Science **265:** 615–621

99 Varani G. and Nagai K. (1998) RNA recognition by RNP proteins during RNA processing. Annu. Rev. Biophys. Biomol. Struct. **27:** 407–445

100 Antson A. A. (2000) Single-stranded-RNA binding proteins. Curr. Opin. Struct. Biol. **10:** 87–94

101 Hall K. B. (2002) RNA-protein interactions. Curr. Opin. Struct. Biol. **12:** 283–288

102 Shamoo Y., Abdul-Manan N. and Williams K. R. (1995) Multiple RNA binding domains (RBDs) just don't add up. Nucleic Acids Res. **23:** 725–728

103 Park-Lee S., Kim S. and Laird-Offringa I. A. (2003) Characterization of the interaction between neuronal RNA-binding protein HuD and AU-rich RNA. J. Biol. Chem. **278:** 39801–39808

104 Banerjee H., Rahn A., Davis W. and Singh R. (2003) Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. RNA **9:** 88–99

105 Lopez de Silanes I., Zhan M., Lal A.,Yang X. and Gorospe M. (2004) Identification of a target RNA motif for RNA-

binding protein HuR. Proc. Natl. Acad. Sci. USA **101:** 2987–8792
106 Ryder S. P., Frater L. A., Abramovitz D. L., Goodwin E. B. and Williamson J. R. (2004) RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. Nat. Struct. Mol. Biol. **11:** 20–28
107 Minovitsky S., Gee S. L., Schokrpur S., Dubchak I. and Conboy J. G. (2005) The splicing regulatory element UGCAUG is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. Nucleic Acids Res. **33:** 714–724
108 Gao F. B., Carson C. C., Levine T. and Keene J. D. (1994) Selection of a subset of mRNAs from combinatorial 3′ untranslated region libraries using neuronal RNA-binding protein Hel-N1. Proc. Natl. Acad. Sci. USA **91:** 11207–11211
109 Tenenbaum S. A., Carson C. C., Lager P. J. and Keene J. D. (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc. Natl. Acad. Sci. USA **97:** 14085–14090
110 Ule J., Jensen K. B., Ruggiu M., Mele A., Ule A. and Darnell R. B. (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science **302:** 1212–1215
111 Blanchette M., Green R. E., Brenner S. E. and Rio D. C. (2005) Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. Genes Dev. **19:** 1306–1314
112 Giot L., Bader J. S., Brouwer C. et al. (2003) A protein interaction map of Drosophila melanogaster. Science **302:** 1727–1736
113 Soller M. and White K. (2005) ELAV multimerizes on conserved $AU_{4-6}$ motifs important for *ewg* splicing regulation. Mol. Cell. Biol. **25:** 7580–7591
114 Tange T. O., Damgaard C. K., Guth S., Valcarcel J. and Kjems J. (2001) The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. EMBO J. **20:** 5748–5758
115 Zhu J., Mayeda A. and Krainer A. R. (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol. Cell **8:** 1351–1361
116 Damgaard C. K., Tange T. O. and Kjems J. (2002) hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. RNA **8:** 1401–1415
117 Soller M. and White K. (2003) ELAV inhibits 3′-end processing to promote neural splicing of ewg pre-mRNA. Genes Dev. **17:** 2526–2538
118 Labourier E., Adams M. D. and Rio D. C. (2001) Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. Mol. Cell **8:** 363–373
119 Markovtsov V., Nikolic J. M., Goldman J. A., Turck C. W., Chou M. Y. and Black D. L. (2000) Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. Mol. Cell. Biol. **20:** 7463–7479
120 Sharma S., Falick A. M. and Black D. L. (2005) Polypyrimidine tract binding protein blocks the 5′ splice site-dependent assembly of U2AF and the prespliceosomal E complex. Mol. Cell **19:** 485–496
121 Izquierdo J. M., Majos N., Bonnal S., Martinez C., Castelo R., Guigo R. et al. (2005) Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. Mol. Cell **19:** 475–484
122 Rocak S. and Linder P. (2004) DEAD-box proteins: the driving forces behind RNA metabolism. Nat. Rev. Mol. Cell Biol. **5:** 232–241
123 Fairman M. E., Maroney P. A., Wang W., Bowers H. A., Gollnick P., Nilsen T. W. et al. (2004) Protein displacement by DExH/D 'RNA helicases' without duplex unwinding. Science **304:** 730–734
124 Abdelhaleem M. (2005) RNA helicases: regulators of differentiation. Clin. Biochem. **38:** 499–503
125 Honig A., Auboeuf D., Parker M. M., O'Malley B. W. and Berget S. M. (2002) Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. Mol. Cell. Biol. **22:** 5698–5707
126 Weidenhammer E. M., Ruiz-Noriega M. and Woolford J. L. Jr. (1997) Prp31p promotes the association of the U4/U6 x U5 tri-snRNP with prespliceosomes to form spliceosomes in Saccharomyces cerevisiae. Mol. Cell. Biol. **17:** 3580–3588
