Review

Author for correspondence:
Shona Murphy
e-mail: shona.murphy@path.ox.ac.uk

Regulation of expression of human RNA polymerase II-transcribed snRNA genes

Joana Guiro and Shona Murphy

Sir William Dunn School of Pathology, University of Oxford, Oxford, UK

JG, 0000-0002-3705-3515

In addition to protein-coding genes, RNA polymerase II (pol II) transcribes numerous genes for non-coding RNAs, including the small-nuclear (sn)RNA genes. snRNAs are an important class of non-coding RNAs, several of which are involved in pre-mRNA splicing. The molecular mechanisms underlying expression of human pol II-transcribed snRNA genes are less well characterized than for protein-coding genes and there are important differences in expression of these two gene types. Here, we review the DNA features and proteins required for efficient transcription of snRNA genes and co-transcriptional 3′ end formation of the transcripts.

1. Introduction

RNA polymerase II (pol II) is responsible for the transcription of all 25 000 or so protein-coding genes in the human genome. Many non-coding RNAs, including ribosomal (r)RNA, transfer (t)RNA, the 7SK RNA that regulates transcription elongation by pol II, and the spliceosomal U6 and U6atac small nuclear (sn)RNAs are transcribed by pol I or pol III [1–4]. However, the non-coding micro-RNAs, small nucleolar (sno)RNAs and the remaining snRNAs are transcribed by pol II [2,5,6]. SnRNAs are short RNAs (less than 350 nts) that associate with proteins to form small nuclear ribonucleoprotein particles (snRNPs) [6].

Due to their roles in pre-mRNA splicing, the pol II-transcribed U1, U2, U4, U4atac, U5, U11 and U12 snRNAs are required for expression of intron-containing protein-coding genes. The major spliceosome includes the U1, U2, U4, U5 and U6 snRNPs and recognizes the canonical GT/AG splice sites flanking introns to assemble in a stepwise manner with the pre-mRNA. Some introns are instead flanked by AT/AC splice sites, which recruit the minor spliceosome containing instead U11, U12, U4atac, U5 and U6atac snRNPs [7]. U7 snRNA is required for 3′ end formation of replication-activated histone mRNA [8] and the U3 sn(o)RNA is required for processing of rRNA [9].

2. DNA sequence elements involved in expression of pol II-transcribed snRNA genes

Although the functions of snRNAs are well understood, the regulation of their expression is still not fully characterized. The human snRNA genes have a much simpler promoter structure than most protein-coding genes; this comprises a distal sequence element (DSE) which acts as an enhancer and an essential snRNA gene-specific proximal sequence element (PSE) which is the core promoter [2,10,11] (figure 1). The transcripts are intronless and non-polyadenylated, and a 3′ box directs formation of the 3′ end of the pre-snRNA, which is further processed to produce the mature snRNA [10].

