An extensive network of coupling among gene expression machines

Tom Maniatis* **& Robin Reed†**

** Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA*
† Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Gene expression in eukaryotes requires several multi-component cellular machines. Each machine carries out a separate step in the gene expression pathway, which includes transcription, several pre-messenger RNA processing steps and the export of mature mRNA to the cytoplasm. Recent studies lead to the view that, in contrast to a simple linear assembly line, a complex and extensively coupled network has evolved to coordinate the activities of the gene expression machines. The extensive coupling is consistent with a model in which the machines are tethered to each other to form 'gene expression factories' that maximize the efficiency and specificity of each step in gene expression.

Eukaryotic gene expression is a complex stepwise process that begins with transcription initiation, elongation and termination (Fig. 1). During transcription, the nascent pre-mRNA is capped at the 5′ end, introns are removed by splicing, and the 3′ end is cleaved and polyadenylated. The mature mRNA is then released from the site of transcription and exported to the cytoplasm for translation. Superimposed on this pathway is an RNA surveillance system that eliminates aberrantly processed or mutant pre-mRNAs and mRNAs. Distinct machines carry out each of the steps in the gene expression pathway. Despite the unique reactions they catalyse, each machine also interfaces both physically and functionally with other machines in the pathway as detailed in several recent reviews[1–5]. Here we discuss evidence that coupling is even more extensive than previously imagined. Indeed coupling occurs not only between sequential steps in the gene expression pathway but also between the earliest and latest steps (see Fig. 1).

In this review we address the reasons why such extensive coupling exists. As discussed below, coupling may solve many of the logistical problems inherent in the gene expression pathway. For example, the production of mature mRNA requires that the nascent pre-mRNA is sufficiently stable to complete its synthesis, processing and export. One of the many functions of the 5′ cap is to protect the pre-mRNA from degradation. By tight coupling between the capping and transcription machineries rapid capping of the nascent pre-mRNA is ensured, thereby protecting it from degradation. Coupling also plays a critical role in gene expression by tethering machines to each other and to their substrates, a mechanism that dramatically increases the rate and specificity of enzymatic reactions. The possible consequences of tethering are illustrated by metazoan pre-mRNA splicing where small exons must be recognized in a vast sea of introns. This recognition problem may be solved at least in part by coupling transcription to splicing, which results in tethering splicing factors directly adjacent to the nascent pre-mRNA as it emerges from the polymerase. Tethering in general is widely used for regulating the activities within individual cellular machines[6]. As with the example of splicing above and as discussed throughout this review tethering is also used to coordinate activities between machines.

RNA polymerase II: coupling to RNA processing

Both the carboxy-terminal domain (CTD) of the RNA polymerase II large subunit and transcription elongation factors play central roles in coupling transcription to pre-mRNA processing. A number of studies have shown that the CTD functions both as an assembly platform for and a regulator of transcription and pre-mRNA processing machines. The CTD consists of 27 heptad repeats in yeast and 52 in humans, and dynamic site-specific phosphorylation and dephosphorylation of these repeats is a critical mechanism for CTD function[7–9]. Several different CTD kinases and at least one phosphatase have been implicated in this mechanism[9–11]. The importance of the CTD in pre-mRNA processing is illustrated by the effects of certain CTD deletions which do not inactivate transcription but significantly decrease the efficiency of capping[12,13], splicing[14] and polyadenylation[14]. The effects of these deletions on splicing and polyadenylation are not simply a secondary

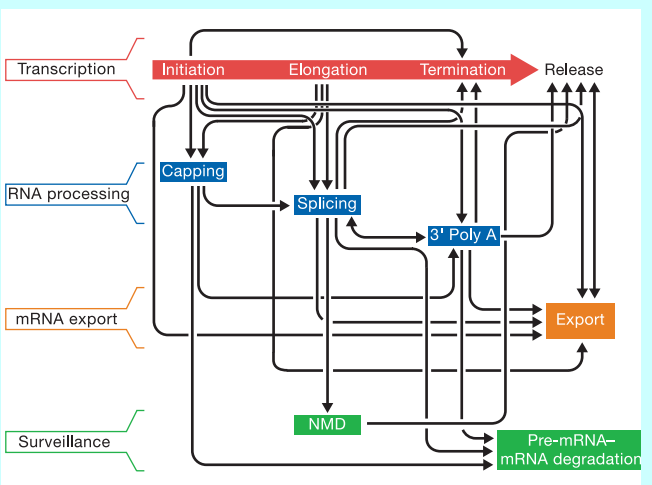

Figure 1 A complex network of coupled interactions in gene expression. The major steps of gene expression are indicated on the left and each stage in transcription is shown along the top in the red arrow. 'Release' indicates release of the mature mRNA from the site of transcription. The steps in RNA processing, mRNA export, nonsense-mediated decay (NMD) and RNA degradation are shown below the arrow. The black arrows indicate physical and/or functional coupling between two steps in gene expression. Each arrow is documented by published studies showing a physical and/or functional interaction between one or more components of gene expression machines. See text for a description of selected examples of coupling, and previous reviews for additional references and discussion[2,3,5].

consequence of the effects on capping[15]. Indeed recent studies indicate that the CTD independently promotes each step of the pre-mRNA processing pathway. Moreover distinct domains of the CTD appear to interact with factors required for different steps in pre-mRNA processing[15], at least in humans.

Insight into the relationship between the CTD and the nascent pre-mRNA was provided by the recent determination of the three-dimensional structure of yeast RNA polymerase II (ref. 16). The 2.8 Å resolution structure revealed that the CTD is directly adjacent to the exit groove for the pre-mRNA. The CTD is 650 Å in length and is connected to the large subunit of RNA polymerase II through a flexible 280 Å linker. Thus the CTD appears to be of sufficient size and flexibility to interact with multiple components of the pre-mRNA processing machinery and to localize this machinery close to the pre-mRNA as it emerges from the exit groove of the polymerase.