127 Blencowe B. J., Bauren G., Eldridge A. G., Issner R., Nickerson J. A., Rosonina E. et al. (2000) The SRm160/300 splicing coactivator subunits. RNA **6:** 111–120
128 Barnard D. C., Ryan K., Manley J. L. and Richter J. D. (2004) Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. Cell **119:** 641–651
129 Neugebauer K. M. (2002) On the importance of being co-transcriptional. J. Cell Sci. **115:** 3865–3871
130 Black D. L. (2003) Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem. **72:** 291–336
131 Cartegni L., Chew S. L. and Krainer A. R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat. Rev. Genet. **3:** 285–298
132 Sorek R., Lev-Maor G., Reznik M., Dagan T., Belinky F., Graur D. and Ast G. (2004) Minimal conditions for exonization of intronic sequences: 5′ splice site formation in alu exons. Mol. Cell **14:** 221–231
133 Caceres J. F., Stamm S., Helfman D. M. and Krainer A. R. (1994) Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. Science **265:** 1706–1709
134 Chen C. D., Kobayashi R. and Helfman D. M. (1999) Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. Genes Dev. **13:** 593–606
135 Zahler A. M., Damgaard C. K., Kjems J. and Caputi M. (2004) SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. J. Biol. Chem. **279:** 10077–11084
136 Kashima T. and Manley J. L. (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. Nat. Genet. **34:** 460–463
137 Cartegni L. and Krainer A. R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat. Genet. **30:** 377–384
138 Wagner E. J. and Garcia-Blanco M. A. (2001) Polypyrimidine tract binding protein antagonizes exon definition. Mol. Cell. Biol. **21:** 3281–3288
139 Wagner E. J. and Garcia-Blanco M. A. (2002) RNAi-mediated PTB depletion leads to enhanced exon definition. Mol. Cell **10:** 943–949
140 Castelo-Branco P., Furger A., Wollerton M., Smith C., Moreira A. and Proudfoot N. (2004) Polypyrimidine tract binding protein modulates efficiency of polyadenylation. Mol. Cell. Biol. **24:** 4174–4183
141 Kanopka A., Muhlemann O. and Akusjarvi G. (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. Nature **381:** 535–538
142 Han K., Yeo G., An P., Burge C. B. and Grabowski P. J. (2005) A combinatorial code for splicing silencing: UAGG and GGGG motifs. PLoS Biol. **3:** e158
143 Philipps D. L., Park J. W. and Graveley B. R. (2004) A computational and experimental approach toward a priori identification of alternatively spliced exons. RNA **10:** 1838–1844
144 Graveley B. R. (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell **123:** 65–73
145 Graveley B. R., Kaur A., Gunning D., Zipursky S. L., Rowen L. and Clemens J. C. (2004) The organization and evolution of

the dipteran and hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. RNA **10:** 1499–1506
146 Yeo G. W., Van Nostrand E., Holste D., Poggio T. and Burge C. B. (2005) Identification and analysis of alternative splicing events conserved in human and mouse. Proc. Natl. Acad. Sci. USA **102:** 2850–2855
147 Dror G., Sorek R. and Shamir R. (2005) Accurate identification of alternatively spliced exons using support vector machine. Bioinformatics **21:** 897–901
148 Maniatis T. and Tasic B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature **418:** 236–243
149 Lopez A. J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. Annu. Rev. Genet. **32:** 279–305
150 Matlin A. J., Clark F. and Smith C. W. (2005) Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol. **6:** 386–398
151 Ladd A. N., Charlet N. and Cooper T. A. (2001) The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol. Cell. Biol. **21:** 1285–1296
152 Charlet B. N., Logan P., Singh G. and Cooper T. A. (2002) Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. Mol. Cell **9:** 649–658
153 Ladd A. N., Taffet G., Hartley C., Kearney D. L. and Cooper T. A. (2005) Cardiac tissue-specific repression of CELF activity disrupts alternative splicing and causes cardiomyopathy. Mol. Cell. Biol. **25:** 6267–6278
154 Jin Y., Suzuki H., Maegawa S., Endo H., Sugano S., Hashimoto K. et al. (2003) A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. EMBO J. **22:** 905–912
155 Gromak N., Matlin A. J., Cooper T. A. and Smith C. W. (2003) Antagonistic regulation of alpha-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein. RNA **9:** 443–456