In vertebrates the genes for the major snRNAs (U1, U2, U4 and U5) are well conserved with recognizable DSE, PSE and 3′ box sequences [2], and *Drosophila* snRNA genes also have PSE-like sequences [16]. The vertebrate snRNA genes are often present in multiple copies. For example, there are four copies of the
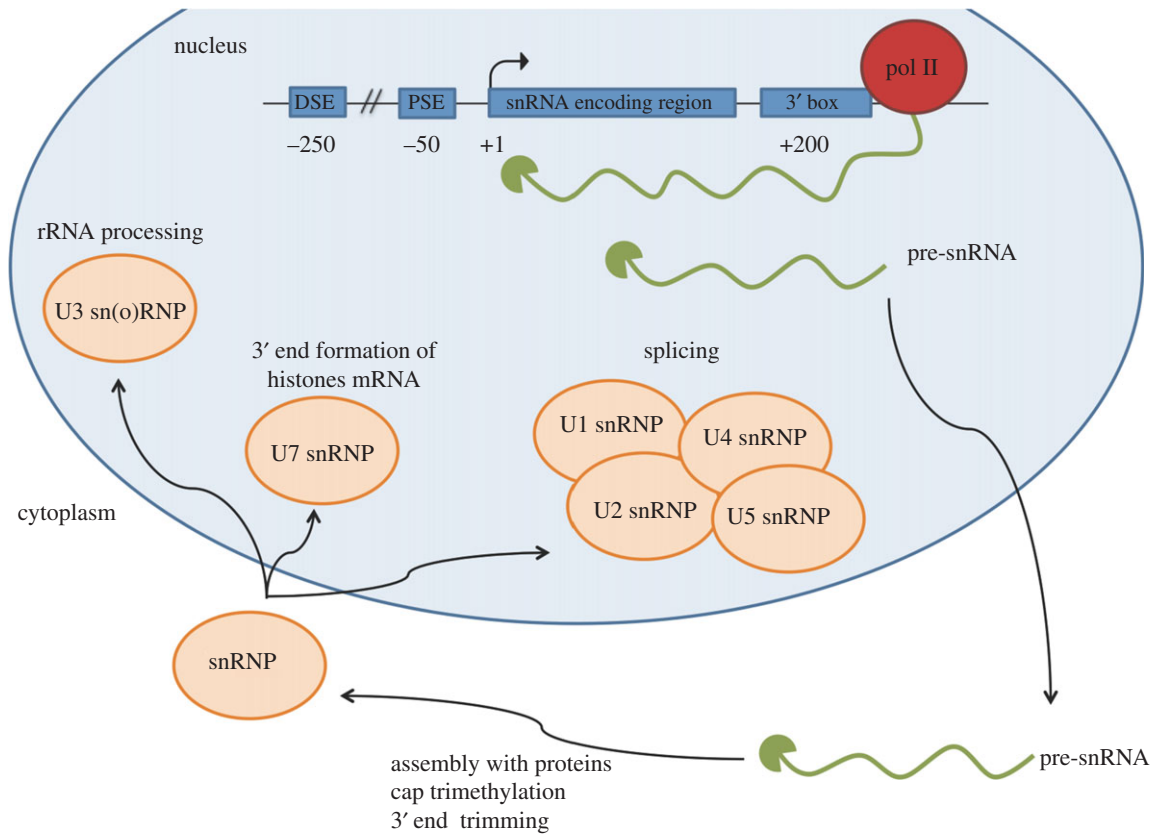
Figure 1. Expression of human snRNA genes. Distal sequence element (DSE) and proximal sequence element (PSE) are the enhancer and promoter elements, respectively. The 3′ box is the snRNA gene-specific RNA processing element. An arrow represents the start site of transcription and the numbers below the line indicate the approximate position of the elements with respect to the transcription start site. DSE comprises the Oct1 binding site ATTTGCAT [12], three Sp1 binding sites (G/T)(G/A)GGCG(G/T)(G/A)(G/A)(G/T) [13] and a STAF binding site YY(A/T)CCC(A/G)N(A/C)AT(G/C)C(A/C)YYRCR [14]. The PSE sequence consensus sequence is TCACCNTNA(G/C)NNNAA(A/T)(G/A)N [2]. The consensus sequence of the 3′ box is GTTYN$_{0-3}$AARRYAGA [15]. The snRNA transcript is represented in green with the cap in the 5′ end. The pre-snRNA is exported to the cytoplasm for maturation by 3′ end trimming, cap trimethylation and assembly with the snRNP proteins [6]. The functions of the various snRNPs after reimport into the nucleus are noted.

U1 snRNA gene/human haploid genome) [17] and 15 copies of the U2 snRNA gene/human haploid genome) [18]. In contrast, the genes for the minor snRNAs (U11, U12 and U4atac) are usually single-copy genes and are missing in many invertebrates including Cnidaria and Annelida [19].

An important feature of snRNA genes is the obligatory coupling between the 3′ box RNA processing element and the snRNA gene-type promoter; replacement of an snRNA gene promoter with the promoter of a protein-coding gene results in failure of RNA 3′ end formation [15,20].

The pol III-transcribed genes for 7SK snRNA and U6 snRNA also have a DSE and PSE, but in addition they have a TATA box at −25 [2]. Placing a TATA box downstream of the PSE in a pol II-transcribed snRNA gene converts transcription from pol II to pol III [2], emphasizing that the DSE and PSE can work as cis-acting elements for either polymerase.

Here, we review what is currently known about expression of human snRNA genes transcribed by pol II and the future prospects of a complete understanding of the mechanisms involved.