The importance of positioning the CTD immediately adjacent to the nascent pre-mRNA is illustrated by the process of 5′ end capping[4]. Capping takes place during the transition from transcription initiation to elongation when the nascent pre-mRNA is only 20–40 nucleotides long. During transcription initiation, serine 5 of the CTD heptad repeat is phosphorylated[9,10]. This phosphorylation triggers a cascade of events beginning with the dissociation of transcription initiation factors from the CTD followed by the recruitment of the capping machinery and the allosteric activation of the capping enzyme[8,17–19]. Serine 5 is subsequently dephosphorylated, resulting in release of the capping machinery[8,18]. Serine 2 of the heptad repeat is then phosphorylated, leading to the recruitment of factors involved in subsequent steps of RNA processing[8,9]. This switch in site-specific phosphorylation functions as a timing mechanism to coordinate the appearance of the 5′ end of the pre-mRNA with capping[8,9].

In mammalian cells the transcription elongation factor P-TEFb is thought to be the serine 2 CTD kinase[9,11,20]. In addition, P-TEFb phosphorylates another elongation factor, hSPT5 (refs 21 and 22). Another level of coupling between capping and transcription was revealed by the finding that hSPT5 interacts directly with and activates the capping enzyme[23]. Thus P-TEFb plays a pivotal role in the temporal coordination of capping and transcriptional elongation[4].

Pre-mRNA splicing is also coupled to transcription elongation in part through interactions between the splicing machinery and the transcription elongation factor TAT-SF1 (ref. 24). These interactions reveal yet another role for P-TEFb in coupling. Specifically, P-TEFb (which is recruited to promoters of cellular genes by transcription factors[25]) recruits TAT-SF1[26], which in turn recruits splicing factors to the nascent pre-mRNA[24]. *In vitro* studies reveal that the splicing machinery recruited by TAT-SF1 strongly stimulates transcription elongation[24]. The role of TAT-SF1 appears to be conserved as the yeast TAT-SF1 homologue (CUS2) associates with splicing factors and regulates the assembly of splicing complexes[27]. The connection between transcription elongation and splicing is further evidenced by *in vitro*[24] and *in vivo* studies showing that genes containing introns are more efficiently transcribed (reviewed in ref. 28). Finally, another transcription elongation factor, TFIIS, was recently shown to associate with a high molecular mass complex containing RNA polymerase II and several splicing factors[29]. Thus, at least three elongation factors have been implicated in coupling transcription elongation to splicing: P-TEFb, TAT-SF1 and TFIIS.

As is the case with capping, transcription is also coupled to both splicing and polyadenylation through interactions between the CTD and the respective processing machineries[2,14]. For example, a number of components of the polyadenylation machinery associate with the CTD[2] and these interactions are required for maximum levels of polyadenylation *in vivo*[14,15]. Similarly a set of splicing-related proteins (SCAFs)[30] and various components of the splicing machinery associate with the CTD[1,2]. Among the CTD-associated splicing factors is Prp40 (ref. 31) which is thought to function in bringing the 5′ and 3′ splice sites together during the splicing reaction[32].

Coupling between the splicing machinery and the transcription initiation and elongation complexes may solve long-standing problems in metazoan pre-mRNA splicing. Most metazoan pre-mRNAs contain very short exons (~140 nucleotides)[33] whereas the introns are often tens of thousands of nucleotides long. Indeed one intron in human neurexin pre-mRNA is approximately 480,000 nucleotides in length (B. Graveley, personal communication). Thus several hours pass between the time of synthesis of the 5′ and 3′ splice sites of this intron. In addition, metazoan pre-mRNAs contain multiple exons and introns, and splice-site sequences are highly degenerate. Thus the splicing machinery must specifically recognize each exon in the context of enormous introns and bring them together in the correct 5′ to 3′ order.

Each of these problems may be solved by tethering the splicing machinery to both the CTD and the transcription elongation complex. This tethering concentrates the splicing machinery directly adjacent to the nascent pre-mRNA as it emerges from the polymerase[2,24,34]. The specific recognition of exons may be facilitated by this tethering, which effectively reduces the kinetics of the interaction between the splicing machinery and the pre-mRNA from first to zero order. With first-order kinetics, factors interact by free diffusion, dramatically decreasing the probability of an interaction. Consequently by coupling transcription to splicing, the splicing machinery has a strong competitive advantage over the abundant non-specific RNA-binding proteins that can contact the pre-mRNA only through free diffusion. As a result of coupling, the splicing machinery has an increased probability of interacting with the pre-mRNA, thereby increasing the specific recognition of exon sequences. Coupling between the transcription and splicing machineries is also likely to be important in ensuring that exons are joined in the correct 5′ to 3′ order during splicing. This order may be established by tethering each newly synthesized exon and the adjacent splice sites to the CTD until the next exon emerges from the exit groove of the polymerase[24,34].

Transcription factors and coactivators: coupling to splicing

Further evidence that transcription is coupled to splicing is provided by recent reports of physical and in some cases functional interactions between the preinitiation complex and the SR protein family of splicing factors[35–38]. This family, which contains arginine/serine (RS) domains[39], binds to specific exon sequences known as splicing enhancers and recruits the splicing machinery to the 5′ and 3′ splice sites[39,40]. Thus SR proteins play a critical role in recognizing exons for inclusion in the spliced mRNA[40]. An example of an SR-like protein thought to have a role in coupling transcription to splicing is the transcriptional coactivator PGC-1. This coactivator, which contains both an RS domain and an RNA-binding domain, is recruited to promoters activated by PPARγ and other nuclear receptors[41]. Neither domain appears to be required for transcription but both are required for the accumulation of mRNAs transcribed from PGC-1-dependent promoters[37]. The PGC-1 protein colocalizes with splicing factors in the nucleus and interacts directly with SR proteins *in vitro*. This and related examples[35,36,38] suggest a common theme for transcription-coupled splicing in which transcription factors recruit SR proteins to the transcription machinery. The proximity of these proteins to the nascent pre-mRNA during transcription may in turn promote splicing.

