

Pre-mRNA Processing in the Nuclear Landscape

Karla M. Neugebauer , Ph.D.*, Kimberly M. Kotovic, Jennifer A. Geiger and David
Stanek, Ph.D.

*corresponding author:

Max Planck Institute for Molecular Cell Biology and Genetics

Pfotenhauerstrasse 108

01307 Dresden

Germany

Tel +49-351-210-2589

Fax +49-351-210-1209

neugebauer@mpi-cbg.de

Table of Contents

1. Introduction
2. Co-transcriptional Pre-mRNA Processing
 - 2.1 RNA polymerase II transcription units
 - 2.2 5' end capping: coordinated by Pol II
 - 2.3 Pre-mRNA splicing: mixed messages
 - 2.4 3' end formation and mRNP release
 - 2.5 Histone 3' end formation
 - 2.6 Fine structure of the Transcription Unit
3. Nuclear Bodies and Pre-mRNA processing
 - 3.1 Nuclear bodies
 - 3.2 SnRNP trafficking to Cajal Bodies
 - 3.3 The sub-nuclear distribution of capping and polyadenylation factors
4. What is Nucleoplasm?
5. Concluding Remarks

1. Introduction

All eukaryotic protein-coding genes are transcribed by RNA polymerase II (pol II), and each mRNA is the product not only of transcription but a variety of pre-mRNA processing events. In humans, every pre-mRNA acquires a methyl-guanosine cap at its 5' end, and nearly every transcript is internally spliced and polyadenylated at its 3' end ¹. The purpose of this chapter is to place these three major pre-mRNA processing steps within the context of the three dimensional space of the cell nucleus. Because capping, splicing and polyadenylation at least begin during RNA synthesis, these reactions occur largely at sites of gene transcription, which are distributed throughout the nucleus and not localized to particular domains or sub-structures ^{2,3}. Splicing and polyadenylation often continue post-transcriptionally, most likely in the interchromatin space. In addition, pre-mRNA processing factors are components of a number of subnuclear structures, such as Cajal Bodies and Cleavage Bodies, suggesting that some functions related to pre-mRNA processing are compartmentalized within the nuclear landscape.

2. Co-transcriptional and Nucleoplasmic Pre-mRNA Processing

2.1 RNA polymerase II transcription units

The term “transcription unit” (TU) describes the smallest unit of gene expression independently expressed by any RNA polymerase. Sometimes multiple genes are contained within a single TU, so the activities of the polymerase and other factors at any given TU include simply the production of the encoded pre-mRNA and, as we will see, RNA processing events which may occur simultaneously. From the point of view of RNA polymerase II (Figure 1A), the transcription process includes pre-initiation complex formation at the promoter, transcription initiation, elongation, termination, and polymerase dissociation from the DNA template. From the point of view of the transcript, pre-mRNA processing includes *i*) 5' end capping, in which the 5' triphosphate of the pre-mRNA is cleaved, and a guanosine monophosphate is added and subsequently methylated to produce

m⁷GpppN, *ii*) editing, in which individual RNA residues are converted to alternative bases (e.g. adenosine to inosine by base deamination) to produce mRNAs encoding distinct protein products, *iii*) splicing, in which introns are removed and exons ligated together by the spliceosome, *iv*) 3' end formation, involving pre-mRNA cleavage and synthesis of the poly(A) tail, and, paradoxically, *v*) degradation. *A priori*, each of these modifications might occur independently of the others, since most can occur in *in vitro* reconstituted systems on purified pre-mRNA substrates.

Over the last two decades, it has become clear that nascent pol II transcripts are pre-mRNA processing substrates, suggesting that transcription units are also RNA processing units ⁴. For example, during its synthesis, pre-mRNA shortening due to intron removal by the spliceosome has been directly visualized ^{5,6}. Such an event is considered “co-transcriptional”, because it occurs before RNA synthesis is complete and while the nascent RNA is still tethered to the DNA by the polymerase. Each of the pre-mRNA processing events described above (with the exception of editing) has been shown to be co-transcriptional at least some of the time, raising the possibility that these chemical reactions are co-regulated. Importantly, a number of trans-acting factors required for pre-mRNA processing directly bind to pol II, pol II has a stimulatory effect on processing, and in some cases, processing feeds back to polymerase activity (see below). This has led to the proposal that transcription and processing occur in a “gene expression factory” composed of machines linked together for the purpose of efficiency and regulation ⁷⁻¹⁰. However, some reactions may occur during transcription, simply because they are relatively fast compared to the time it takes to transcribe the gene to its end. Some reactions clearly continue post-transcriptionally.

2.2 5' end capping: coordinated by Pol II

RNA polymerase II specifies 5' capping of mRNA by binding to and recruiting all three capping activities to transcription units ^{11,12}. This explains why only RNAs transcribed by pol II are capped at their 5' ends. Two proteins in humans and three in yeast are responsible for

the triphosphatase, guanylyltransferase and methyltransferase activities. When RNA polymerase II transitions from initiation to processive elongation, the C-terminal domain (CTD) of the large subunit of pol II becomes hyperphosphorylated on Ser2 and Ser5 of the heptad YSPTSPS, which is repeated 26 times in yeast and 52 times in humans. The hyperphosphorylated form of the CTD has affinity for both human (Hce1 and Hcm1) and two of the three yeast factors (Ceg1 the guanylyltransferase and Abd1 the methyltransferase). Interestingly, in yeast the triphosphatase activity encoded by Cet1 binds to Ceg1 with two consequences: 1) Cet1, like Ceg1 and Abd1, becomes bound to the polymerase, and 2) Cet1 stimulates Ceg1 activity by an allosteric interaction^{13,14}. In Hce1, triphosphatase activity is dependent on an active guanylyltransferase domain, and guanylyltransferase activity is in turn stimulated by phosphorylation of Ser5 of the CTD to which it is bound¹⁵. Reflecting this link with initiation and the speed of the capping reactions, capping occurs when the nascent RNA is only 20-40 nucleotides long¹⁶. Consistent with the early occurrence of capping, all three enzymes are concentrated at the promoter-proximal regions of yeast TUs (Figure 1A), and Ceg1 and Cet1 dissociate from TUs downstream of the promoter due to CTD dephosphorylation during elongation^{17,18}. The recruitment of the capping enzymes by pol II and their regulation via CTD binding provide a complete explanation for the capping of pol II transcripts.

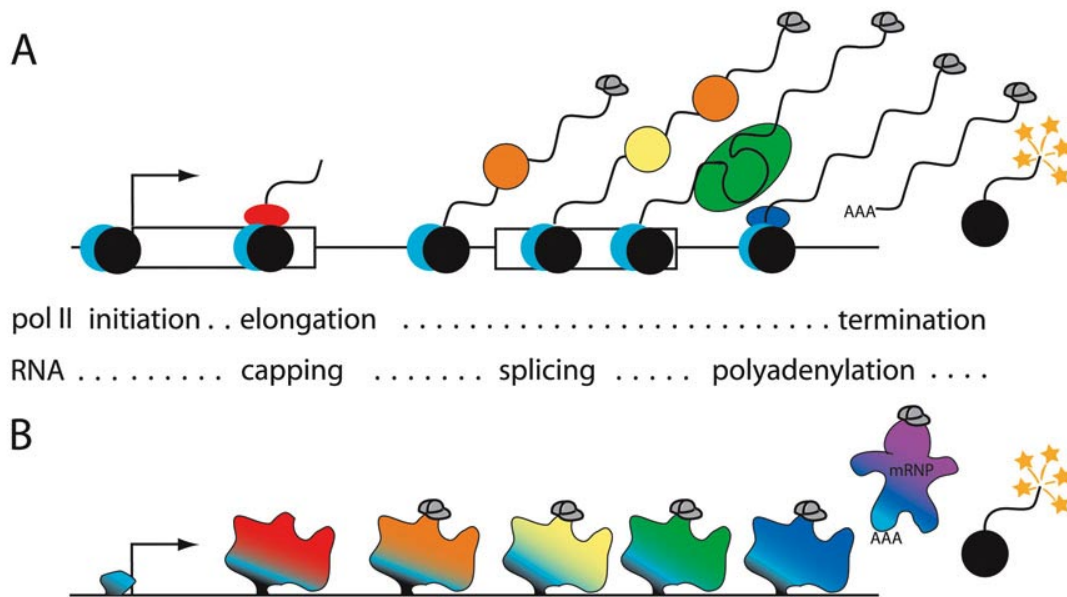


Figure 1. Co-transcriptional pre-mRNA processing

(A) Schematic representation of transcription and pre-mRNA processing events at pol II transcription units (TUs). Pol II (black ball) initiates transcription at the promoter (arrow) and proceeds along the TU during elongation phase, terminating and releasing from the DNA template following passage through the polyadenylation signals. Several polyadenylation factors, such as CPSF and CstF, bind directly to pol II and are shown all along the TU as a blue ball adjacent to the black one. Capping enzymes (red oval) are bound to pol II as it enters elongation phase and then fall off the TU. The 5' cap added by the capping enzymes is symbolized by the baseball cap. Because splicing is co-transcriptional, we have hypothetically placed splicing factors recognizing the 5' and 3' splice sites (orange and yellow balls, respectively) and the assembly of the spliceosome (green oval) within the body of the TU. Additional polyadenylation factors are recruited to downstream regions, as shown by the additional dark blue ball. At termination, pol II is released from the template and recycled, and the fragment of cleaved nascent RNA remaining will be degraded. The mRNP is released from the template and undergoes nuclear transport. (B) Hypothetical scheme for the fine structure of active transcription units based on the tomographic imaging of nascent transcript and splicing (NTS) complexes¹¹⁶. NTS complexes are tethered to the DNA axis by pol II (black shading) and contain processing factors required for the processing steps, which take place at different positions along the TU. NTS complexes along the TU are differently colored to reflect their heterogeneous composition of pre-mRNA processing factors.