156 Gromak N., Rideau A., Southby J., Scadden A. D., Gooding C., Huttelmaier S. et al. (2003) The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing. EMBO J. **22:** 6356–6364
157 Zhang W., Liu H., Han K. and Grabowski P. J. (2002) Region-specific alternative splicing in the nervous system: implications for regulation by the RNA-binding protein NAPOR. RNA **8:** 671–685
158 Dredge B. K. and Darnell R. B. (2003) Nova regulates GABA(*a*) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. Mol. Cell. Biol. **23:** 4687–4700
159 Dredge B. K., Stefani G., Engelhard C. C. and Darnell R. B. (2005) Nova autoregulation reveals dual functions in neuronal splicing. EMBO J. **24:** 1608–1620
160 Schutt C. and Nothiger R. (2000) Structure function and evolution of sex-determining systems in Dipteran insects. Development **127:** 667–677
161 Beckmann K., Grskovic M., Gebauer F. and Hentze M. W. (2005) A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in Drosophila. Cell **122:** 529–540
162 Demir E. and Dickson B. J. (2005) fruitless splicing specifies male courtship behavior in Drosophila. Cell **121:** 785–794
163 Usui-Aoki K., Ito H., Ui-Tei K., Takahashi K., Lukacsovich T., Awano W. et al. (2000) Formation of the male-specific muscle in female Drosophila by ectopic fruitless expression. Nat. Cell. Biol. **2:** 500–506
164 Manoli D. S., Foss M., Villella A., Taylor B. J., Hall J. C. and Baker B. S. (2005) Male-specific fruitless specifies the neural substrates of Drosophila courtship behaviour. Nature **436:** 395–400
165 Labourier E., Blanchette M., Feiger J. W., Adams M. D. and Rio D. C. (2002) The KH-type RNA-binding protein PSI is required for Drosophila viability male fertility and cellular mRNA processing. Genes Dev. **16:** 72–84
166 Horabin J. I. and Schedl P. (1993) Regulated splicing of the Drosophila sex-lethal male exon involves a blockage mechanism. Mol. Cell. Biol. **13:** 1408–1414
167 Sakamoto H., Inoue K., Higuchi I., Ono Y. and Shimura Y. (1992) Control of Drosophila Sex-lethal pre-mRNA splicing by its own female- specific product. Nucleic Acids Res. **20:** 5533–5540
168 Wang J. and Bell L. R. (1994) The Sex-lethal amino terminus mediates cooperative interactions in RNA binding and is essential for splicing regulation. Genes Dev. **8:** 2072–2085
169 Lallena M. J., Chalmers K. J., Llamazares S., Lamond A. I. and Valcarcel J. (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45. Cell **109:** 285–296
170 Nagengast A. A., Stitzinger S. M., Tseng C. H., Mount S. M. and Salz H. K. (2003) Sex-lethal splicing autoregulation *in vivo*: interactions between SEX-LETHAL the U1 snRNP and U2AF underlie male exon skipping. Development **130:** 463–471
171 Lisbin M. J., Qiu J. and White K. (2001) The neuron-specific RNA-binding protein ELAV regulates neuroglian alternative splicing in neurons and binds directly to its pre-mRNA. Genes Dev. **15:** 2546–2561
172 Park J. W., Parisky K., Celotto A. M., Reenan R. A. and Graveley B. R. (2004) Identification of alternative splicing regulators by RNA interference in Drosophila. Proc. Natl. Acad. Sci. USA **101:** 15974–15979
173 Kumar S. and Lopez A. J. (2005) Negative feedback regulation among SR splicing factors encoded by Rbp1 and Rbp1-like in Drosophila. EMBO J. **24:** 2646–2655
174 Blanchette M. and Chabot B. (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. EMBO J. **18:** 1939–1952
175 Samson M. L. (1998) Evidence for 3′ untranslated region-dependent autoregulation of the Drosophila gene encoding the neuronal nuclear RNA-binding protein ELAV. Genetics **150:** 723–733
176 Gunderson S. I., Beyer K., Martin G., Keller W., Boelens W. C. and Mattaj L. W. (1994) The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(*a*) polymerase. Cell **76:** 531–541
177 Sureau A., Gattoni R., Dooghe Y., Stevenin J. and Soret J. (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. EMBO J. **20:** 1785–1796
178 Wollerton M. C., Gooding C., Wagner E. J., Garcia-Blanco M. A. and Smith C. W. (2004) Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol. Cell **13:** 91–100
179 Le Guiner C., Lejeune F., Galiana D., Kister L., Breathnach R., Stevenin J. et al. (2001) TIA-1 and TIAR activate splicing of alternative exons with weak 5′ splice sites followed by a U-rich stretch on their own pre-mRNAs. J. Biol. Chem. **276:** 40638–40646