Coupling of transcription to splicing was also revealed in studies showing that transcription of pre-mRNA by different promoters can generate different alternatively spliced mRNAs[42,43]. Two models which are not mutually exclusive have been proposed to explain this promoter-specific alternative splicing. In the first, particular

SR protein family members are differentially recruited to promoters and are then handed off to cognate splicing enhancers to promote inclusion of specific exons in the mRNA[43]. In the second, the pattern of splicing is determined by the rate of transcription elongation which is dictated in turn by the particular promoter[34,44]. In the second model different pre-mRNA secondary structures would be generated depending on the rate of transcription. The alternate structures could determine the availability of sequence elements recognized by the splicing machinery. This model is consistent with the observation that the pattern of alternative splicing can differ depending on whether the splicing machinery acts on a fully synthesized pre-mRNA, or whether splicing occurs co-transcriptionally[34,45]. In addition, transcriptional pausing or conditions that can decrease the rate of elongation can affect the pattern of alternative splicing[34,44,46].

Coupling capping, splicing and polyadenylation

As proposed for coupling between transcription and pre-mRNA processing, coupling among all of the different pre-mRNA processing steps also seems to be critical for gene expression. For example, proteins that bind to the 5′ cap of pre-mRNAs interact with splicing factors and promote recognition of the cap-proximal 5′ splice site[47,48]. Splicing factors that associate with the terminal 3′ intron also interact with downstream polyadenylation factors and promote 3′ end cleavage and polyadenylation[49,50]. Reciprocally, polyadenylation complexes can promote splicing of the upstream 3′ terminal intron[33]. These interactions, in conjunction with the recruitment of the splicing machinery to internal exons in the pre-mRNA[51], play an important role in ensuring that the correct splice sites are recognized and engaged in the splicing reaction, and that capping and polyadenylation are timely, accurate and efficient.

Coupling the earliest and latest steps of gene expression

Polyadenylation, transcription termination (release of RNA polymerase from the DNA) and release of the RNA from the site of transcription are the final steps in the production of mature mRNA. A premature event in any of these steps would result in the release of an incompletely synthesized or partially processed pre-mRNA. This problem seems to be circumvented by extensive coupling among the late steps in pre-mRNA synthesis. For example, both a functional polyadenylation signal and polyadenylation factors are required for normal transcription termination[52,53]. Moreover RNA polymerase II CTD truncations that are defective in polyadenylation are likewise defective in termination[14,54]. A number of models for coupling polyadenylation and termination have been proposed but a general view has not emerged, owing at least in part to differences in assay systems[2,52,55–57].

Recent studies have shown that transcription initiation and termination may also be coupled and a polyadenylation factor may mediate this coupling[58,59]. The human transcriptional coactivator protein PC4, which interacts directly with the polyadenylation factor CstF, provides an example. The yeast homologues of these proteins interact genetically, and both are required for efficient transcription termination[59]. On the basis of these observations, it was proposed that PC4 functions as an anti-terminator by inhibiting the termination activity of CstF during transcription elongation. PC4 and CstF would then dissociate from each other at the polyadenylation site, allowing the CstF to function in both polyadenylation and termination[59]. This speculative model is attractive because events that occur during the earliest steps of gene expression would ensure the correct temporal order of downstream steps in the pathway.

Subsequent to transcription termination, the mature mRNA is released from the site of synthesis and transported to nuclear pores. Even this release step is coupled to both upstream and downstream events[60]. For example, unlike wild-type pre-mRNAs, pre-mRNAs containing splicing mutations accumulate at the site of transcription. The mutant pre-mRNAs may form abortive splicing complexes that fail to release normally or splicing itself may be required for release. The coupling of release and mRNA export is also suggested by the observation that RNA transcripts accumulate at the site of synthesis in a number of mRNA export mutants[61]. Whether or not this is a direct effect of coupling remains to be established.

Coupling and the gene expression factory model

The physical coupling between transcription and pre-mRNA processing components may also play an important role in the assembly, storage, and intranuclear transport of different gene expression machines. Transcription and splicing factors are concentrated in two types of nuclear regions known as Cajal bodies and nuclear speckles[62,63]. Time course studies suggest that transcription and pre-mRNA processing components newly imported into the nucleus first localize to Cajal bodies, and are ultimately stored in the nuclear speckles[62,64]. To activate gene expression, transcription and pre-mRNA processing components move from speckles to the gene[65–67]. As with several of the coupling steps mentioned above this intranuclear transport of splicing components requires an intact CTD[67]. Although the mechanism of transport is not understood it appears that phosphorylation of SR proteins is required for their release from speckles[66], and the movement of SR proteins within the nucleus is rapid and ATP-independent[68]. Thus SR proteins diffuse freely in the nucleus, rapidly moving from storage depots to their sites of action. Coupling between machines in such an environment could have an essential role in coordinating their assembly and localization.

An intriguing question is whether the transcription machinery moves along the DNA carrying all of the coupled machines with it or is anchored in an immobilized 'gene expression factory'[14,69] and reels in the template DNA as the pre-mRNA is continuously extruded in a highly localized manner[70]. At present it is not possible to distinguish rigorously between these two models. However the immobilized gene expression factor model[70] has considerable appeal from the perspective of the extensive coupling in gene expression. Consistent with this model, there is evidence that both the transcription and splicing machineries associate with the nuclear substructure, and that newly synthesized pre-mRNA is concentrated in discrete nuclear foci[63,70]. These foci may correspond to the gene expression factories, containing all of the tightly coupled machines required to produce mature mRNA. The CTD would then function as the linchpin of these factories, playing a central role in the assembly and coordination of all of the machines.

A working model for coupling transcription and pre-mRNA processing based on the gene expression factory model is shown in Fig. 2. In this model capping (CAP), splicing (SF) and polyadenylation (pA) factors are recruited to the transcription preinitiation complex (PIC) (see Fig. 2a). At this stage transcription factors (TF) are bound to the unphosphorylated CTD. Upon transcription initiation the CTD is phosphorylated, the PIC is released and capping, splicing and polyadenylation components associate with the CTD (Fig. 2b). Shortly after pre-mRNA synthesis begins, the 5′ end is capped, and the capping machinery then dissociates from the CTD (Fig. 2b). As the DNA is reeled into the gene expression factory, and the nascent pre-mRNA is extruded through the exit channel of the RNA polymerase, splicing factors are systematically transferred from the CTD to the exons (Fig. 2c), where they recruit the rest of the splicing machinery. Splicing factors are also transferred via the elongation complex (not shown). The interaction between splicing factors bound to the CTD with those bound to the exon functions to tether the exon to the CTD until the next exon emerges from the exit pore of the polymerase (Fig. 2d). This tethering plays a role in recognizing exons, and in ensuring the correct 5′ and 3′ association of sequentially synthesized exons. The splicing factors transferred from the CTD to the nascent pre-mRNA

are rapidly replaced by new splicing factors which are optimally positioned for the appearance of the next exon. This reloading mechanism is consistent with electron tomography studies of Balbiani ring chromosomes showing that splicing factors are continuously added to and released from nascent transcripts during transcription elongation[71]. The splicing machinery associates with the exons in a 5′ to 3′ order as they are synthesized. However as indicated in the figure the covalent joining of exons does not necessarily proceed in a 5′ to 3′ direction owing to inherent differences in rates of splicing catalysis between each exon pair.