The 5' cap modification itself renders pre-mRNA and mRNA resistant to the action of 5' to 3' exonucleases. In addition, the cap serves as a binding site for two important factors: the Cap Binding Complex (CBC) in the nucleus and the translation initiation factor eIF4E in the cytoplasm¹⁹. Like capping, CBC binding is co-transcriptional²⁰, but there is no evidence to date that recruitment of CBC to the cap requires any specific coupling to the transcription

machinery. CBC is composed of 2 subunits, CBP80 and CBP20, and plays a role in splicing of the first intron ²¹⁻²³, promotes the nucleocytoplasmic export of U snRNAs ²⁴, and supports a “pioneer round” of mRNA translation in the cytoplasm before CBC is exchanged for eIF4E ^{25,26}. Thus, the rapid and highly specific addition of the 5′ cap to pol II transcribed RNAs has important consequences for the lifetime of the (pre)-mRNA, and this cascade of events can be attributed to the initial interaction of the capping enzymes with pol II.

2.3 Pre-mRNA splicing: mixed messages

In pre-mRNA splicing, introns are removed and exons ligated together by a two-step transesterification reaction carried out by the spliceosome, a dynamic 60S ribonucleoprotein particle the size and molecular complexity of the ribosome ^{27,28}. The active spliceosome contains at least 70 polypeptides ²⁹, including the spliceosomal small nuclear ribonucleoprotein particles (snRNPs) and a number of non-snRNP splicing factors, and many additional non-snRNP splicing factors are essential for splicing even though they are not detectable in spliceosomes ^{30,31}. Formation of the spliceosome at particular splice junctions is triggered by recognition of the 5′ splice site by the U1 snRNP and the 3′ splice site by U2AF, followed by the U2 snRNP. It is unclear whether the spliceosome is assembled from larger complexes, such as the recently identified penta-snRNP containing U1, U2, U4, U5, and U6 snRNAs ³² or the 200S InRNP (large nuclear ribonucleoprotein particle) containing additional non-snRNP RNA processing factors ³³, or by the sequential addition of snRNP and non-snRNP factors as was previously supposed.

Pre-mRNA splicing begins co-transcriptionally and often continues post-transcriptionally, as exemplified by the Balbiani ring genes of *Chironomus tentans*, in which a high proportion of nascent RNAs lack introns at their 5′ ends but still contain terminal introns ^{34,35}. Co-transcriptional splicing has also been documented in *Drosophila* ^{5,6,36} and humans ^{37,38} and is likely to occur in yeast ³⁹. Moreover, snRNPs and non-snRNP splicing factors, such as SR proteins, are concentrated at active sites of transcription (Figure 2) ⁴⁰⁻⁴².

However, co-transcriptional splicing is not obligatory, and it may be important that some splicing events occur post-transcriptionally (see below).

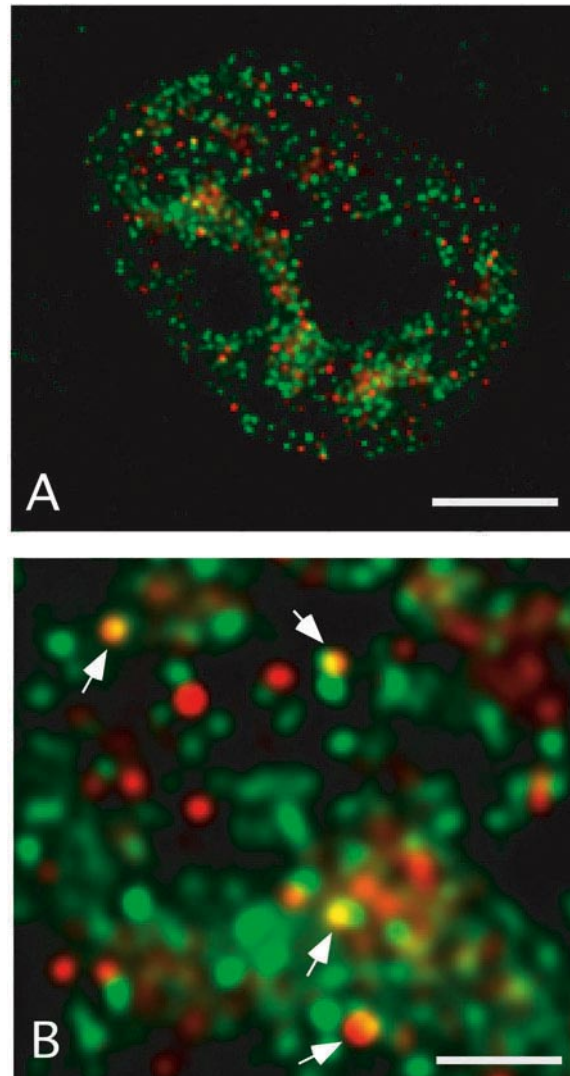


Figure 2. Co-localization of SR protein splicing factors with sites of mRNA synthesis

Double-staining of HeLa cells following run-on transcription in the presence of BrUTP with anti-SR (red) and rat anti-BrUTP (green). Panel A shows the entire nucleus; panel B shows an enlarged area of the same nucleus. Even though they are transcriptionally active, the nucleoli appear as dark circles, because under the conditions used here, they are impermeable to anti-Br-UTP. Many of the overlapping particles (yellow) appear identical in three-dimensional space (arrows). The lowest arrow points to a particle recognized by anti-SR that overlaps a differently shaped object detected by anti-BrUTP. Scale bars 4 μm in A and 1 μm in B.

Understanding how splicing is integrated with transcription is more complicated than the case of capping, because metazoan genes contain multiple introns (an average of 9 per gene in humans), which cannot serve as splicing substrates until both the 5' and 3' ends of each intron are synthesized. Thus, the time that it takes for pol II to synthesize each intron defines a minimal time and distance along the gene in which splicing factors can be recruited and spliceosomes formed. The time that it takes for pol II to reach the end of the TU defines the maximal time in which splicing could occur co-transcriptionally. In general, pol II moves along the DNA template at a rate of 1-1.5 kb/min. In humans, introns (avg. 3,300 bp) are ten times longer than exons (avg. 300 bp)¹ corresponding to ~3 min transcription time for introns and only ~30s for exons. RNP formation at 3' splice sites in *Drosophila* is observed 48s after 3' splice site synthesis with intron removal occurring 3 minutes later⁵. If these rates are similar in humans, then by the time the 3' splice site is recognized, the next exon may already be finished, and by the time splicing could occur 3 minutes later the next intron will have been completed. This opens up the possibility for competition among splice sites in alternative splicing. Indeed, intron removal does not always occur in the order of intron synthesis, indicating that some splicing events occur much more rapidly than others and that slower splicing events may occur post-transcriptionally in the nucleoplasm^{35,36}. Evidence for the interplay between transcription and splicing kinetics comes from experiments in humans and yeast, in which changes in transcription rate by introduction of transcriptional pause sites or the mutation of elongation factors results in the alternative selection of splice sites⁴³. The demonstration that transcriptional activators influence alternative splicing by modulating pol II elongation rates⁴⁴ provides physiological relevance for this kinetic relationship and suggests that alternative splicing *in vivo* may in part be due to transcriptional rather than splicing regulation per se. It will be of interest to learn whether members of an increasing number of trans-acting elongation factors also regulate splice site choice by a similar mechanism. Undoubtedly, a component of this type of regulation is the question of how much

time the nascent RNA has to bind trans-acting splicing factors before the next binding site or splice site is made.

Overlaid on this kinetic link between transcription and splicing is the distinct possibility that a physical link(s) may also exist. The pivotal observation is that the pol II CTD stimulates splicing in human cells, independent of its effects on capping or 3' end formation ⁴⁵. Addition of RNA polymerase II or the CTD alone also stimulates splicing *in vitro* ^{46,47}, but the molecular mechanism underlying this stimulation is unknown. While the search for such a link has focused on a proposed role for the CTD in directly binding to splicing factors ^{48,49}, to date the only bona fide splicing factor shown to bind the CTD *in vitro* is the yeast U1snRNP component Prp40p which has no known homologue in metazoans ⁵⁰. Although a search for direct binding partners of the CTD revealed a set of proteins containing arginine-rich domains similar to those present in non-snRNP splicing factors, it is noteworthy that splicing factors with demonstrated splicing activity were not detected in those assays ⁵¹. Within the Balbiani Ring genes, snRNP and non-snRNP splicing factors are concentrated in intron-rich regions and relatively depleted in regions lacking introns ⁵², suggesting that splicing factors do not travel with pol II within the TU. It is important to note that, in contrast to capping, splicing of at least some pre-mRNAs in fission and budding yeast can occur efficiently following synthesis by pol III ^{53,54}, T7 RNA polymerase ⁵⁵ or a CTD-less pol II ⁵⁶. Therefore, the stimulatory effect of the CTD on splicing may not be essential.