3. Transcription factors associated with snRNA genes

Transcription factors Oct1, Sp1, NF1 and Staf bind to sequences in the DSE, which has the properties of a transcriptional enhancer [2,21]. Oct1 enhances snRNA gene expression by stabilizing binding of the PSE-binding protein/PSE-binding transcription factor/snRNA activating protein complex, PTF (also known as PBP and SNAPc) [12,22,23], to the PSE of both pol II- and pol II-transcribed snRNA genes through direct interaction [12,24,25]. In turn, PTF helps to recruit the TATA-binding protein (TBP) and the TBP-associated factors (TAFs) that make up the snRNA TAF complex (snTAFc) to the pol II-transcribed snRNA genes [22,26,27]. The snTAFc on the U2 snRNA gene, which comprises TAF5, TAF6, TAF8, TAF9, TAF11 and TAF13 [27], is a subset of the TAFs found in TFIID, the TBP/TAF-containing complex required for transcription of many protein-coding genes [28]. However, the U1, U4, U5 and U11 snRNA genes appear to have a different TAF complex, which includes TAF7 [29]. TAF7 may help to recruit the large multisubunit Mediator complex, which is required for efficient transcription of snRNA genes [29]. TAF7 has also been implicated in regulation of promoter escape (see below). However, TAF7 appears to be absent from the U2 snRNA genes [27], suggesting that these genes employ a different mechanism of Mediator recruitment and promoter escape. The general protein-coding gene transcription factors TFIIA, TFIIB, TFIIE, TFIIF and TFIIH are also required for transcription of snRNA genes [30,31]. Figure 2 indicates which factors are shared between snRNA genes and protein-coding genes and which are specific to each class of gene. As noted above, the DSE and PSE are interchangeable
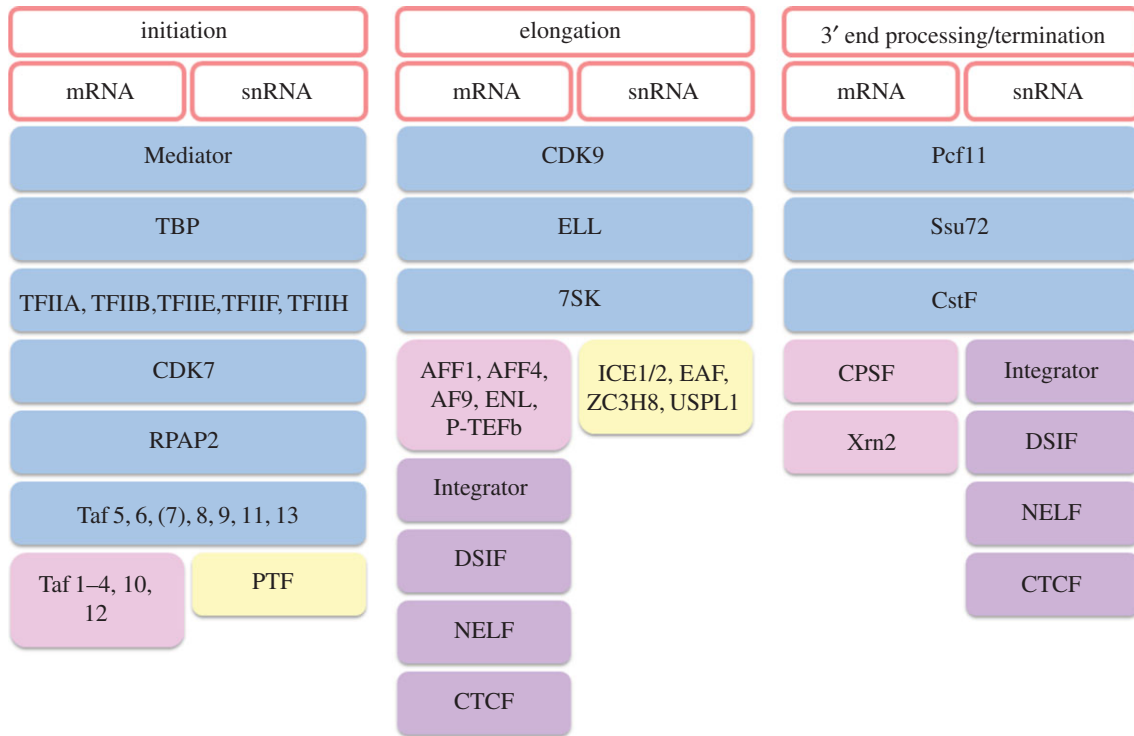
Figure 2. Factors regulating expression of snRNA and protein-coding genes. In blue are the factors that have roles in the same event of transcription in both classes of genes. In pink and yellow are the factors unique to protein-coding genes and snRNA genes, respectively. Factors that regulate both genes but at different steps are shown in purple [22,24,26 – 55].

between the pol II- and pol III-transcribed snRNA genes [2]. However, in the presence of a TATA box in the promoter, a pol III- and snRNA gene-specific TBP-containing complex is recruited [2].