180 Kraus M. E. and Lis J. T. (1994) The concentration of B52 an essential splicing factor and regulator of splice site choice *in vitro* is critical for Drosophila development. Mol. Cell. Biol. **14:** 5360–5370
181 Labourier E., Bourbon H. M., Gallouzi I. E., Fostier M., Allemand E. and Tazi J. (1999) Antagonism between RSF1 and SR proteins for both splice-site recognition *in vitro* and Drosophila development. Genes Dev. **13:** 740–753
182 Edwalds-Gilbert G., Veraldi K. L. and Milcarek C. (1997) Alternative poly(*a*) site selection in complex transcription units: means to an end? Nucleic Acids Res. **25:** 2547–2561

183 Peterson M. L. (1994) RNA processing and expression of immunoglobulin genes. In: Handbook of B and T lymphocytes, pp. 321–342, Snow E. C. (ed.), Academic Press, San Diego
184 Edwalds-Gilbert G. and Milcarek C. (1995) Regulation of poly(*a*) site use during mouse B-cell development involves a change in the binding of a general polyadenylation factor in a B-cell stage-specific manner. Mol. Cell. Biol. **15:** 6420–6429
185 Takagaki Y., Seipelt R. L., Peterson M. L. and Manley J. L. (1996) The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell **87:** 941–952
186 Takagaki Y. and Manley J. L. (1998) Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. Mol. Cell **2:** 761–771
187 Veraldi K. L., Arhin G. K., Martincic K., Chung-Ganster L. H., Wilusz J. and Milcarek C. (2001) hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. Mol. Cell. Biol. **21:** 1228–1238
188 Phillips C., Pachikara N. and Gunderson S. I. (2004) U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(*a*) site by binding between the two downstream GU-rich regions. Mol. Cell. Biol. **24:** 6162–6171
189 Phillips C. and Gunderson S. (2003) Sequences adjacent to the 5′ splice site control U1A binding upstream of the IgM heavy chain secretory poly(*a*) site. J. Biol. Chem. **278:** 22102–22111
190 Phillips C., Jung S. and Gunderson S. I. (2001) Regulation of nuclear poly(*a*) addition controls the expression of immunoglobulin M secretory mRNA. EMBO J. **20:** 6443–6452
191 Gunderson S. I., Vagner S., Polycarpou-Schwarz M. and Mattaj I. W. (1997) Involvement of the carboxyl terminus of vertebrate poly(*a*) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. Genes Dev. **11:** 761–773
192 Gunderson S. I., Polycarpou-Schwarz M. and Mattaj I. W. (1998) U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(*a*) polymerase. Mol. Cell **1:** 255–264
193 Peterson M. L., Bertolino S. and Davis F. (2002) An RNA polymerase pause site is associated with the immunoglobulin mus poly(*a*) site. Mol. Cell. Biol. **22:** 5606–5615
194 Koushika S. P., Soller M. and White K. (2000) The neuron-enriched splicing pattern of Drosophila erect wing is dependent on the presence of ELAV protein. Mol. Cell. Biol. **20:** 1836–1845
195 Shin C. and Manley J. L. (2004) Cell signalling and the control of pre-mRNA splicing. Nat. Rev. Mol. Cell Biol. **5:** 727–738
196 Stamm S. (2002) Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. Hum. Mol. Genet. **11:** 2409–2416
197 Bedford M. T. and Richard S. (2005) Arginine methylation an emerging regulator of protein function. Mol. Cell **18:** 263–272
198 Kanopka A., Muhlemann O., Petersen-Mahrt S., Estmer C., Ohrmalm C. and Akusjarvi G. (1998) Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. Nature **393:** 185–187
199 Shin C., Feng Y. and Manley J. L. (2004) Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. Nature **427:** 553–558
200 Konig H., Ponta H. and Herrlich P. (1998) Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. EMBO J. **17:** 2904–2913
201 Weg-Remers S., Ponta H., Herrlich P. and Konig H. (2001) Regulation of alternative pre-mRNA splicing by the ERK MAP-kinase pathway. EMBO J. **20:** 4194–4203
202 Matter N., Herrlich P. and Konig H. (2002) Signal-dependent regulation of splicing via phosphorylation of Sam68. Nature **420:** 691–695
203 Habelhah H., Shah K., Huang L., Ostareck-Lederer A., Burlingame A. L., Shokat K. M. et al. (2001) ERK phosphorylation drives cytoplasmic accumulation of hnRNP-K and inhibition of mRNA translation. Nat. Cell. Biol. **3:** 325–330
204 Xie J. and Black D. L. (2001) A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels. Nature **410:** 936–939