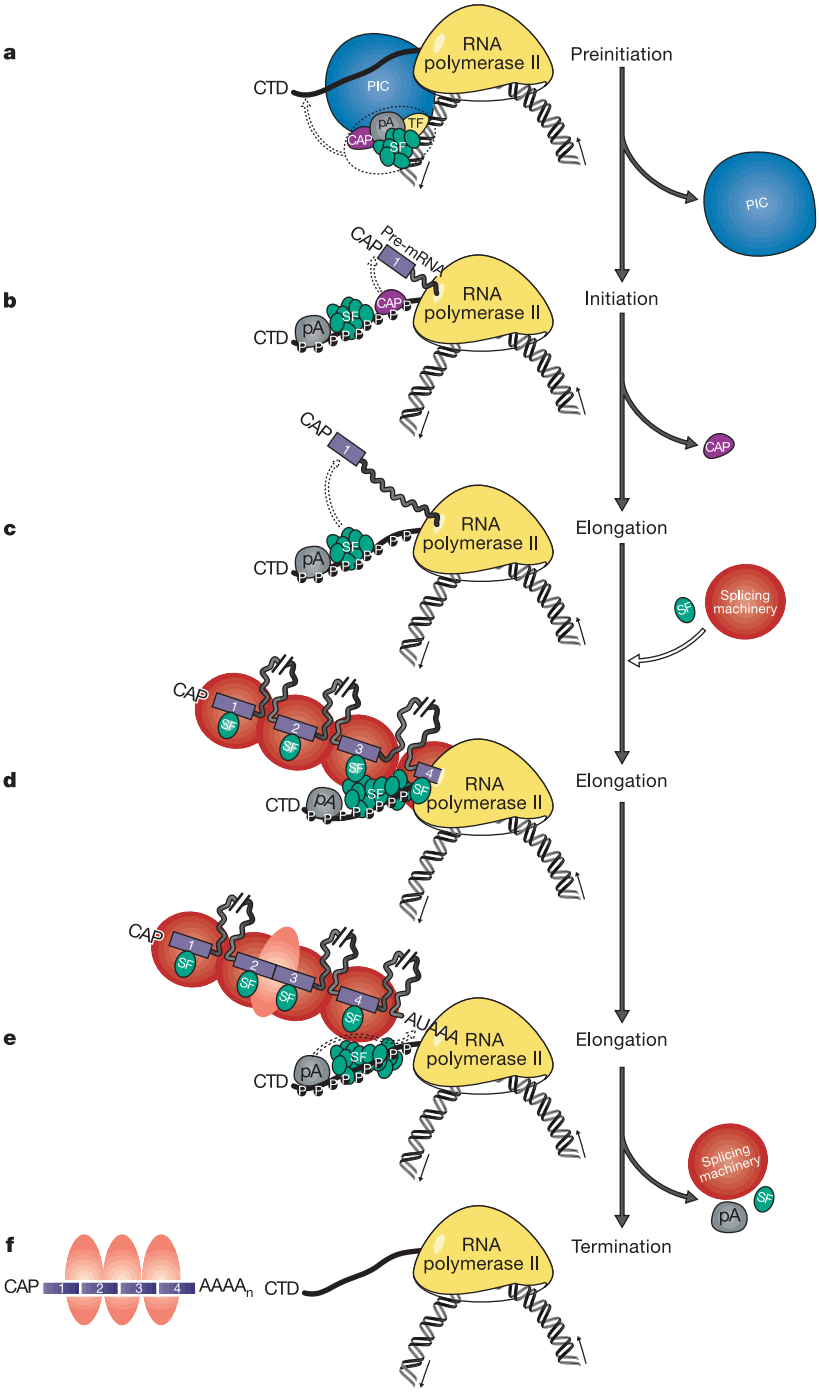

Figure 2 Gene expression factory model for coupling steps in gene expression. In this model the gene expression factory is anchored to the nuclear substructure and the DNA is reeled through the RNA polymerase as the nascent RNA is extruded through its exit channel. The machineries involved in transcription, capping, splicing and polyadenylation are shown. The shaded pink ovals over the spliced exons represent the mRNA complexes formed near the exon–exon junctions during the splicing reaction. See text for details of parts **a**–**f**. PIC, preinitiation complex; TF, transcription factors; CTD, carboxy-terminal domain; CAP, capping factor; SF, splicing factor; pA, polyadenylation factor; P, phosphorylated CTD.

Splicing leads to the formation of a specific complex of proteins on the spliced exons (messenger RNA–ribonucleoprotein (mRNP) complex, described below). When the polymerase approaches the end of the transcript, the polyadenylation factors function (Fig. 2e). Finally the CTD is dephosphorylated, the processing factors are released and the mature mRNP is released from the site of transcription (Fig. 2f).