A recent study suggests that the effects of the CTD on splicing efficiency may be indirect, via an interaction of splicing snRNPs with pol II elongation factors ⁵⁷. This study shows that snRNPs as well as introns within the transcription template stimulate pol II elongation by the direct binding of snRNPs to the elongation factor TAT-SF1; TAT-SF1 in turn binds P-TEFb, which phosphorylates the CTD and remains associated with it during elongation ⁵⁷. One implication of this finding is that pol II elongation machinery might bring snRNPs to active genes. This may explain why a gene transcribed by CTD-less pol II fails to accumulate snRNPs or members of the SR protein family of non-snRNP splicing factors at the light microscopic level ⁴²; however, because the nascent RNA produced by the CTD-less

pol II most likely also lacks the 5' cap and CBC, these data remain open to other interpretations. Indeed, intronless genes transcribed by wild-type pol II fail to recruit SR proteins in similar assays, suggesting that the nascent RNA plays an important role in splicing factor recruitment^{58,59}. Importantly, if the CTD were pre-loaded with snRNPs directly or indirectly via P-TEFb/TAT-SF1, it would be difficult to understand how introns could further increase elongation rate. Taken together, the simplest explanation for this set of observations is that snRNP and TAT-SF1 recruitment to TUs is enhanced by the cooperative binding of snRNPs to splicing signals within the nascent RNA and of TAT-SF1 to P-TEFb. Additional alternatives to generic splicing factor recruitment by the CTD are provided by the findings that the SR protein family member SF2 binds directly to the transcriptional co-activator p52⁶⁰ and that alternative splicing can be influenced by promoter identity^{61,62}.

Despite the lack of evidence for direct binding of snRNP or non-snRNP splicing factors to the CTD, prevailing models of transcription and splicing coupling in the literature are based on it^{9,63}. The underlying logic of the model is that the crystal structure of pol II places the CTD at the exit groove of pol II from which the nascent RNA emerges⁶⁴, and placement of splicing factors at the outlet would promote the efficient recruitment of splicing factors to cognate RNA binding sites as they are made. However, splicing factors such as snRNPs and SR proteins are expressed at quite high concentrations in HeLa cell nuclei (1-10 μ M for U1, U2, U4, U5, and U6 snRNPs⁶⁵, and 10-100 μ M for the SR protein SF2^{66,67}). The affinity of at least one SR protein, SRp55, for its binding site in the alternatively spliced cTNT pre-mRNA is 60 nM⁶⁸, so a compelling argument for why further concentration of splicing factors would be advantageous has yet to be made. Additional open questions not addressed by the model include differences in splicing rates between introns, differences in the order of intron removal, and how alternative splicing could occur in the context of such recruitment directed by pol II. Finally, it is unclear whether all of the components of the spliceosome and/or every alternative splicing regulator should be positioned at every gene, or whether genes accumulate factors differentially to reflect their particular biosynthetic requirements.

While coupling between transcription and splicing can be important, it may be equally important for some transcripts that splicing continues post-transcriptionally. The *Drosophila* Ubx pre-mRNA, contains a 75kb intron which is recursively spliced, such that the first splicing event creates new splice sites which are subsequently recognized, and the intron is spliced again ⁶⁹. This chain of events could occur co-transcriptionally, but a strict coupling between splice site synthesis and splicing factor binding must be ruled out. A similar complication arises through RNA editing by the ADAR family of adenosine deaminases, because editing sites occur at splice junctions where intron sequences basepair with upstream exon sequences to produce a characteristic stem loop ⁷⁰. By definition, this must occur before splicing, and indeed editing can alter splice site sequences to produce alternative splicing ⁷¹. Thus, depending on the site and kinetics of editing, splicing of edited transcripts may be either co- or post-transcriptional. The proposal that alternative splicing may occur more slowly than constitutive splicing and result in the splicing of some introns post-transcriptionally ⁷² is supported by microscopic studies which detect slow-splicing introns away from the site of synthesis ⁷³⁻⁷⁵. The movement of (pre)-mRNA away from the gene is not thought to represent vectorial transport to the nuclear envelope, because the rates and trajectories of mRNP movement are consistent with diffusion ⁷⁶⁻⁸⁰; rather, the diffusion of such transcripts to the envelope may provide additional time for post-transcriptional splicing to occur.

2.4 3' end formation and mRNP release

All mRNAs, with the exception of the replication-dependent histone mRNAs, are modified by the addition of a tract of adenosine residues at their 3' ends ^{11,81,82}. The poly(A) tail is important for mRNA stability and the regulation of translation. Polyadenylation involves first the cleavage of the pre-mRNA at a site located between the canonical AAUAAA sequence where cleavage and polyadenylation specificity factor (CPSF) binds, and a downstream G/U-rich region where cleavage stimulatory factor (CstF) binds. Cleavage is accomplished by

cleavage factors I and II (CFI and CFII), and nuclear polyadenylation is accomplished by poly(A) polymerase (PAP) bound to CPSF and the nuclear poly(A) binding protein.

Transcription termination and release of pol II from the DNA template depends on transcription through a functional polyadenylation signal, which in humans can be up to 1500 bp upstream of the termination site ^{10,11,81}. This intimate relationship between termination and polyadenylation suggests that transcription is coupled with at least some of the steps leading to polyadenylation. In human cells, the CTD of pol II is specifically required for efficient polyadenylation ⁴⁵, and pol II itself stimulates polyadenylation *in vitro* ⁸³. Physical links to pol II are abundant, as several components are bound to the CTD (e.g. CPSF, cleavage/polyadenylation factor IA) and to other components of the pol II holoenzyme (e.g. CPSF to TFIID, CstF to the transcriptional coactivator PC4) ⁸⁴⁻⁸⁷. Thus, extensive protein-protein interactions among the polyadenylation factors themselves and with pol II may help to coordinate termination and polyadenylation. In *Chironomus*, these two events are temporally correlated ⁸⁸, and polyadenylation cleavage factors are required for efficient termination in yeast ⁸⁹. However, direct visualization of nascent transcripts in *Xenopus* and *Drosophila* show that cleavage often occurs after the release of pol II from the DNA ^{90,91}, suggesting that a substantial fraction of polyadenylation may occur post-transcriptionally.

In contrast to capping, polyadenylation is not solely specified by pol II. A small but significant set of pol II transcripts, such as histone mRNAs, snRNAs and snoRNAs are not polyadenylated and undergo alternative mechanisms of 3' end formation ¹⁰; likewise, rRNA, normally the product of pol I, is not polyadenylated when synthesized by pol II ⁹². Thus, polyadenylation targeting by pol II can at least be overridden by other processing signals. Along these lines, polyadenylation signals in the nascent yeast RNAs support partial polyadenylation of mRNAs transcribed by pol I, T7 RNA polymerase or a CTD-less pol II ^{55,93,94}, confirming that a strict coupling between pol II and polyadenylation is not required. Another case of modulation of polyadenylation function occurs in alternative terminal exon usage in which polyadenylation sites in upstream exons are not used in favor of those sites found in downstream alternative exons. This points to the importance of the strength of the

polyadenylation signals described above which can determine the rate of assembly of polyadenylation complexes on the nascent transcripts ⁹⁵, suggesting that assembly of the polyadenylation complexes at alternative terminal exons or unpolyadenylated genes may be relatively slow compared to the rates of splicing or alternative 3' end formation. Thus, signals in the nascent RNA play a defining role in where and whether the transcript is polyadenylated.

The interdependence of terminal intron splicing and polyadenylation ⁹⁶ and their temporal coincidence ⁸⁸ suggest a kinetic and/or physical link between the two processes. The splicing factor U2AF65 binds the polypyrimidine tract at all 3' splice sites where it promotes annealing of the U2 snRNA with the branchpoint. Interestingly, U2AF65 also binds to the C-terminus of PAP ⁹⁷, and this additional binding interaction likely promotes definition of the terminal exon for splicing and assembly of the polyadenylation machinery within the exon. The U1 snRNP binds at 5' splice sites and inhibits PAP, perhaps suppressing premature polyadenylation/termination in long introns or before the synthesis of the terminal exon ⁹⁸. In alternative terminal exon usage in the IgM pre-mRNA, the kinetics of polyadenylation likely plays a role, since elevated levels of CstF-64 in plasma cells promotes the recognition of the weaker upstream polyadenylation signal and precludes splicing to the downstream 3' splice site ⁹⁹. Conversely, the calcitonin/CGRP pre-mRNA undergoes alternative terminal exon usage through the action of a splicing factor SRp20, which promotes splicing and polyadenylation at upstream sites ¹⁰⁰. These physical and kinetic links between polyadenylation and splicing indicate that these two processes co-evolved. Because polyadenylation is linked with termination, interactions with the splicing machinery may, on the one hand, have put pressure on splicing to occur co-transcriptionally and, on the other hand, may have selected for splicing to occur slowly enough to permit assembly of downstream complexes on polyadenylation sites that might be otherwise spliced out too quickly.

Because the pre-mRNA travels with pol II significantly beyond the polyadenylation signal, there may be additional time for co-transcriptional processing of the nascent RNA before it is released from the TU. Several recent studies in yeast suggest a possible mRNA processing surveillance mechanism, which takes place at the TU prior to mRNP release. First, mutations in mRNA nuclear export factors lead to the retention of mRNAs at their sites of transcription, and these mRNAs become hyperadenylated ¹⁰¹. Second, (pre)-mRNAs which are cleaved but not polyadenylated due to a mutation in PAP also accumulate at TUs and can be released upon inactivation of components of the nuclear exosome ¹⁰² which are thought to mainly function as 3' to 5' exonucleases ¹⁰³. Interestingly, retention of transcripts aberrantly processed at their 3' ends does not depend on pol II, since retention also occurs when the transcript is synthesized by T7 RNA polymerase ⁵⁵. These findings suggest a novel function of components of the nuclear exosome in regulating mRNP release from the TU. Interestingly, a previous study implicated the exosome in monitoring pre-mRNA splicing ¹⁰⁴.