205 Stallings-Mann M. L., Ludwiczak R. L., Klinger K. W. and Rottman F. (1996) Alternative splicing of exon 3 of the human growth hormone receptor is the result of an unusual genetic polymorphism. Proc. Natl. Acad. Sci. USA **93:** 12394–12399
206 Reugels A. M., Kurek R., Lammermann U. and Bunemann H. (2000) Mega-introns in the dynein gene DhDhc7(Y) on the heterochromatic Y chromosome give rise to the giant threads loops in primary spermatocytes of Drosophila hydei. Genetics **154:** 759–769
207 Misra S., Crosby M. A., Mungall C. J., Matthews B. B., Campbell K. S., Hradecky P. et al. (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. Genome Biol. **3:** RESEARCH0083
208 Reenan R. A., Hanrahan C. J. and Barry G. (2000) The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na+ channel transcript in a region of RNA editing. Neuron **25:** 139–149
209 Suter D., Tomasini R., Reber U., Gorman L., Kole R. and Schumperli D. (1999) Double-target antisense U7 snRNAs promote efficient skipping of an aberrant exon in three human beta-thalassemic mutations. Hum. Mol. Genet. **8:** 2415–2423
210 Dreyfuss G., Kim V. N. and Kataoka N. (2002) Messenger-RNA-binding proteins and the messages they carry. Nat. Rev. Mol. Cell Biol. **3:** 195–205
211 Chou M. Y., Underwood J. G., Nikolic J., Luu M. H. and Black D. L. (2000) Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. Mol. Cell **5:** 949–957
212 Amir-Ahmady B., Boutz P. L., Markovtsov V., Phillips M. L. and Black D. L. (2005) Exon repression by polypyrimidine tract binding protein. RNA **11:** 699–716
213 Hatton A. R., Subramaniam V. and Lopez A. J. (1998) Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. Mol. Cell **2:** 787–796
214 Burnette J. M., Miyamoto-Sato E., Schaub M. A., Conklin J. and Lopez A. J. (2005) Subdivision of large introns in Drosophila by recursive splicing at nonexonic elements. Genetics **170:** 661–674
215 Ott S., Tamada Y., Bannai H., Nakai K. and Miyano S. (2003) Intrasplicing – analysis of long intron sequences. Pac. Symp. Biocomput. 339–350
216 Lopez P. J. and Seraphin B. (2000) Uncoupling yeast intron recognition from transcription with recursive splicing. EMBO Rep. **1:** 334–339
217 Blumenthal T. and Gleason K. S. (2003) Caenorhabditis elegans operons: form and function. Nat. Rev. Genet. **4:** 112–120
218 Nilsen T. W. (2001) Evolutionarily origin of SL-addition trans-splicing: still an enigma. Trends Genet. **17:** 678–680
219 Krauss V. and Dorn R. (2004) Evolution of the trans-splicing Drosophila locus mod(mdg4) in several species of Diptera and Lepidoptera. Gene **331:** 165–176
220 Dorn R., Reuter G. and Loewendorf A. (2001) Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in Drosophila. Proc. Natl. Acad. Sci. USA **98:** 9724–9729
221 Gabler M., Volkmar M., Weinlich S., Herbst A., Dobberthien P., Sklarss S. et al. (2005) Trans-splicing of the mod(mdg4) complex locus is conserved between the distantly related

species Drosophila melanogaster and D. virilis. Genetics **169:** 723–736
222 Horiuchi T., Giniger E. and Aigaki T. (2003) Alternative trans-splicing of constant and variable exons of a Drosophila axon guidance gene lola. Genes Dev. **17:** 2496–2501
223 Tasic B., Nabholz C. E., Baldwin K. K., Kim Y., Rueckert E. H., Ribich S. A. et al. (2002) Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. Mol. Cell **10:** 21–33
224 Sage B. T. and Csink A. K. (2003) Heterochromatic self-association a determinant of nuclear organization does not require sequence homology in Drosophila. Genetics **165:** 1183–1193
225 Kim M., Krogan N. J., Vasiljeva L., Rando O. J., Nedea E., Greenblatt J. F. et al. (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature **432:** 517–522
226 Kaneko S. and Manley J. L. (2005) The mammalian RNA polymerase II C-terminal domain interacts with RNA to suppress transcription-coupled 3′ end formation. Mol. Cell **20:** 91–103
227 Kiss T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell **109:** 145–148
228 Huang Z. P., Zhou H., He H. L., Chen C. L., Liang D. and Qu L. H. (2005) Genome-wide analyses of two families of snoRNA genes from Drosophila melanogaster demonstrating the extensive utilization of introns for coding of snoRNAs. RNA **11:** 1303–1316