Coupling splicing to mRNA export

Splicing occurs at the site of pre-mRNA synthesis before the mature mRNA is released for nuclear export. Despite the different nuclear locations of splicing and export, recent studies indicate that these two processes are directly coupled. Evidence for this coupling initially came from the observation that mRNAs generated by splicing are more efficiently exported than their identical counterparts transcribed from a complementary DNA[72]. This effect of splicing on export was explained by the finding that spliced mRNAs (but not cDNA transcripts) are assembled into a distinct mRNP complex that promotes efficient export[72]. This complex, or at least certain components of it, was subsequently shown to assemble adjacent to newly formed exon–exon junctions[73]. The increased export efficiency of the spliced mRNP is thought to be due to recruitment of the mRNA export factor ALY to the mRNA during the splicing reaction[74,75]. The splicing factor UAP56, which interacts directly with ALY, plays a role in recruiting ALY to the spliced mRNA[76–78]. The mRNP complex also contains a number of other factors[74,75,79–83], possibly including TAP[79], a protein that associates with the nuclear pore and is thought to be an mRNA export receptor[84,85]. Studies in yeast indicate that both TAP (Mex67) and UAP56 (Sub2) interact directly with the same region on ALY (Yra1)[77]. This observation together with other work suggests a model in which ALY is first recruited to the mRNA by interactions with UAP56 during the splicing reaction[76–78,86]. In a subsequent step, a hand-off occurs in which the ALY/TAP interaction is established, thus delivering the mRNP to the nuclear pore for export[77].

As mentioned above, metazoan introns are usually large, whereas exons are minute by comparison. Thus, specific mechanisms must exist to ensure that the vast amount of intron sequences are retained in the nucleus and degraded, while the relatively small amount of mature mRNA is exported. The coupling between splicing and export may provide another mechanism for this discrimination[87]. Specifically the mRNA is packaged into a splicing-dependent complex containing export factors while the large intron sequences are packaged into hnRNP complexes, which may play a role in nuclear retention. Thus, the coupling between splicing and mRNA export may provide a critical checkpoint for proper gene expression[87].

A small number of metazoan and most yeast pre-mRNAs lack introns. It is nevertheless possible that the same conserved mRNA export machinery described above is involved in intronless mRNA export, but is recruited by other mechanisms. In support of this notion, UAP56 is required for export of both spliced mRNAs and intronless mRNAs[76–78,86]. In metazoan intronless mRNAs, specific mRNA sequence elements are required for export. In one such mRNA, this element associates with members of the SR family of splicing factors, which are thought to mediate export of the intronless mRNA[88]. The SR proteins may either recruit the conserved export machinery or play a direct role in export themselves. In both yeast and metazoans the export of intronless mRNAs may also be coupled to polyadenylation[89].

The coupling between splicing and export provides an example of linking two machineries with spatially distinct locations, one at the site of transcription and the other at the nuclear pore. This long-range coupling is further illustrated by a recent study in yeast indicating that mRNA export is also coupled to transcription[90]. In this case Np13, an SR-like protein thought to function in export, was shown to interact genetically and physically with components of the transcription machinery. In addition, both Np13 and the mRNA export factor Yra1 (that is, yeast ALY) are thought to associate with the pre-mRNA during its synthesis[90]. Another example of long-range coupling is provided by the Hrp1 protein, which may associate with pre-mRNA near or at the site of transcription initiation[8]. This protein is also required for polyadenylation[91], mRNA export[92], and nonsense-mediated decay (NMD, see below)[93].

The coupling of export to both transcription and pre-mRNA processing requires that export machinery components bind to the nascent pre-mRNA at its site of synthesis and processing, and then remain bound until the mature mRNA reaches the nuclear pores. These interactions have several interesting implications. First, the proteins involved must form highly stable complexes with the pre-mRNA in order to remain bound throughout the entire pathway. Second, these proteins must also have high affinity for their ultimate targets at the nuclear pores so that they are recruited efficiently. Finally, mRNA export proteins bound to pre-mRNA so early in the gene expression pathway must be prevented from functioning at the incorrect time and exporting nascent pre-mRNA or pre-mRNA that is not fully processed. The binding of late-acting proteins at such early steps in gene expression would increase the efficiency and specificity of gene expression.

Coupling pre-mRNA processing to nonsense-mediated decay

During the course of mRNA production, specific RNA surveillance mechanisms operate to detect and eliminate defective pre-mRNA/mRNA[94]. The best characterized of these mechanisms is NMD, in which pre-mRNAs containing premature termination codons are targeted for degradation. This process was first linked to splicing by the observation that pre-mRNAs containing stop codons located greater than 50 nucleotides (nts) upstream from an intron are targeted for destruction[94]. This finding implicated the exon junction and hence splicing in NMD. The recent identification of the spliced mRNP complex[72], and the observation that it forms near exon–exon junctions[73,75,79,81,83], have provided new insights into the mechanism for coupling splicing to NMD. These studies revealed that in addition to the mRNA export factors ALY and TAP, the exon junction complex contains factors required for NMD[73,75,79,80,82,83], including hUPF3, the human homologue of an essential yeast NMD protein[94–96]. The hUPF3 protein is recruited to the exon junction complex by the splicing factor RNPS1 (ref. 81), and another protein called Y14 (ref. 83). Importantly, these interactions persist in the cytoplasm, thereby establishing a direct link between pre-mRNA splicing in the nucleus and NMD in the cytoplasm[79,81,82].

The exon junction-marking proteins (for example, Y14, RNPS1 and UPF3) play a critical role in the cytoplasmic degradation of transcripts containing a premature termination codon. These proteins, which are normally stripped from the mRNA during the first round of translation, instead remain bound because the movement of the ribosome is stopped by the premature termination codon[79,81,82]. The presence of these proteins then leads to the recruitment of factors that trigger rapid mRNA decay. Thus, it has been proposed that coupling between the splicing and NMD machineries carries a molecular memory of the intron to the cytoplasm[79,81,82]. A model for coupling splicing, mRNA export and NMD is shown in Fig. 3.

In addition to NMD in the cytoplasm, many pre-mRNAs containing nonsense mutations are eliminated in the nucleus before they can be exported to the cytoplasm[97]. Although the mechanism for this nuclear NMD is not well understood it appears that NMD is coupled to the release of the nascent pre-mRNA from the site of transcription[98]. Moreover, mRNAs containing premature termination codons may be subject to surveillance as they move from the nucleus to the cytoplasm[99].