There is evidence that a nucleosome between the DSE and PSE in the U1, U2, 7SK and U6 snRNA genes brings those two regions together, facilitating the interaction between Oct1 and PTF [56–59]. In addition, PTF plays a role in depletion of histones from the transcription unit of the U2 gene [32]. Thus, chromatin structure plays a role in the regulation of expression of human snRNA genes. Expression of the pol III-transcribed U6 snRNA gene is regulated by interaction of a protein involved in chromatin modification, the chromodomain-helicase-DNA binding protein 8 (CHD8), with Staf [60]. As CHD8 has also been shown to interact with the elongating form of pol II [60], it may also regulate transcription of pol II-transcribed snRNA genes.

4. Pol II carboxyl terminal domain phosphorylation and snRNA gene expression

The largest subunit of mammalian pol II has a carboxyl terminal domain (CTD) comprising 52 repeats of the consensus heptapeptide $Y_1S_2P_3T_4S_5P_6S_7$. During transcription, the phosphorylation of the CTD helps to recruit transcription and RNA processing factors at the right point of the transcription cycle [61].

Ser7 is phosphorylated during transcription of all human genes. However its phosphorylation appears to be essential only for expression of snRNA genes [33]. Ser5 and Ser7 are both phosphorylated by the cyclin-dependent kinase 7

(CDK7) subunit of TFIIH soon after initiation of transcription [34] and phospho-Ser7 (Ser7P) mediates recruitment of the phospho-Ser5 (Ser5P) phosphatase RPAP2 [62]. RPAP2 not only dephosphorylates Ser5P soon after initiation of transcription but also helps to recruit subunits Int1, Int4, Int5, Int6 and Int7 of the large multisubunit Integrator complex [62]. The complete Integrator complex comprises 14 subunits and is responsible for the recognition of the 3' box and RNA cleavage that produces pre-snRNA [35]. However, the catalytic subunit Int11 is absent when Integrator is associated with RPAP2 [62], and Int9 is also likely to be absent as it interacts strongly with Int11 [63]. The CDK9 kinase subunit of the positive-transcription elongation factor b (P-TEFb), which is involved in the expression of both protein-coding genes and of snRNA genes, phosphorylates Ser2 (Ser2P) of the pol II CTD [61] and Ser7 and Ser2 phosphorylation together recruit the missing subunits Int9/11 to activate 3' box recognition and RNA 3' end processing [64]. Accordingly, CDK9 inhibitors have a drastic effect on 3' end processing of snRNA gene transcripts [36]. This sequence of events is shown in figure 3.

Ser2 can also be phosphorylated by CDK12 [65,66] and CDK12 knockdown affects snRNA gene expression [67]. Phosphorylation of the DRB-sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) by P-TEFb are also required for productive elongation of protein-coding genes [37,38,68]. However, CDK9 inhibitors have little effect on transcription of the U1 and U2 genes [39].