It is challenging to explain how mutations in export factors produce retention of transcripts at TUs. However, several lines of evidence link mRNA transport with transcription. First, pre-mRNA splicing deposits a set of proteins called the exon-junction complex (EJC) on mRNA, and this complex promotes the nucleocytoplasmic transport of the mRNP ¹⁰⁵. Second, even in the absence of splicing, two nuclear export factors in yeast, Yra1p and Sub2p, and their human counterparts, Aly and UAP56, are recruited to TUs via direct binding to the THO transcription elongation complex ^{106,107}. This evolutionarily conserved complex, named TREX for transcription/export complex, is detectable throughout the TU ¹⁰⁷, while Yra1p was detected in downstream regions of the TU in a separate study ¹⁰⁶. The co-transcriptional binding of these factors to nascent RNA raises the possibility of a feedback mechanism active at the TU. This link between RNA processing, mRNP release, and nuclear export is reminiscent of studies in human cells, showing that transcripts defective in splicing or polyadenylation are retained at the TU ¹⁰⁸⁻¹¹⁰. It remains to be determined how transcripts are retained at TUs and whether mRNP retention in humans depends on components of the nuclear exosome.

2.5 Histone 3' end formation

Unlike all other mRNAs, the replication-dependent histone mRNAs are cleaved at their 3' ends and are generally not polyadenylated. These mRNAs contain the Histone Downstream Element (HDE), a 26-nucleotide sequence including a 16-nucleotide stem loop, which bind stem loop binding protein (SLBP) and the U7 snRNP. These transacting factors guide the endonucleolytic cleavage of histone mRNA 3' ends through the ATP-independent action of an as yet unidentified cleavage factor ¹¹¹. SLBP remains associated with histone mRNPs and travels with them to the cytoplasm where it regulates translation and stability ^{112,113}. Although a separate class of histone pre-mRNAs are intron-containing, polyadenylated, and expressed throughout interphase, the replication-dependent histone genes are expressed specifically during S-phase, likely due to the elevated expression of SLBP at the G1-S boundary ¹¹⁴. Thus, the functions of histone 3' end formation are to co-ordinate histone gene expression with DNA replication, to prolong the half-life of histone mRNAs, and to promote translation. The HDE occurs 30 nucleotides downstream of the termination codon, and mRNA cleavage may signal transcription termination as well as histone mRNP release ¹¹⁵. Despite this possible connection to pol II transcription, there is currently no evidence for a physical link between histone 3' end formation and termination. In addition, the replication-dependent histone genes lack introns, and therefore, these pol II TUs are likely to differ significantly from the bulk of protein-coding TUs in both composition and regulation.

2.6 Fine structure of the Transcription Unit

What do transcription units look like? Classic electron microscopic images of pol II transcription units ^{5,6} reveal the typical “lampbrush” display of RNA as it becomes progressively longer with increasing distance from the promoter (see also chapter 6); additional electron dense blobs correlate with the positions of splice junctions. However, because nascent RNA is expected to have a high degree of secondary structure and to be

fairly coated with protein, one would expect nascent RNA *in vivo* to occur with associated proteins in large particles adjacent to the DNA axis rather than dangling from it. A recent electron tomographic study has indeed visualized such “nascent transcript and splicing” (NTS) complexes as regularly sized and shaped objects attached to the DNA ¹¹⁶. The NTS complexes were shown to contain at least RNA polymerase II and the U2 snRNP by immunolabelling. The imaging of the NTS complex combined with our current understanding of co-transcriptional RNA processing suggests that NTS complexes along the transcription unit may be heterogeneous, containing the factors which have assembled locally, via protein-protein interactions with pol II and/or protein-RNA binding (Figure 1B). For example, NTS complexes near the promoter are more likely to contain the capping enzymes, whereas downstream they may be enriched in splicing, polyadenylation and/or nuclear transport factors. This raises the question of whether the entire transcription unit should be considered a gene expression “factory” or whether each nascent transcript represents a distinct dynamic machine, in which each of the tools for mRNA synthesis and processing is recruited as needed by pol II and/or the nascent transcript. The latter model is a better fit to the existing experimental evidence, since there is currently no indication that NTS complexes at distinct positions along the TU influence the composition or activities of their neighbors.

3. Nuclear Bodies and Pre-mRNA processing

3.1 Nuclear Bodies

The cell nucleus does not contain any further membrane-bound organelles, but a number of substructures within the nucleus have been described. The term “nuclear body” describes a variety of distinct nuclear objects $\geq 0.5 \mu\text{m}$ in diameter, which are detectable by light and/or electron microscopy and enriched in specific sets of nuclear factors ¹¹⁷⁻¹¹⁹. The largest (up to $5.0 \mu\text{m}$ diameter in some cells) and best understood nuclear body is the nucleolus, which is the site of rDNA transcription by RNA polymerase I, rRNA processing, and assembly of pre-ribosomal subunits (chapter 8) ¹¹⁸. Because the nucleolus is not only a structural but a functional unit, the rationale for concentrating pol I transcription factors, rRNA processing

factors, and ribosomal proteins in nucleoli is clear. Many nuclear bodies contain pre-mRNA processing factors (Table I), although others not discussed here contain transcription factors. Because the pre-mRNA processing events described above occur co-transcriptionally or at least begin at the transcription unit, it is perhaps surprising that some processing factors are concentrated in nuclear bodies in addition to their nucleoplasmic distribution. Proposed functions for nuclear bodies in general include the storage of inactive factors and the assembly and/or recycling of multi-component RNA processing complexes. It is currently a challenge for cell biologists to distinguish among these possibilities.

3.2 snRNP trafficking to Cajal Bodies

Cajal Bodies were discovered ~100 years ago by Santiago Ramon y Cajal who named them accessory bodies, because of their position adjacent to nucleoli in the nuclei of central nervous system neurons ^{117,120}. Since their initial description, the structure and molecular composition of Cajal Bodies (CBs) has been explored, but their function remains obscure (chapter 10). Prominent components of CBs include the spliceosomal snRNPs (Figure 3A), the U7 snRNP and SLBP important in histone 3' end formation, as well as a variety of other components implicated in transcription and mRNA processing (Table I) ¹²¹. CBs are electron dense ~0.5 μm structures, and the identification of the protein p80-coilin as a molecular marker for CBs (previously known as coiled bodies) has facilitated studies of CBs in a variety of species and cell types. Live imaging studies have shown that CBs are highly mobile and are capable of fusing and splitting ^{122,123}. Most somatic cells contain one or two CBs, while some cells contain nucleoplasmic coilin but no CBs per se and transformed cell lines can contain many CBs. The nuclei of amphibian and insect oocytes contain 50-100 CBs. Because oocyte nuclei are thought to stockpile factors required for the early cleavage divisions of the embryo, one might suppose that the amplified numbers of CBs in oocytes reflect the need to store particular RNA processing factors. On the other hand, recent evidence suggests that some factors, such as the spliceosomal snRNPs, may transit through Cajal Bodies as a normal step in their assembly and maturation. If this is true, then the elevated numbers of CBs in

particular cells, such as transformed cell lines and/or oocytes, may reflect an elevated biosynthetic requirement for RNA processing factors in highly metabolically active cells. An additional possibility is that CBs function in some aspect(s) of snRNA precursor processing and/or histone mRNA 3' end formation, since some CBs associate with snRNA and histone genes in a transcription-dependent manner¹²⁴⁻¹²⁸.

Table I. Sub-nuclear Localization of Pre-mRNA Processing Factors

<u>Pre-mRNA processing factors</u>	<u>Nuclear localization</u>
Capping	
capping enzymes	?
cap-binding complex	nucleoplasm
Splicing factors	
snRNPs	nucleoplasm (SFCs), Cajal bodies
SR proteins	nucleoplasm (SFCs)
U2AF	nucleoplasm (SFCs)
PTB (hnRNP I)	nucleoplasm, perinucleolar compartment
Polyadenylation	
Poly(A) polymerase	nucleoplasm
CPSF-100	nucleoplasm, cleavage bodies
CstF-64	nucleoplasm, cleavage bodies
CstF-50	nucleoplasm
PAB II	nucleoplasm (SFCs)
Histone 3' end formation	
U7 snRNP	Cajal bodies
SLBP	Cajal bodies

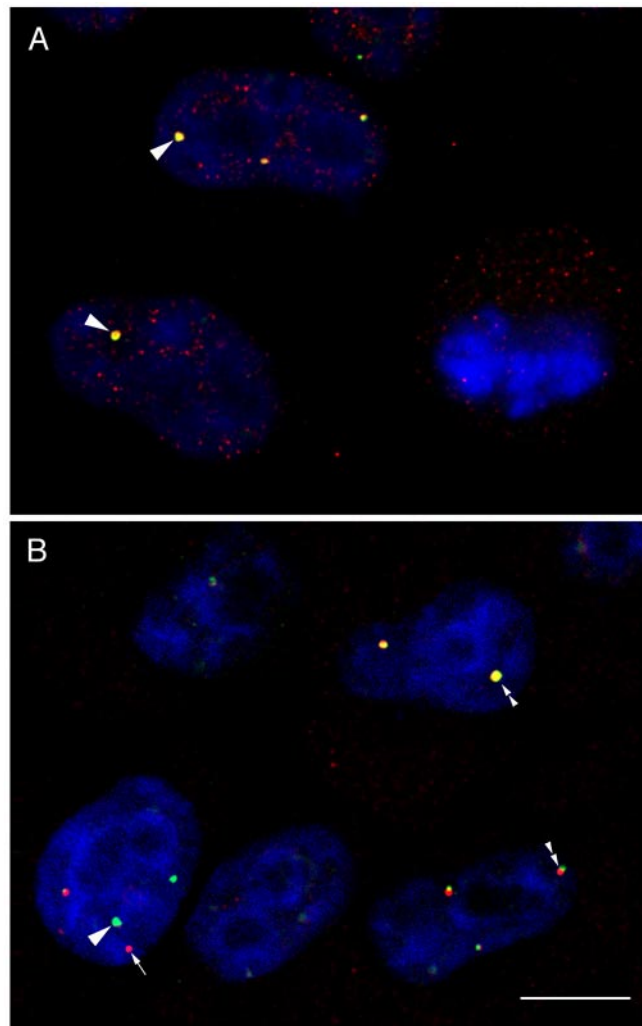


Figure 3. Spliceosomal snRNPs are distributed throughout the nucleoplasm and in Cajal bodies

(A) HeLa cells labeled with anti-coilin antibody (green) and anti-2,2,7-trimethylguanosine cap, a marker of snRNAs (red). The snRNAs are distributed throughout the nucleoplasm and concentrated in Cajal bodies (arrowheads). Cajal bodies are not detectable in mitotic cells (right bottom cell) and snRNAs are distributed in the cytoplasm. (B) Coilin (green) and survival motor neurons protein (SMN, red) in HeLa cells. SMN is found in the cytoplasm and in the nucleus where it concentrates in gems (arrow). Gems often associate with Cajal bodies (double arrowheads). A Cajal body not associated with gems is marked by an arrowhead. Bar 10μm.