In yeast, where most genes lack introns, specific RNA sequence

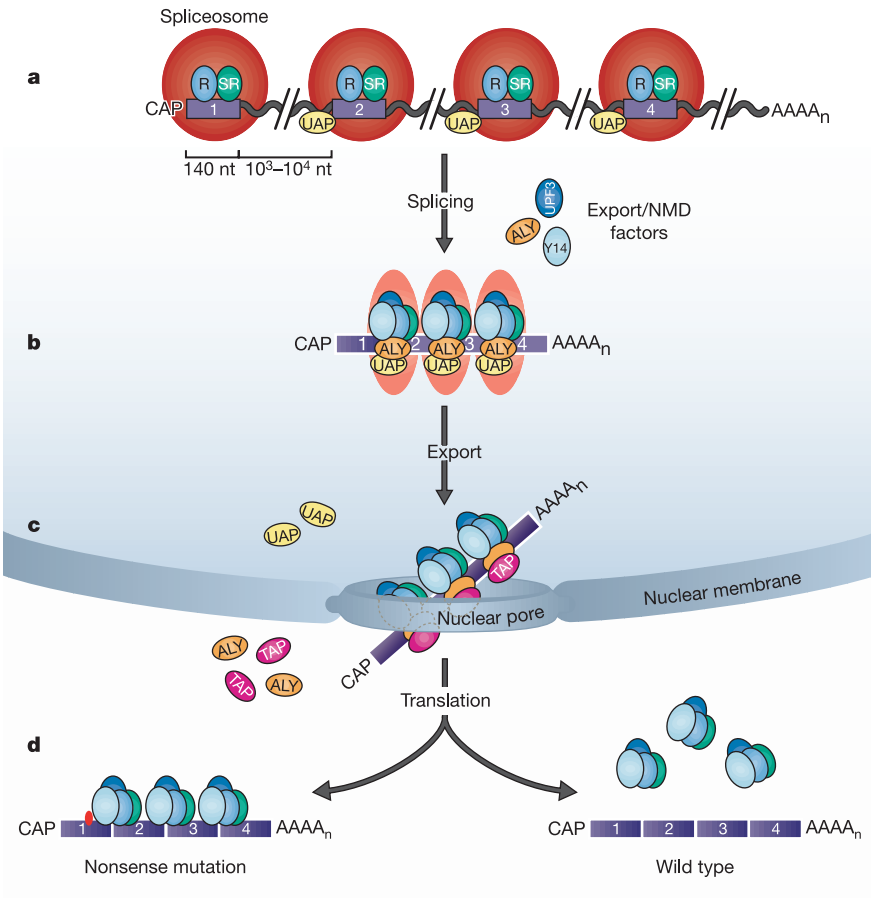

Figure 3 Model for coupling splicing to mRNA export and nonsense-mediated decay. In this model the splicing factor UAP56 (UAP, in yellow) associates with the pre-mRNA during spliceosome assembly along with other splicing factors including RNPS1 (R) and other SR proteins (**a**). UAP56 then recruits the mRNA export factor ALY along with NMD components including Y14 and UPF3 (**b**). All of these proteins form a complex that associates with the RNA near the exon–exon junction. UAP56 may then dissociate from ALY and be replaced by the export factor TAP, which mediates export of the mRNP through the nuclear pore complex (**c**). In the cytoplasm the export machinery is released and the NMD machinery remains associated with the exon–exon junctions. When the mRNA is wild type, the NMD machinery is displaced by the translating ribosome. In contrast, in the presence of a premature stop codon (represented by the red dot) (**d**) upstream from an exon–exon junction translation is arrested and the NMD machinery is not released but instead recruits additional components that trigger rapid degradation of the mRNA. nt, nucleotides.

elements called DSEs (downstream sequence elements) are required for NMD[100]. DSEs, which are located within protein coding sequences, act on upstream premature termination codons. Recently, Hrp1, the same protein implicated in both polyadenylation and mRNA export, was shown to bind specifically to DSEs, and to be required for NMD[93]. In addition, Hrp1 specifically interacts with other proteins required for NMD. Thus, similarly to the EJC complex, which marks intron-containing pre-mRNAs for export and NMD, the Hrp1 protein appears to form an export/NMD complex on DSEs[93]. Polyadenylation, mRNA export and NMD are therefore coupled through the multiple activities of the Hrp1 protein.

Discussion

The vast network of coupling discussed here and illustrated in Fig. 1 reveals that virtually every step in gene expression, from the earliest to the latest, is coupled. The emerging picture is that gene expression is carried out in immobilized gene expression factories consisting of a large number of interacting machines that orchestrate the multiple steps in the gene expression pathway. Understanding the inner workings of these factories will reveal an even greater number of interactions between the components of gene expression machines. □

1. Bentley, D. Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr. Opin. Cell Biol.* **11,** 347–351 (1999).
2. Hirose, Y. & Manley, J. L. RNA polymerase II and the integration of nuclear events. *Genes Dev.* **14,** 1415–1429 (2000).
3. Proudfoot, N. Connecting transcription to messenger RNA processing. *Trends Biochem. Sci.* **25,** 290–293 (2000).
4. Shatkin, A. J. & Manley, J. L. The ends of the affair: capping and polyadenylation. *Nature Struct. Biol.* **7,** 838–842 (2000).
5. Cramer, P. *et al.* Coordination between transcription and pre-mRNA processing. *FEBS Lett.* **498,** 179–182 (2001).
6. Ptashne, M. & Gann, A. *Genes and Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002).
7. Dahmus, M. E. Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *J. Biol. Chem.* **271,** 19009–19012 (1996).
8. Komarnitsky, P., Cho, E. J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14,** 2452–2460 (2000).
9. Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J. & Buratowski, S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.* **15,** 3319–3329 (2001).
10. Trigon, S. *et al.* Characterization of the residues phosphorylated *in vitro* by different C-terminal domain kinases. *J. Biol. Chem.* **273,** 6769–6775 (1998).
11. Price, D. H. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell. Biol.* **20,** 2629–2634 (2000).