229 Hirose T. and Steitz J. A. (2001) Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. Proc. Natl. Acad. Sci. USA **98:** 12914–12919
230 Hirose T., Shu M. D. and Steitz J. A. (2003) Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. Mol. Cell **12:** 113–123
231 Wiemann S., Kolb-Kokocinski A. and Poustka A. (2005) Alternative pre-mRNA processing regulates cell-type specific expression of the IL4l1 and NUP62 genes. BMC Biol. **3:** 16
232 Proudfoot N. J., Furger A. and Dye M. J. (2002) Integrating mRNA processing with transcription. Cell **108:** 501–512
233 Calvo O. and Manley J. L. (2003) Strange bedfellows: polyadenylation factors at the promoter. Genes Dev. **17:** 1321–1327
234 Kornblihtt A. R., de la Mata M., Fededa J. P., Munoz M. J. and Nogues G. (2004) Multiple links between transcription and splicing. RNA **10:** 1489–1498
235 McCracken S., Fong N., Yankulov K., Ballantyne S., Pan G., Greenblatt J., Patterson S. D. et al. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature **385:** 357–361
236 Misteli T. and Spector D. L. (1999) RNA polymerase II targets pre-mRNA splicing factors to transcription sites *in vivo*. Mol. Cell **3:** 697–705
237 Bentley D. L. (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr. Opin. Cell Biol. **17:** 251–256
238 Ares M. Jr. and Proudfoot N. J. (2005) The spanish connection: transcription and mRNA processing get even closer. Cell **120:** 163–166
239 Fong N. and Bentley D. L. (2001) Capping splicing and 3′ processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev. **15:** 1783–1795
240 Fong N., Bird G., Vigneron M. and Bentley D. L. (2003) A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription splicing and 3′ end processing. EMBO J. **22:** 4274–4282
241 Chapman R. D., Palancade B., Lang A., Bensaude O. and Eick D. (2004) The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability. Nucleic Acids Res. **32:** 35–44
242 Bird G., Zorio D. A. and Bentley D. L. (2004) RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3′-end formation. Mol. Cell. Biol. **24:** 8963–8969
243 Xu Y. X., Hirose Y., Zhou X. Z., Lu K. P. and Manley J. L. (2003) Pin1 modulates the structure and function of human RNA polymerase II. Genes Dev. **17:** 2765–2776
244 Hirose Y., Tacke R. and Manley J. L. (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. Genes Dev. **13:** 1234–1239
245 Hirose Y. and Manley J. L. (1998) RNA polymerase II is an essential mRNA polyadenylation factor. Nature **395:** 93–96
246 Zeng C. and Berget S. M. (2000) Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. Mol. Cell. Biol. **20:** 8290–8301
247 Cramer P., Pesce C. G., Baralle F. E. and Kornblihtt A. R. (1997) Functional association between promoter structure and transcript alternative splicing. Proc. Natl. Acad. Sci. USA **94:** 11456–11460
248 Fededa J. P., Petrillo E., Gelfand M. S., Neverov A. D., Kadener S., Nogues G. et al. (2005) A polar mechanism coordinates different regions of alternative splicing within a single gene. Mol. Cell **19:** 393–404
249 Howe K. J., Kane C. M. and Ares M. Jr. (2003) Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. RNA **9:** 993–1006
250 Lorincz M. C., Dickerson D. R., Schmitt M. and Groudine M. (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat. Struct. Mol. Biol. **11:** 1068–1075
251 Auboeuf D., Dowhan D. H., Kang Y. K., Larkin K., Lee J. W., Berget S. M. et al. (2004) Differential recruitment of nuclear receptor coactivators may determine alternative RNA splice site choice in target genes. Proc. Natl. Acad. Sci. USA **101:** 2270–2274
252 Auboeuf D., Honig A., Berget S. M. and O'Malley B. W. (2002) Coordinate regulation of transcription and splicing by steroid receptor coregulators. Science **298:** 416–419
253 Auboeuf D., Dowhan D. H., Li X., Larkin K., Ko L., Berget S. M. et al. (2004) CoAA a nuclear receptor coactivator protein at the interface of transcriptional coactivation and RNA splicing. Mol. Cell. Biol. **24:** 442–453
254 Dowhan D. H., Hong E. P., Auboeuf D., Dennis A. P., Wilson M. M., Berget S. M. et al. (2005) Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta. Mol. Cell **17:** 429–439
255 Rosonina E., Bakowski M. A., McCracken S. and Blencowe B. J. (2003) Transcriptional activators control splicing and 3′-end cleavage levels. J. Biol. Chem. **278:** 43034–43040
256 Daneholt B. (2001) Assembly and transport of a premessenger RNP particle. Proc. Natl. Acad. Sci. USA **98:** 7012–7017