The spliceosomal snRNPs are essential components of the spliceosome^{27,65}. The five major snRNPs U1, U2, U4, U5, and U6 function in the splicing of the majority of introns, which begin with GU and end with AG nucleotides. In metazoans, a minor class of snRNPs including U11, U12, U4atac and U6atac snRNPs function at introns beginning with AU and

ending in AC. Each snRNP is a complex of a single species of snRNA, a set of seven shared proteins called Sm (in the U1, U2, U4, U5, U11, U12, and U4atac snRNPs) or Lsm (in the U6 and U6 atac snRNPs), and a variety of snRNP-specific proteins¹²⁹. The snRNAs are highly structured within each snRNP, due to intramolecular base-pairing that creates specific helices and stem-loops. During the splicing reaction, the U2, U4, U5, and U6 snRNPs (and their minor snRNP equivalents) undergo extensive rearrangements through the unwinding and re-annealing of base-pairing interactions. Because snRNPs are re-used for subsequent rounds of splicing, they must be re-assembled into functional snRNPs in a process termed the spliceosome cycle.

The U1, U2, U4, and U5 snRNA genes are transcribed by pol II, and consequently these snRNAs receive a mono-methyl cap at their 5' ends. The snRNAs are transported to the cytoplasm where they are bound by the Sm proteins and other snRNP-specific proteins. The Sm proteins B/B', D, and D3 are modified by symmetrical arginine dimethylation by the 20S methylosome¹³⁰, creating a binding site for the SMN complex, so named for the Survival of Motor Neurons protein, which is mutant in the inherited disease Spinal Muscular Atrophy (see chapter 12)¹³¹⁻¹³⁵. Modification of the Sm proteins by arginine di-methylation and snRNP assembly by the SMN complex is thought to occur in the cytoplasm; however, SMN is also detectable in nuclei in structures known as gems that often overlap with CBs (Figure 3B). The function of the nuclear pool of SMN is currently unknown.

After association with the Sm proteins, the 5' ends of the snRNAs are hypermethylated in the cytoplasm to produce a tri-methylguanosine (TMG) cap¹³⁶, which together with the core snRNP structure serves as an important nuclear import signal¹³⁷⁻¹⁴⁰. Additional snRNA modifications include site-specific nucleotide 2'-O-ribose methylation and isomerization of uridines to pseudo-uridine residues¹⁴¹. A set of small RNAs (U85, U87, U88, and U89) was recently shown to guide these snRNA modifications, and these guide RNAs have been termed scaRNAs to reflect their localization in CBs¹⁴². The concentration of the scaRNAs along with their spliceosomal snRNA substrates in CBs suggests that these

guided modifications may occur in CBs, but it is currently unclear whether they occur before or after snRNP assembly and re-import. The finding that newly synthesized Sm proteins are detectable in CBs before they accumulate in the nucleoplasm¹⁴³⁻¹⁴⁵ suggests that CBs serve as an entry point for snRNPs imported from the cytoplasm and may be the site where the final steps of snRNP maturation take place.

In contrast to other snRNAs, the U6 gene is transcribed by RNA polymerase III and lacks a 5' TMG cap. U6 snRNA also lacks the Sm binding site, but it interacts instead with 7 proteins named Lsm2-Lsm8. The U6 snRNA does not leave the nucleus during its biogenesis; instead, U6 is transiently localized in the nucleolus¹⁴⁶ where U6 snRNA presumably undergoes snoRNA mediated 2'-O-methylation and pseudouridylation^{147,148}. It is not known when or where U6 snRNA first associates with Lsm and other U6 specific proteins. Interestingly Lsm4 was detected in Cajal bodies (D.S. and K.N. unpublished results) as well as U6 snRNA^{149,150}.

An extensive rearrangement of snRNPs occurs during the assembly of an active spliceosome. The most dramatic rearrangements reflect the dynamic function of the U2, U4 and U6 snRNAs. Before spliceosome assembly, the U2 snRNA base-pairs with the 3' splice site of the pre-mRNA. When the U4/U6•U5 tri-snRNP joins the spliceosome, base-pairing between U4 and U6 is disrupted, U2 base-pairs with U6 to form a helix, and U6 contacts the pre-mRNA to perform its catalytic role in the first and second trans-esterification steps of splicing catalysis. Following splicing, U2, U4, U5, and U6 snRNPs are released from the spliceosome and must be regenerated before the next round of splicing. Little is known about the regeneration of the U2 snRNP, but the yeast protein Cus2p has been shown to rescue a mutation in U2 snRNA predicted to influence proper U2 folding¹⁵¹.

How and where are the U4, U5, and U6 snRNPs recycled? The yeast protein Prp24p is required for U4/U6 snRNP regeneration after splicing *in vivo*¹⁵², and was shown to promote U4/U6 re-annealing in an *in vitro* assay¹⁵³. Recently, the protein SART3/p110 was identified as the human homologue of Prp24p and its activity as a U4/U6 annealing factor was shown *in vitro*¹⁵⁴. In addition, the U6-specific Lsm proteins have been shown to promote

U4/U6 annealing in humans ¹⁵⁵ and interact directly with yeast Prp24p ¹⁵⁶, suggesting that Prp24p/SART3 and Lsm proteins co-operate during U4/U6 snRNP assembly and recycling. Although a pool of U4/U6 snRNP is measurable in cells, it associates with the U5 snRNP before addition to the spliceosome. The U4/U6•U5 tri-snRNP-specific protein 61K was recently shown to bridge U5 with the U4/U6 snRNP and thus may be important for tri-snRNP assembly ¹⁵⁷. Unlike the Lsm proteins or 61K, Prp24p/SART3 is not a stable component of any snRNP and is thought to associate only transiently with its substrates. Prp24p/SART3 is not detectable in spliceosomes, suggesting that snRNP recycling occurs away from sites of transcription and splicing. Localization of U4, U5, and U6 snRNAs and the 61K protein to CBs raises the possibility that CBs are the site of snRNP recycling as well as assembly. If the nuclear pool of SMN has a role in snRNP recycling, then the concentration of SMN in some CBs would support this hypothesis.

3.3 The sub-nuclear distribution of capping and polyadenylation factors

Although the capping enzymes have been localized by in vivo crosslinking to TUs, their sub-cellular distribution has not yet been examined by light or electron microscopy. The cap-binding proteins CBP80 and CBP20 are prominently detected throughout the nucleoplasm of HeLa cells as bright dots, consistent with their expected concentration at TUs and on mRNPs transiting the nucleus, and faintly detectable in the cytoplasm where they are expected to support the first round of translation ^{20,158}. Consistent with this, CBP20 is present throughout the nucleoplasm and faintly cytoplasmic in the salivary glands of *Chironimus tentans*; CBP20 is highly enriched at sites of transcription in chromosome spreads and is seen to transit to nuclear pores with mRNP complexes by electron microscopy ²⁰. Thus, the subcellular localization of the CBC mirrors its biochemically and genetically defined functions at TUs, in mRNP nucleo-cytoplasmic transport, and in translation.

Several factors involved in polyadenylation, including poly(A) polymerase (PAP), CstF-50 subunit, and poly(A) binding protein (PAB II), are distributed throughout the nucleoplasm ^{159,160}, consistent with their activities at TUs and post-transcriptionally. However,

CstF-64 subunit and CPSF-100 have been detected both diffusely in the nucleoplasm and in a relatively novel nuclear body, termed the cleavage body ¹⁶¹. Cleavage bodies were found to sometimes overlap with Cajal Bodies, particularly during the G1 phase of the cell cycle. During S-phase, cleavage bodies are often located adjacent to CBs and overlapping with replication-dependent histone gene clusters ¹²⁶. This suggests that CstF-64 and CPSF and/or as yet unidentified components of the cleavage body may have a role in histone mRNA processing. At present, this interpretation remains paradoxical, since these histone mRNAs are not polyadenylated; U7 snRNP and SLBP, factors required for histone mRNA 3' end formation, are instead concentrated in CBs, and CstF and CPSF have no known function in histone mRNA processing. It will be of interest to determine whether polyadenylation cleavage factors themselves are localized to cleavage bodies at histone genes. If so, this may suggest that the elusive histone 3' end cleavage factor might be shared with the polyadenylation machinery and that the overlap of CstF-64 and CPSF with histone genes may be fortuitous. Because cleavage bodies are largely transcriptionally inactive ¹⁶¹, they may be the sites of post-transcriptional polyadenylation, and this possibility remains to be tested.