12. McCracken, S. *et al.* 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 (1997).
13. Cho, E. J., Takagi, T., Moore, C. R. & Buratowski, S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **11,** 3319–3326 (1997).
14. McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385,** 357–361 (1997).
15. Fong, N. & Bentley, D. L. 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 (2001).
16. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292,** 1863–1876 (2001).
17. Ho, C. K. & Shuman, S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell* **3,** 405–411 (1999).
18. Schroeder, S. C., Schwer, B., Shuman, S. & Bentley, D. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.* **14,** 2435–2440 (2000).
19. Cho, E. J., Rodriguez, C. R., Takagi, T. & Buratowski, S. Allosteric interactions between capping enzyme subunits and the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **12,** 3482–3487 (1998).
20. Taube, R., Lin, X., Irwin, D., Fujinaga, K. & Peterlin, B. M. Interaction between P-TEFb and the C-terminal domain of RNA polymerase II activates transcriptional elongation from sites upstream or downstream of target genes. *Mol. Cell. Biol.* **22,** 321–331 (2002).
21. Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D. & Handa, H. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription *in vitro*. *EMBO J.* **17,** 7395–7403 (1998).
22. Kim, J. B. & Sharp, P. A. Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *J. Biol. Chem.* **276,** 12317–12323 (2001).
23. Wen, Y. & Shatkin, A. J. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* **13,** 1774–1779 (1999).
24. Fong, Y. W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414,** 929–933 (2001).
25. Barboric, M., Nissen, R. M., Kanazawa, S., Jabrane-Ferrat, N. & Peterlin, B. M. NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol. Cell* **8,** 327–337 (2001).
26. Kim, J. B., Yamaguchi, Y., Wada, T., Handa, H. & Sharp, P. A. Tat-SF1 protein associates with RAP30 and human SPT5 proteins. *Mol. Cell. Biol.* **19,** 5960–5968 (1999).
27. Yan, D. *et al.* CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. *Mol. Cell. Biol.* **18,** 5000–5009 (1998).
28. Ares, M. Jr, Grate, L. & Pauling, M. H. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **5,** 1138–1139 (1999).
29. Robert, F., Blanchette, M., Maes, O., Chabot, B. & Coulombe, B. A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly. *J. Biol. Chem.* **277,** 9302–9306 (2002).
30. Corden, J. L. & Patturajan, M. A CTD function linking transcription to splicing. *Trends Biochem. Sci.* **22,** 413–416 (1997).
31. Morris, D. P. & Greenleaf, A. L. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **275,** 39935–39943 (2000).
32. Abovich, N. & Rosbash, M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89,** 403–412 (1997).
33. Berget, S. M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270,** 2411–2414 (1995).
34. Goldstrohm, A. C., Greenleaf, A. L. & Garcia-Blanco, M. A. Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene* **277,** 31–47 (2001).
35. Lai, M. C., Teh, B. H. & Tarn, W. Y. A human papillomavirus E2 transcriptional activator. The interactions with cellular splicing factors and potential function in pre-mRNA processing. *J. Biol. Chem.* **274,** 11832–11841 (1999).
36. Ge, H. Si, Y. & Wolffe, A. P. A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. *Mol. Cell* **2,** 751–759 (1998).
37. Monsalve, M. *et al.* Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. *Mol. Cell* **6,** 307–316 (2000).
38. Martinez, E. *et al.* Human staga complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors *in vivo*. *Mol. Cell. Biol.* **21,** 6782–6795 (2001).
39. Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA* **6,** 1197–1211 (2000).
40. Graveley, B. R., Hertel, K. J. & Maniatis, T. SR proteins are 'locators' of the RNA splicing machinery. *Curr. Biol.* **9,** R6–R7 (1999).
41. Puigserver, P. *et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92,** 829–839 (1998).
42. Cramer, P., Pesce, C. G., Baralle, F. E. & Kornblihtt, A. R. Functional association between promoter structure and transcript alternative splicing. *Proc. Natl Acad. Sci. USA* **94,** 11456–11460 (1997).
43. Cramer, P. *et al.* Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol. Cell* **4,** 251–258 (1999).
44. Kadener, S. *et al.* Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *EMBO J.* **20,** 5759–5768 (2001).
45. Eperon, L. P., Graham, I. R., Griffiths, A. D. & Eperon, I. C. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell* **54,** 393–401 (1988).
46. Roberts, G. C., Gooding, C., Mak, H. Y., Proudfoot, N. J. & Smith, C. W. Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res.* **26,** 5568–5572 (1998).
47. Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. & Mattaj, I. W. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5′ splice site. *Genes Dev.* **10,** 1683–1698 (1996).
48. Colot, H. V., Stutz, F. & Rosbash, M. The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes Dev.* **10,** 1699–1708 (1996).
49. Vagner, S., Vagner, C. & Mattaj, I. W. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3′-end processing and splicing. *Genes Dev.* **14,** 403–413 (2000).
50. McCracken, S., Lambermon, M. & Blencowe, B. J. SRm160 splicing coactivator promotes transcript 3′-end cleavage. *Mol. Cell. Biol.* **22,** 148–160 (2002).
51. Reed, R. Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12,** 340–345 (2000).
52. Proudfoot, N. J. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.* **14,** 105–110 (1989).
53. Birse, C. E., Minvielle-Sebastia, L., Lee, B. A., Keller, W. & Proudfoot, N. J. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* **280,** 298–301 (1998).
54. McCracken, S. *et al.* Role of RNA polymerase II carboxy-terminal domain in coordinating transcription with RNA processing. *Cold Spring Harbor Symp. Quant. Biol.* **63,** 301–309 (1998).
55. Dye, M. J. & Proudfoot, N. J. Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* **105,** 669–681 (2001).
56. Osheim, Y. N., Proudfoot, N. J. & Beyer, A. L. EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol. Cell* **3,** 379–387 (1999).
57. Tran, D. P., Kim, S. J., Park, N. J., Jew, T. M. & Martinson, H. G. Mechanism of poly(A) signal transduction to RNA polymerase II *in vitro*. *Mol. Cell. Biol.* **21,** 7495–7508 (2001).
58. Aranda, A. & Proudfoot, N. Transcriptional termination factors for RNA polymerase II in yeast. *Mol. Cell* **7,** 1003–1011 (2001).
59. Calvo, O. & Manley, J. L. Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol. Cell* **7,** 1013–1023 (2001).
60. Custodio, N. *et al.* Inefficient processing impairs release of RNA from the site of transcription. *EMBO J.* **18,** 2855–2866 (1999).
61. Jensen, T. H., Patricio, K., McCarthy, T. & Rosbash, M. A block to mRNA nuclear export in *S. cerevisiae* leads to hyperadenylation of transcripts that accumulate at the site of transcription. *Mol. Cell* **7,** 887–898 (2001).
62. Gall, J. G. A role for Cajal bodies in assembly of the nuclear transcription machinery. *FEBS Lett.* **498,** 164–167 (2001).
63. Dundr, M. & Misteli, T. Functional architecture in the cell nucleus. *Biochem. J.* **356,** 297–310 (2001).
64. Sleeman, J. E. & Lamond, A. I. Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.* **9,** 1065–1074 (1999).
65. Misteli, T., Caceres, J. F. & Spector, D. L. The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387,** 523–527 (1997).
66. Misteli, T. *et al.* Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription *in vivo*. *J. Cell. Biol.* **143,** 297–307 (1998).
67. Misteli, T. & Spector, D. L. RNA polymerase II targets pre-mRNA splicing factors to transcription sites *in vivo*. *Mol. Cell* **3,** 697–705 (1999).
68. Phair, R. D. & Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* **404,** 604–609 (2000).
69. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *J. Cell Sci.* **109,** 1427–1436 (1996).
70. Cook, P. R. The organization of replication and transcription. *Science* **284,** 1790–1795 (1999).
71. Wetterberg, I., Zhao, J., Masich, S., Wieslander, L. & Skoglund, U. *In situ* transcription and splicing in the Balbiani ring 3 gene. *EMBO J.* **20,** 2564–2574 (2001).
72. Luo, M. J. & Reed, R. Splicing is required for rapid and efficient mRNA export in metazoans. *Proc. Natl Acad. Sci. USA* **96,** 14937–14942 (1999).
73. Le Hir, H., Moore, M. J. & Maquat, L. E. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon–exon junctions. *Genes Dev.* **14,** 1098–1108 (2000).
74. Zhou, Z. *et al.* The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407,** 401–405 (2000).
75. Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. *EMBO J.* **19,** 6860–6869 (2000).
76. Luo, M. L. *et al.* Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413,** 644–647 (2001).
77. Strasser, K. & Hurt, E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* **413,** 648–652 (2001).
78. Gatfield, D. *et al.* The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. *Curr. Biol.* **11,** 1716–1721 (2001).
79. Le Hir, H., Gatfield, D., Izaurralde, E. & Moore, M. J. The exon–exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20,** 4987–4997 (2001).
80. Kataoka, N. *et al.* Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol. Cell* **6,** 673–682 (2000).
81. Lykke-Andersen, J., Shu, M. D. & Steitz, J. A. Communication of the position of exon–exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293,** 1836–1839 (2001).
82. Kim, V. N. *et al.* The Y14 protein communicates to the cytoplasm the position of exon-exon junctions. *EMBO J.* **20,** 2062–2068 (2001).
83. Kim, V. N., Kataoka, N. & Dreyfuss, G. Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon–exon junction complex. *Science* **293,** 1832–1836 (2001).
84. Bachi, A. *et al.* The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. *RNA* **6,** 136–158 (2000).
85. Segref, A. *et al.* Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. *EMBO J.* **16,** 3256–3271 (1997).
86. Jensen, T. H., Boulay, J., Rosbash, M. & Libri, D. The DECD box putative ATPase Sub2p is an early mRNA export factor. *Curr. Biol.* **11,** 1711–1715 (2001).
87. Reed, R. & Magni, K. A new view of mRNA export: Separating the wheat from the chaff. *Nature Cell Biol.* **3,** E201–E204 (2001).
88. Huang, Y. & Steitz, J. A. Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol. Cell* **7,** 899–905 (2001).
89. Hilleren, P., McCarthy, T., Rosbash, M., Parker, R. & Jensen, T. H. Quality control of mRNA 3′-end processing is linked to the nuclear exosome. *Nature* **413,** 538–542 (2001).
90. Lei, E. P., Krebber, H. & Silver, P. A. Messenger RNAs are recruited for nuclear export during transcription. *Genes Dev.* **15,** 1771–1782 (2001).
91. Gross, S. & Moore, C. L. Rna15 interaction with the a-rich yeast polyadenylation signal is an essential step in mRNA 3′-end formation. *Mol. Cell. Biol.* **21,** 8045–8055 (2001).