257 Hieronymus H. and Silver P. A. (2004) A systems view of mRNP biology. Genes Dev. **18:** 2845–2860
258 Aguilera A. (2005) Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. Curr. Opin. Cell Biol. **17:** 242–250
259 Huertas P. and Aguilera A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol. Cell **12:** 711–721
260 Li X. and Manley J. L. (2005) Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability. Cell **122:** 365–378
261 Luna R., Jimeno S., Marin M., Huertas P., Garcia-Rubio M. and Aguilera A. (2005) Interdependence between transcrip-

tion and mRNP processing and export and its impact on genetic stability. Mol. Cell **18:** 711–722
262 Jensen T. H., Boulay J., Olesen J. R., Colin J., Weyler M. and Libri D. (2004) Modulation of transcription affects mRNP quality. Mol. Cell **16:** 235–244
263 Muhlemann O., Mock-Casagrande C. S., Wang J., Li S., Custodio N., Carmo-Fonseca M. et al. (2001) Precursor RNAs harboring nonsense codons accumulate near the site of transcription. Mol. Cell **8:** 33–43
264 Hilleren P., McCarthy T., Rosbash M., Parker R. and Jensen T. H. (2001) Quality control of mRNA 3′-end processing is linked to the nuclear exosome. Nature **413:** 538–542
265 Maquat L. E. (2004) Nonsense-mediated mRNA decay: splicing translation and mRNP dynamics. Nat. Rev. Mol. Cell Biol. **5:** 89–99
266 Ishigaki Y., Li X., Serin G. and Maquat L. E. (2001) Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell **106:** 607–617
267 Tange T. O., Nott A. and Moore M. J. (2004) The ever-increasing complexities of the exon junction complex. Curr. Opin. Cell Biol. **16:** 279–284
268 Lewis B. P., Green R. E. and Brenner S. E. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc. Natl. Acad. Sci. USA **100:** 189–192
269 Lau P. P., Xiong W. J., Zhu H. J., Chen S. H. and Chan L. (1991) Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation. J. Biol. Chem. **266:** 20550–20554
270 Kuersten S. and Goodwin E. B. (2005) Linking nuclear mRNP assembly and cytoplasmic destiny. Biol. Cell. **97:** 469–478
271 Hachet O. and Ephrussi A. (2004) Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. Nature **428:** 959–963
272 Kaplan D. L. and O'Donnell M. (2003) Rho factor: transcription termination in four steps. Curr. Biol. **13:** R714–R716
273 Furger A., O'Sullivan J. M., Binnie A., Lee B. A. and Proudfoot N. J. (2002) Promoter proximal splice sites enhance transcription. Genes Dev. **16:** 2792–2799
274 Dichtl B., Blank D., Ohnacker M., Friedlein A., Roeder D., Langen H. et al. (2002) A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. Mol. Cell **10:** 1139–1150
275 Krishnamurthy S., He X., Reyes-Reyes M., Moore C. and Hampsey M. (2004) Ssu72 Is an RNA polymerase II CTD phosphatase. Mol. Cell **14:** 387–394
276 He X., Khan A. U., Cheng H., Pappas D. L. Jr., Hampsey M. and Moore C. L. (2003) Functional interactions between the transcription and mRNA 3′ end processing machineries mediated by Ssu72 and Sub1. Genes Dev. **17:** 1030–1042
277 Ahn S. H., Kim M. and Buratowski S. (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. Mol. Cell **13:** 67–76
278 Ni Z., Schwartz B. E., Werner J., Suarez J. R. and Lis J. T. (2004) Coordination of transcription RNA processing and surveillance by P-TEFb kinase on heat shock genes. Mol. Cell **13:** 55–65
279 Enriquez-Harris P., Levitt N., Briggs D. and Proudfoot N. J. (1991) A pause site for RNA polymerase II is associated with termination of transcription. EMBO J. **10:** 1833–1842
280 Zhang Z., Fu J. and Gilmour D. S. (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor Pcf11. Genes Dev. **19:** 1572–1580
281 Lee H., Kraus K. W., Wolfner M. F. and Lis J. T. (1992) DNA sequence requirements for generating paused polymerase at the start of hsp70. Genes Dev. **6:** 284–295
282 Alen C., Kent N. A., Jones H. S., O'Sullivan J., Aranda A. and Proudfoot N. J. (2002) A role for chromatin remodeling in transcriptional termination by RNA polymerase II. Mol. Cell **10:** 1441–1452
283 West S., Gromak N. and Proudfoot N. J. (2004) Human 5′→3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature **432:** 522–525
284 Dye M. J. and Proudfoot N. J. (2001) Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. Cell **105:** 669–681