4. What is Nucleoplasm?

Like cytoplasm, “nucleoplasm” generally describes the internal contents of the cell nucleus, distinct from discrete structures such as chromosomes, nuclear bodies, and the nuclear envelope. However, uncondensed chromatin in the interphase nucleus is interspersed with soluble molecules, and thus nucleoplasm at interphase must be imagined as being quite mixed with chromatin, providing the environment in which gene expression and DNA replication occurs. Factors involved in transcription and splicing diffuse relatively rapidly throughout the nucleus as if few physical or chemical barriers exist ⁶⁷. The free diffusion of these nucleoplasmic factors is in contrast to the restricted movement of the histones, which like the mobile factors, are distributed throughout the nucleoplasm but are nevertheless anchored to relatively immobile chromatin ⁶⁷. These recent mobility studies carried out in live cells have

contributed a new perspective to the meaning of “nucleoplasm”, containing distinct populations of mobile and immobile molecular constituents ^{162,163}.

It has been proposed that nucleoplasm is subcompartmentalized into domains, some of which are enriched in particular sets of pre-mRNA processing factors ¹⁶⁴(Table I). Recently, the perinulceolar compartment (PNC) has been described, containing the RNA-binding and splicing factor PTB/hnRNP-I and five RNA pol III transcripts ^{165,166}. The PNC and several other perinucleolar compartments, namely Sam68 bodies containing RNA-binding proteins and another body containing hnRNP-L, have been described, and the functions for all three remain unknown ¹⁶⁷. The prototypical nucleoplasmic compartment is the “nuclear speckle” or splicing factor compartment (SFC) containing a soluble pool of snRNP and non-snRNP splicing factors, which appears as an interconnected meshwork penetrating the nucleoplasm at the light microscopic level ^{67,168-170}. At the EM level, the SFC corresponds to electron-lucent regions of chromatin-poor nucleoplasm, including perichromatin fibrils, thought to consist of nascent transcripts, and interchromatin granules ^{164,171}.

Because splicing is largely co-transcriptional, it was perplexing that many sites of transcription and splicing did not overlap with SFCs ^{172,173}. This paradox was largely resolved by the finding that dilution of antibodies specific for snRNP and non-snRNP splicing factors produces a punctate staining pattern that significantly overlaps with sites of pol II transcription (Figure 3) ⁴¹. Antibodies that recognize single species of splicing factors, rather than entire families of factors, reveal only the punctate staining pattern at sites of transcription ⁴¹. Although SFCs appear to contain dramatic concentrations of some splicing factors by fluorescence microscopy, quantitative measurement of splicing factors detected within and outside of SFCs reveals only a modest concentration (1.6- to 3-fold elevations) of splicing factors in SFCs ^{67,174}. Taken together, the simplest view of the SFC is that it represents the interchromatin space in which splicing factors may occur at elevated concentrations, perhaps because they are undiluted by chromatin. As described above, some splicing is likely to occur in the SFC, for those TUs located within or near the SFC or for transcripts that undergo post-transcriptional splicing. The SFC is expected to provide splicing factors, such as SR proteins,

to transcription sites⁴⁰ and may also be a site where inactive snRNP and non-snRNP splicing factors (e.g. incorrectly phosphorylated or de-phosphorylated SR proteins^{175,176} are sequestered and/or recycled.

5. Concluding Remarks

In this chapter, we have addressed how and where pre-mRNA is processed. Pre-mRNA processing begins at transcription units and sometimes continues post-transcriptionally. Co-transcriptional mRNA processing is likely the outcome of *i*) the relatively fast kinetics of processing reactions compared with the relatively long time that it takes to synthesize an entire pre-mRNA and *ii*) direct binding of some RNA processing factors to the transcriptional machinery. Many of the mechanistic details of processing factor recruitment and function, particularly with respect to splicing, have yet to be elucidated. Here we propose a model in which nascent mRNPs still attached to the DNA template by pol II contain heterogeneous constellations of processing factors based on the locations within TUs at which processing reactions occur. TUs are distributed throughout the nucleoplasm, which is characterized by highly mobile sets of soluble factors and a limited set of restricted elements, such as chromatin. However, the concentration of many pre-mRNA processing factors in nuclear bodies and sub-domains points to the possible compartmentalization of some processing reactions and/or the assembly of processing complexes away from sites of transcription. Understanding the function of nuclear bodies and what possible advantages they may confer to cells remains a major challenge for cell biologists.

Acknowledgements

Work in our laboratory has been supported by grants from the National Science Foundation (to KN) and from the American Cancer Society (to KN), and by the Max Planck Gesellschaft.

References

- 1 E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chisoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, *Nature* **409**, 860 (2001).
- 2 D. A. Jackson, A. B. Hassan, R. J. Errington, and P. R. Cook, *Embo J.* **12**, 1059 (1993)
- 3 D. G. Wansink, W. Schul, I. van der Kraan, B. van Steensel, R. van Driel, and L. de Jong, *J. Cell Biol.* **122**, 283 (1993)

- 4 K. M. Neugebauer and M. B. Roth, *Genes Dev.* **11**, 3279 (1997)
- 5 A. L. Beyer and Y. N. Osheim, *Genes Dev.* **2**, 754 (1988)
- 6 Y. N. Osheim, O. L. Miller, Jr., and A. L. Beyer, *Cell* **43**, 143 (1985)
- 7 D. Bentley, *Curr. Opin. Cell Biol.* **14**, 336 (2002)
- 8 P. R. Cook, *Science* **284**, 1790 (1999)
- 9 T. Maniatis and R. Reed, *Nature* **416**, 499 (2002)
- 10 N. J. Proudfoot, A. Furger, and M. J. Dye, *Cell* **108**, 501 (2002)
- 11 A. J. Shatkin and J. L. Manley, *Nat. Struct. Biol.* **7**, 838 (2000)
- 12 S. Shuman, *Prog. Nucleic Acid Res. Mol. Biol.* **66**, 1 (2001)
- 13 E. J. Cho, C. R. Rodriguez, T. Takagi, and S. Buratowski, *Genes Dev.* **12**, 3482 (1998)
- 14 C. K. Ho, B. Schwer, and S. Shuman, *Mol. Cell. Biol.* **18**, 5189 (1998)
- 15 C. K. Ho and S. Shuman, *Mol Cell* **3**, 405-11. (1999).
- 16 E. B. Rasmussen and J. T. Lis, *Proc. Natl. Acad. Sci .USA* **90**, 7923 (1993)
- 17 S. C. Schroeder, B. Schwer, S. Shuman, and D. Bentley, *Genes Dev.* **14**, 2435 (2000)
- 18 P. Komarnitsky, E. J. Cho, and S. Buratowski, *Genes Dev.* **14**, 2452 (2000)
- 19 J. D. Lewis and E. Izaurralde, *Eur. J. Biochem.* **247**, 461 (1997)
- 20 N. Visa, E. Izaurralde, J. Ferreira, B. Daneholt, and I. W. Mattaj, *J. Cell Biol.* **133**, 5 (1996)
- 21 H. V. Colot, F. Stutz, and M. Rosbash, *Genes Dev.* **10**, 1699 (1996)
- 22 J. D. Lewis, D. Gorlich, and I. W. Mattaj, *Nucleic Acids Res.* **24**, 333 (1996).
- 23 J. D. Lewis, E. Izaurralde, A. Jarmolowski, C. McGuigan, and I. W. Mattaj, *Genes Dev.* **10**, 1683 (1996)
- 24 D. Gorlich, R. Kraft, S. Kostka, F. Vogel, E. Hartmann, R. A. Laskey, I. W. Mattaj, and E. Izaurralde, *Cell* **87**, 21 (1996)
- 25 P. Fortes, T. Inada, T. Preiss, M. W. Hentze, I. W. Mattaj, and A. B. Sachs, *Mol. Cell* **6**, 191 (2000)
- 26 Y. Ishigaki, X. Li, G. Serin, and L. E. Maquat, *Cell* **106**, 607 (2001)

- 27 J. P. Staley and C. Guthrie, *Cell* **92**, 315 (1998)
- 28 M. S. Jurica, L. J. Licklider, S. R. Gygi, N. Grigorieff, and M. J. Moore, *Rna* **8**, 426 (2002)
- 29 M. D. Chiara, O. Gozani, M. Bennett, P. Champion-Arnaud, L. Palandjian, and R. Reed, *Mol. Cell. Biol.* **16**, 3317 (1996)
- 30 R. Reed, *Curr. Opin. Cell Biol.* **12**, 340 (2000)
- 31 C. W. Smith and J. Valcarcel, *Trends Biochem. Sci.* **25**, 381 (2000)
- 32 S. W. Stevens, D. E. Ryan, H. Y. Ge, R. E. Moore, M. K. Young, T. D. Lee, and J. Abelson, *Mol. Cell* **9**, 31 (2002)
- 33 S. Yitzhaki, E. Miriami, R. Sperling, and J. Sperling, *Proc .Natl. Acad .Sci. USA* **93**, 8830 (1996)
- 34 G. Bauren and L. Wieslander, *Cell* **76**, 183 (1994)
- 35 I. Wetterberg, G. Bauren, and L. Wieslander, *Rna* **2**, 641 (1996)
- 36 M. F. LeMaire and C. S. Thummel, *Mol. Cell. Biol.* **10**, 6059 (1990)
- 37 J. Wuarin and U. Schibler, *Mol. Cell. Biol.* **14**, 7219 (1994)
- 38 C. N. Tennyson, H. J. Klamut, and R. G. Worton, *Nat. Genet.* **9**, 184 (1995)
- 39 D. J. Elliott and M. Rosbash, *Exp. Cell Res.* **229**, 181 (1996)
- 40 T. Misteli, J. F. Caceres, and D. L. Spector, *Nature* **387**, 523 (1997)
- 41 K. M. Neugebauer and M. B. Roth, *Genes Dev.* **11**, 1148 (1997)
- 42 T. Misteli and D. L. Spector, *Mol. Cell* **3**, 697 (1999)
- 43 G. C. Roberts, C. Gooding, H. Y. Mak, N. J. Proudfoot, and C. W. Smith, *Nucleic Acids Res.* **26**, 5568 (1998)
- 44 S. Kadener, P. Cramer, G. Nogues, D. Cazalla, M. de La Mata, J. P. Fededa, S. E. Werbajh, A. Srebrow, and A. R. Kornblihtt, *Embo J.* **20**, 5759 (2001)
- 45 N. Fong and D. L. Bentley, *Genes Dev.* **15**, 1783 (2001)
- 46 Y. Hirose, R. Tacke, and J. L. Manley, *Genes Dev.* **13**, 1234 (1999)
- 47 C. Zeng and S. M. Berget, *Mol. Cell Biol.* **20**, 8290 (2000)
- 48 J. L. Corden, *Trends Biochem. Sci.* **15**, 383 (1990)