92. Kessler, M. M. *et al.* Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3′-end formation in yeast. *Genes Dev.* **11,** 2545–2556 (1997).
93. Gonzalez, C. I., Ruiz-Echevarria, M. J., Vasudevan, S., Henry, M. F. & Peiltz, S. W. The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. *Mol. Cell* **5,** 489–499 (2000).
94. Maquat, L. E. & Carmichael, G. G. Quality control of mRNA function. *Cell* **104,** 173–176 (2001).
95. Serin, G., Gersappe, A., Black, J. D., Aronoff, R. & Maquat, L. E. Identification and characterization of human orthologues to *Saccharomyces cerevisiae* Upf2 protein and Upf3 protein (*Caenorhabditis elegans* SMG-4). *Mol. Cell. Biol.* **21,** 209–223 (2001).
96. Lykke-Andersen, J., Shu, M. D. & Steitz, J. A. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **103,** 1121–1131 (2000).
97. Hentze, M. W. & Kulozik, A. E. A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* **96,** 307–310 (1999).
98. Muhlemann, O. *et al.* Precursor RNAs harboring nonsense codons accumulate near the site of transcription. *Mol. Cell* **8,** 33–43 (2001).
99. Ishigaki, Y., Li, X., Serin, G. & Maquat, L. E. 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 (2001).
100. Czaplinski, K., Ruiz-Echevarria, M. J., Gonzalez, C. I. & Peltz, S. W. Should we kill the messenger? The role of the surveillance complex in translation termination and mRNA turnover. *Bioessays* **21,** 685–696 (1999).

Acknowledgements

We thank R. Axel, D. Bentley, S. Buratowski, B. Graveley, J. Manley, M. Ptashne and members of our labs for their comments on the manuscript. We also thank R. Hellmiss for the illustrations.

Correspondence and requests for materials should be addressed to T.M. (e-mail: maniatis@mcb.harvard.edu) or R.R. (e-mail: rreed@hms.harvard.edu).