285 Teixeira A., Tahiri-Alaoui A., West S., Thomas B., Ramadass A., Martianov I. et al. (2004) Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. Nature **432:** 526–530
286 Yelin R., Dahary D., Sorek R., Levanon E. Y., Goldstein O., Shoshan A. et al. (2003) Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. **21:** 379–386
287 Jen C. H., Michalopoulos I., Westhead D. R. and Meyer P. (2005) Natural antisense transcripts with coding capacity in Arabidopsis may have a regulatory role that is not linked to double-stranded RNA degradation. Genome Biol. **6:** R51
288 Bartel D. P. (2004) MicroRNAs: genomics biogenesis mechanism and function. Cell **116:** 281–297
289 Hastings M. L., Ingle H. A., Lazar M. A. and Munroe S. H. (2000) Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA. J. Biol. Chem. **275:** 11507–11513
290 Cawley S., Bekiranov S., Ng H. H., Kapranov P., Sekinger E. A., Kampa D. et al. (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell **116:** 499–509
291 Spector D. L. (2003) The dynamics of chromosome organization and gene regulation. Annu. Rev. Biochem. **72:** 573–608
292 Carmo-Fonseca M. (2002) The contribution of nuclear compartmentalization to gene regulation. Cell **108:** 513–521
293 Misteli T. (2005) Concepts in nuclear architecture. Bioessays **27:** 477–487
294 Oliver B. and Misteli T. (2005) A non-random walk through the genome. Genome Biol. **6:** 214
295 Olson M. O., Dundr M. and Szebeni A. (2000) The nucleolus: an old factory with unexpected capabilities. Trends Cell Biol. **10:** 189–196
296 Misteli T. (2001) The concept of self-organization in cellular architecture. J. Cell. Biol. **155:** 181–185
297 Cook P. R. (2002) Predicting three-dimensional genome structure from transcriptional activity. Nat. Genet. **32:** 347–352
298 Jackson D. A. (2003) The principles of nuclear structure. Chromosome Res. **11:** 387–401
299 Nayler O., Stratling W., Bourquin J. P., Stagljar I., Lindemann L., Jasper H. et al. (1998) SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. Nucleic Acids Res. **26:** 3542–3549
300 Patel A. A., McCarthy M. and Steitz J. A. (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression. EMBO J. **21:** 3804–3815
301 Otake L. R., Scamborova P., Hashimoto C. and Steitz J. A. (2002) The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in Drosophila. Mol. Cell **9:** 439–446
302 Kim M., Ahn S. H., Krogan N. J., Greenblatt J. F. and Buratowski S. (2004) Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. EMBO J. **23:** 354–364
303 O'Sullivan J. M., Tan-Wong S. M., Morillon A., Lee B., Coles J., Mellor J. et al. (2004) Gene loops juxtapose promoters and terminators in yeast. Nat. Genet. **36:** 1014–1018
304 Carter D., Chakalova L., Osborne C. S., Dai Y. F. and Fraser P. (2002) Long-range chromatin regulatory interactions *in vivo*. Nat. Genet. **32:** 623–626

305 Sawado T., Halow J., Bender M. A. and Groudine M. (2003) The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. Genes Dev. **17:** 1009–1018
306 Chambeyron S. and Bickmore W. A. (2004) Does looping and clustering in the nucleus regulate gene expression? Curr. Opin. Cell Biol. **16:** 256–262
307 Matzke M. A. and Birchler J. A. (2005) RNAi-mediated pathways in the nucleus. Nat. Rev. Genet. **6:** 24–35
308 Ule J., Ule A., Spencer J., Williams A., Hu J. S., Cline M. et al. (2005) Nova regulates brain-specific splicing to shape the synapse. Nat. Genet. **37:** 844–852
309 Huang C. S., Shi S. H., Ule J., Ruggiu M., Barker L. A., Darnell R. B. et al. (2005) Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. Cell **123:** 105–118
310 Keene J. D. and Tenenbaum S. A. (2002) Eukaryotic mRNPs may represent posttranscriptional operons. Mol. Cell **9:** 1161–1167
311 Andolfatto P. (2005) Adaptive evolution of non-coding DNA in *Drosophila*. Nature **437:** 1149–1152
312 Mattick J. S. (2004) RNA regulation: a new genetics? Nat. Rev. Genet. **5:** 316–323
313 Roberts G. C. and Smith C. W. (2002) Alternative splicing: combinatorial output from the genome. Curr. Opin. Chem. Biol. **6:** 375–283
314 Ponta H., Sherman L. and Herrlich P. A. (2003) CD44: from adhesion molecules to signalling regulators. Nat. Rev. Mol. Cell Biol. **4:** 33–45
315 Wetterberg I., Zhao J., Masich S., Wieslander L. and Skoglund U. (2001) In situ transcription and splicing in the Balbiani ring 3 gene. EMBO J. **20:** 2564–2574