- 49 A. L. Greenleaf, Trends Biochem. Sci. **18**, 117 (1993)
- 50 D. P. Morris and A. L. Greenleaf, J. Biol. Chem. **275**, 39935 (2000)
- 51 A. Yuryev, M. Patturajan, Y. Litingtung, R. V. Joshi, C. Gentile, M. Gebara, and J. L. Corden, Proc. Natl. Acad. Sci. USA **93**, 6975 (1996)
- 52 E. Kiseleva, T. Wurtz, N. Visa, and B. Daneholt, Embo J. **13**, 6052 (1994)
- 53 K. Kohrer, K. Vogel, and H. Domdey, Embo J. **9**, 705 (1990)
- 54 T. Tani and Y. Ohshima, Genes Dev. **5**, 1022 (1991)
- 55 K. Dower and M. Rosbash, Rna **8**, 686 (2002)
- 56 D. D. Licatosi, Geiger, G., Minet, M., Schroder, S., Cilli, K., McNeil, J.B., and D.L. Bentley, Mol. Cell **9**, 1101 (2002)
- 57 Y. W. Fong and Q. Zhou, Nature **414**, 929 (2001)
- 58 S. Huang and D. L. Spector, J. Cell Biol. **133**, 719 (1996)
- 59 C. Jolly, C. Vourc'h, M. Robert-Nicoud, and R. I. Morimoto, J. Cell Biol. **145**, 1133 (1999)
- 60 H. Ge, Y. Si, and A. P. Wolffe, Mol. Cell **2**, 751 (1998)
- 61 P. Cramer, J. F. Caceres, D. Cazalla, S. Kadener, A. F. Muro, F. E. Baralle, and A. R. Kornblihtt, Mol. Cell **4**, 251 (1999)
- 62 P. Cramer, C. G. Pesce, F. E. Baralle, and A. R. Kornblihtt, Proc. Natl. Acad. Sci. USA **94**, 11456 (1997)
- 63 A. C. Goldstrohm, A. L. Greenleaf, and M. A. Garcia-Blanco, Gene **277**, (2001)
- 64 P. Cramer, D. A. Bushnell, and R. D. Kornberg, Science **292**, 1863 (2001)
- 65 Y.-T. Yu, E.C. Sharl, C.M. Smith and J.A. Steitz., in *The RNA World*, 2 ed., edited by C. Gesteland, and Atkins, Cold Spring harbor Laboratory Press, Cold Spring harbor, New York (1999) p.487
- 66 A. Hanamura, J. F. Caceres, A. Mayeda, B. R. Franza, Jr., and A. R. Krainer, Rna **4**, 430 (1998)
- 67 R. D. Phair and T. Misteli, Nature **404**, 604 (2000)
- 68 R. J. Nagel, A. M. Lancaster, and A. M. Zahler, Rna **4**, 11 (1998)

- 69 A. R. Hatton, V. Subramaniam, and A. J. Lopez, *Mol. Cell* **2**, 787 (1998)
- 70 L. P. Keegan, A. Gallo, and M. A. O'Connell, *Nat. Rev. Genet.* **2**, 869 (2001)
- 71 S. M. Rueter, T. R. Dawson, and R. B. Emeson, *Nature* **399**, 75 (1999)
- 72 I. Melcak and I. Raska, *J. Struct. Biol.* **117**, 189 (1996)
- 73 R. W. Dirks, K. C. Daniel, and A. K. Raap, *J. Cell Sci.* **108**, 2565 (1995)
- 74 Z. Zachar, J. Kramer, I. P. Mims, and P. M. Bingham, *J. Cell Biol.* **121**, 729 (1993)
- 75 C. Johnson, D. Primorac, M. McKinstry, J. McNeil, D. Rowe, and J. B. Lawrence, *J. Cell Biol.* **150**, 417 (2000)
- 76 O. P. Singh, B. Bjorkroth, S. Masich, L. Wieslander, and B. Daneholt, *Exp. Cell Res.* **251**, 135 (1999)
- 77 J. C. Politz, R. A. Tuft, T. Pederson, and R. H. Singer, *Curr. Biol.* **9**, 285 (1999)
- 78 J. C. Politz, E. S. Browne, D. E. Wolf, and T. Pederson, *Proc. Natl. Acad. Sci. USA* **95**, 6043 (1998)
- 79 G. S. Wilkie and I. Davis, *Cell* **105**, 209 (2001)
- 80 I. Melcak, S. Cermanova, K. Jirsova, K. Koberna, J. Malinsky, and I. Raska, *Mol. Biol. Cell* **11**, 497 (2000)
- 81 E. Wahle and U. Ruegsegger, *FEMS Microbiol. Rev.* **23**, 277 (1999)
- 82 N. Proudfoot, *Trends Biochem. Sci.* **25**, 290 (2000)
- 83 Y. Hirose and J. L. Manley, *Nature* **395**, 93 (1998)
- 84 O. Calvo and J. L. Manley, *Mol. Cell* **7**, 1013 (2001)
- 85 J. C. Dantonel, K. G. Murthy, J. L. Manley, and L. Tora, *Nature* **389**, 399 (1997)
- 86 S. McCracken, N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S. D. Patterson, M. Wickens, and D. L. Bentley, *Nature* **385**, 357 (1997)
- 87 D. Barilla, B. A. Lee, and N. J. Proudfoot, *Proc. Natl. Acad. Sci. USA* **98**, 445 (2001)
- 88 G. Bauren, S. Belikov, and L. Wieslander, *Genes Dev* **12**, 2759-69. (1998).
- 89 C. E. Birse, L. Minvielle-Sebastia, B. A. Lee, W. Keller, and N. J. Proudfoot, *Science* **280**, 298 (1998)
- 90 Y. N. Osheim, N. J. Proudfoot, and A. L. Beyer, *Mol. Cell* **3**, 379 (1999)

- 91 Y. N. Osheim, Zikes, M.L., and A.L. Beyer, *Chromosoma* **111**, 1 (2002)
- 92 Y. Nogi, R. Yano, and M. Nomura, *Proc. Natl. Acad. Sci. USA* **88**, 3962 (1991)
- 93 J. B. McNeil, H. Agah, and D. Bentley, *Genes Dev.* **12**, 2510 (1998)
- 94 H. J. Lo, H. K. Huang, and T. F. Donahue, *Mol. Cell. Biol.* **18**, 665 (1998)
- 95 L. C. Chao, A. Jamil, S. J. Kim, L. Huang, and H. G. Martinson, *Mol. Cell. Biol.* **19**, 5588 (1999)
- 96 M. Niwa, C. C. MacDonald, and S. M. Berget, *Nature* **360**, 277 (1992)
- 97 S. Vagner, C. Vagner, and I. W. Mattaj, *Genes Dev.* **14**, 403 (2000)
- 98 S. I. Gunderson, S. Vagner, M. Polycarpou-Schwarz, and I. W. Mattaj, *Genes Dev.* **11**, 761 (1997)
- 99 Y. Takagaki and J. L. Manley, *Mol. Cell* **2**, 761 (1998)
- 100 H. Lou, K. M. Neugebauer, R. F. Gagel, and S. M. Berget, *Mol. Cell. Biol.* **18**, 4977 (1998)
- 101 T. H. Jensen, K. Patricio, T. McCarthy, and M. Rosbash, *Mol. Cell* **7**, 887 (2001)
- 102 P. Hilleren, T. McCarthy, M. Rosbash, R. Parker, and T. H. Jensen, *Nature* **413**, 538 (2001)
- 103 P. Mitchell and D. Tollervey, *Curr. Opin. Cell Biol.* **13**, 320 (2001)
- 104 C. Bousquet-Antonelli, C. Presutti, and D. Tollervey, *Cell* **102**, 765 (2000)
- 105 R. Reed and E. Hurt, *Cell* **108**, 523 (2002)
- 106 E. P. Lei, H. Krebber, and P. A. Silver, *Genes Dev.* **15**, 1771 (2001)
- 107 K. Strasser, S. Masuda, P. Mason, J. Pfannstiel, M. Oppizzi, S. Rodriguez-Navarro, A. G. Rondon, A. Aguilera, K. Struhl, R. Reed, and E. Hurt, *Nature* **28**, 28 (2002)
- 108 N. Custodio, M. Carmo-Fonseca, F. Geraghty, H. S. Pereira, F. Grosveld, and M. Antoniou, *Embo J.* **18**, 2855 (1999)
- 109 O. Muhlemann, C. S. Mock-Casagrande, J. Wang, S. Li, N. Custodio, M. Carmo-Fonseca, M. F. Wilkinson, and M. J. Moore, *Mol. Cell* **8**, 33 (2001)
- 110 D. S. Horowitz, E. J. Lee, S. A. Mabon, and T. Misteli, *Embo J.* **21**, 470 (2002)
- 111 Z. Dominski and W. F. Marzluff, *Gene* **239**, 1 (1999)

- 112 A. S. Williams, T. C. Ingledue, 3rd, B. K. Kay, and W. F. Marzluff, *Nucleic Acids Res.* **22**, 4660 (1994)
- 113 D. R. Gallie, N. J. Lewis, and W. F. Marzluff, *Nucleic Acids Res.* **24**, 1954 (1996)
- 114 M. L. Whitfield, L. X. Zheng, A. Baldwin, T. Ohta, M. M. Hurt, and W. F. Marzluff, *Mol. Cell. Biol.* **20**, 4188 (2000)
- 115 N. Chodchoy, N. B. Pandey, and W. F. Marzluff, *Mol. Cell. Biol.* **11**, 497 (1991)
- 116 I. Wetterberg, J. Zhao, S. Masich, L. Wieslander, and U. Skoglund, *Embo J.* **20**, 2564 (2001)
- 117 A. G. Matera, *Trends Cell Biol.* **9**, 302 (1999)
- 118 M. Dundr and T. Misteli, *Biochem J.* **356**, 297 (2001)
- 119 D. L. Spector, *J. Cell Sci.* **114**, 2891 (2001)
- 120 J. G. Gall, *Annu. Rev. Cell Dev. Biol.* **16**, 273 (2000)
- 121 J. G. Gall, M. Bellini, Z. Wu, and C. Murphy, *Mol. Biol. Cell* **10**, 4385 (1999)
- 122 M. Platani, I. Goldberg, J. R. Swedlow, and A. I. Lamond, *J. Cell Biol.* **151**, 1561 (2000)
- 123 S. Snaar, K. Wiesmeijer, A. G. Jochemsen, H. J. Tanke, and R. W. Dirks, *J. Cell Biol.* **151**, 653 (2000)
- 124 K. P. Smith, K. C. Carter, C. V. Johnson, and J. B. Lawrence, *J. Cell Biochem.* **59**, 473 (1995)
- 125 W. Schul, R. van Driel, and L. de Jong, *Mol. Biol. Cell* **9**, 1025 (1998)
- 126 W. Schul, I. van Der Kraan, A. G. Matera, R. van Driel, and L. de Jong, *Mol. Biol. Cell* **10**, 3815 (1999)
- 127 M. R. Frey, A. D. Bailey, A. M. Weiner, and A. G. Matera, *Curr Biol.* **9**, 126 (1999)
- 128 M. R. Frey and A. G. Matera, *J. Cell Biol.* **154**, 499 (2001)
- 129 C. Schneider, C. L. Will, O. V. Makarova, E. M. Makarov, and R. Luhrmann, *Mol. Cell. Biol.* **22**, 3219 (2002)
- 130 W. J. Friesen, S. Paushkin, A. Wyce, S. Massenet, G. S. Pesiridis, G. Van Duyne, J. Rappsilber, M. Mann, and G. Dreyfuss, *Mol. Cell. Biol.* **21**, 8289 (2001)

- 131 H. Brahms, L. Meheus, V. de Brabandere, U. Fischer, and R. Luhrmann, *Rna* **7**, 1531 (2001)
- 132 W. J. Friesen, S. Massenet, S. Paushkin, A. Wyce, and G. Dreyfuss, *Mol. Cell* **7**, 1111 (2001)
- 133 G. Meister, D. Buhler, R. Pillai, F. Lottspeich, and U. Fischer, *Nat. Cell Biol.* **3**, 945 (2001)
- 134 G. Meister, C. Eggert, D. Buhler, H. Brahms, C. Kambach, and U. Fischer, *Curr. Biol.* **11**, 1990 (2001)
- 135 S. Paushkin, A. K. Gubitza, S. Massenet, and G. Dreyfuss, *Curr. Opin. Cell Biol.* **14**, 305 (2002)
- 136 I. W. Mattaj, *Cell* **46**, 905 (1986)
- 137 J. Hamm, E. Darzynkiewicz, S. M. Tahara, and I. W. Mattaj, *Cell* **62**, 569 (1990)
- 138 U. Fischer, E. Darzynkiewicz, S. M. Tahara, N. A. Dathan, R. Luhrmann, and I. W. Mattaj, *J. Cell Biol.* **113**, 705 (1991)
- 139 U. Fischer and R. Luhrmann, *Science* **249**, 786 (1990)
- 140 U. Fischer, V. Sumpster, M. Sekine, T. Satoh, and R. Luhrmann, *Embo J.* **12**, 573 (1993)
- 141 S. Massenet, A. Mougin, and C. Branlant, in *Modification and Editing of RNA*, edited by H. G. a. R. Benne, ASM Press, Washington, DC (1998) p.201
- 142 X. Darzacq, B. E. Jady, C. Verheggen, A. M. Kiss, E. Bertrand, and T. Kiss, *Embo J.* **21**, 2746 (2002)
- 143 J. E. Sleeman, P. Ajuh, and A. I. Lamond, *J. Cell Sci.* **114**, 4407 (2001)
- 144 J. E. Sleeman and A. I. Lamond, *Curr. Biol.* **9**, 1065 (1999)
- 145 T. Carvalho, F. Almeida, A. Calapez, M. Lafarga, M. T. Berciano, and M. Carmo-Fonseca, *J. Cell Biol.* **147**, 715 (1999)
- 146 T. S. Lange and S. A. Gerbi, *Mol. Biol. Cell* **11**, 2419 (2000)
- 147 K. T. Tycowski, Z. H. You, P. J. Graham, and J. A. Steitz, *Mol. Cell* **2**, 629 (1998)

- 148 P. Ganot, B. E. Jady, M. L. Bortolin, X. Darzacq, and T. Kiss, *Mol. Cell. Biol.* **19**, 6906 (1999)
- 149 A. G. Matera and D. C. Ward, *J. Cell Biol.* **121**, 715 (1993)
- 150 M. Carmo-Fonseca, R. Pepperkok, M. T. Carvalho, and A. I. Lamond, *J. Cell Biol.* **117**, 1 (1992)
- 151 D. Yan, R. Perriman, H. Igel, K. J. Howe, M. Neville, and M. Ares, Jr., *Mol. Cell. Biol.* **18**, 5000 (1998)
- 152 K. W. Shannon and C. Guthrie, *Genes Dev.* **5**, 773 (1991)
- 153 P. L. Raghunathan and C. Guthrie, *Science* **279**, 857 (1998)
- 154 M. Bell, S. Schreiner, A. Damianov, R. Reddy, and A. Bindereif, *Embo J.* **21**, 2724 (2002)
- 155 T. Achsel, H. Brahm, B. Kastner, A. Bachi, M. Wilm, and R. Luhrmann, *Embo J.* **18**, 5789 (1999)
- 156 M. Fromont-Racine, A. E. Mayes, A. Brunet-Simon, J. C. Rain, A. Colley, I. Dix, L. Decourty, N. Joly, F. Ricard, J. D. Beggs, and P. Legrain, *Yeast* **17**, 95 (2000)
- 157 O. V. Makarova, E. M. Makarov, S. Liu, H. P. Vornlocher, and R. Luhrmann, *Embo J.* **21**, 1148 (2002)
- 158 E. Izaurralde, J. Lewis, C. McGuigan, M. Jankowska, E. Darzynkiewicz, and I. W. Mattaj, *Cell* **78**, 657 (1994)
- 159 W. Schul, R. van Driel, and L. de Jong, *Exp. Cell Res.* **238**, 1 (1998)
- 160 D. Cmarko, P. J. Verschure, T. E. Martin, M. E. Dahmus, S. Krause, X. D. Fu, R. van Driel, and S. Fakan, *Mol. Biol. Cell* **10**, 211 (1999)
- 161 W. Schul, B. Groenhout, K. Koberna, Y. Takagaki, A. Jenny, E. M. Manders, I. Raska, R. van Driel, and L. de Jong, *Embo J.* **15**, 2883 (1996)
- 162 T. Misteli, *Science* **291**, 843 (2001)
- 163 M. Carmo-Fonseca, *Cell* **108**, 513 (2002)
- 164 D. L. Spector, *Annu. Rev. Cell Biol.* **9**, 265 (1993)

- 165 A. Ghetti, S. Pinol-Roma, W. M. Michael, C. Morandi, and G. Dreyfuss, *Nucleic Acids Res.* **20**, 3671 (1992)
- 166 A. G. Matera, M. R. Frey, K. Margelot, and S. L. Wolin, *J. Cell Biol.* **129**, 1181 (1995)
- 167 S. Huang, *J. Struct. Biol.* **129**, 233 (2000)
- 168 X. D. Fu and T. Maniatis, *Nature* **343**, 437 (1990)
- 169 D. L. Spector, *Proc. Natl. Acad. Sci. USA* **87**, 147 (1990)
- 170 M. Gama-Carvalho, R. D. Krauss, L. Chiang, J. Valcarcel, M. R. Green, and M. Carmo-Fonseca, *J. Cell Biol.* **137**, 975 (1997)
- 171 S. Fakan, *Trends Cell. Biol.* **4**, 86 (1994)
- 172 G. Zhang, K. L. Taneja, R. H. Singer, and M. R. Green, *Nature* **372**, 809 (1994)
- 173 K. P. Smith, P. T. Moen, K. L. Wydner, J. R. Coleman, and J. B. Lawrence, *J. Cell Biol.* **144**, 617 (1999)
- 174 F. S. Fay, K. L. Taneja, S. Shenoy, L. Lifshitz, and R. H. Singer, *Exp. Cell Res.* **231**, 27 (1997)
- 175 T. Misteli and D. L. Spector, *Mol. Biol. Cell* **7**, 1559 (1996)
- 176 T. Misteli, J. F. Caceres, J. Q. Clement, A. R. Krainer, M. F. Wilkinson, and D. L. Spector, *J. Cell Biol.* **143**, 297-307. (1998).