After initiation of transcription of protein-coding genes, the capping machinery is recruited to the 5' end of the nascent transcript and allosterically activated by Ser5P [69]. As snRNA gene transcripts are also co-transcriptionally capped [70], it is likely that Ser5P also ensures capping of pre-snRNAs.
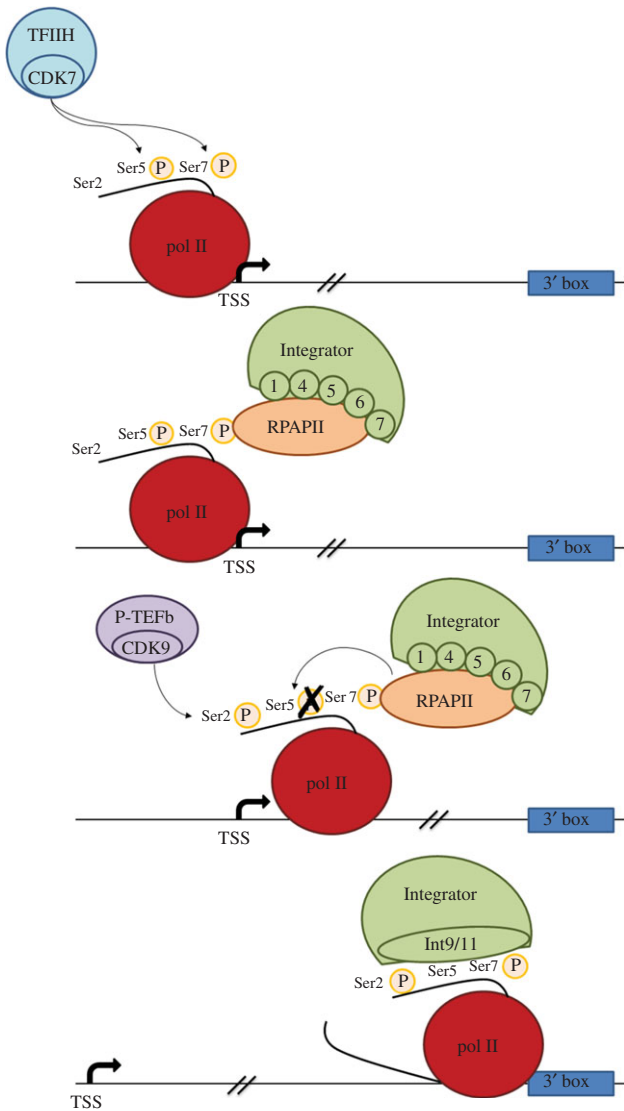
Figure 3. Pol II CTD phosphorylation events in snRNA gene transcription. Initially, the cyclin-dependent kinase (CDK)7 subunit of TFIIH phosphorylates Ser5 and Ser7. RPAP2 interacts with Ser7P. RPAP2 in turn recruits the Integrator subunits Int1, Int4, Int5, Int6 and Int7. RPAP2 dephosphorylates Ser5P, and positive-transcription elongation factor b (P-TEFb) is recruited by a mechanism still unknown. The P-TEFb subunit CDK9 phosphorylates Ser2. The double phosphorylation on Ser2 and Ser7 recruits Int9/11, which allows RNA processing to occur [62].

5. The role of Mediator in expression of snRNA genes

Association of the large 26-subunit Mediator complex with protein-coding genes has been well studied. This complex is recruited to the pre-initiation complex (PIC) and can bind to several transcription factors at the same time. Mediator has an important role as a binding platform for interaction between the transcription factors bound to sequences in protein-coding gene promoters and pol II [40].

The Mediator complex is made up of modules, with the head module comprising the subunits Med6, Med8, Med11, Med17, Med18, Med20 and Med22 [71]. The head module is responsible for the interaction with the pol II CTD through Med17 [72]. Med17 is a subunit on the 'nose' of the head module whereas Med18 and Med20 are located on the 'jaw'. A mutation on the alpha helix of Med18 is predicted to strengthen the interaction between the jaw and the rest of the module. The jaw is therefore thought to be the moveable element that regulates binding to the pol II CTD [71]. This flexibility may also have a role in interactions between Mediator and pol II and also between Mediator and transcription factors such as TFIIH [73] and TBP [74].

The association of Mediator subunits with snRNA genes was only described for the first time recently. Med1, Med23 and Med26 are associated with a subset of snRNA genes (U1, U4 and U5) and in significantly higher levels than in the promoter region of some protein-coding genes [29]. However, the ChIP experiments did not define the region of the snRNA genes associated with Mediator. Med26 is required for the expression of U1, U2, U4, U5 and U11 genes [29] and is essential for the recruitment of the little elongation complex (LEC) [29]. Interestingly, Med26 also interacts with TAF7, which represses initiation and blocks LEC recruitment. In addition, TAF7 has protein sequence similarities to EAF, which is part of the LEC. Accordingly, it has been suggested that EAF could snatch Med26 from TAF7 to facilitate elongation [29]. Med26 may therefore act as a molecular switch triggering elongation after promoter escape, at least for the snRNA genes where TAF7 is recruited.

6. The little elongation complex and elongation of transcription

Gene expression is regulated not only at transcription initiation, but also during elongation. Pol II pauses soon after initiation of transcription of protein-coding genes at an early elongation checkpoint (EEC) due to the negative elongation factors, NELF and DSIF. Phosphorylation of the pol II CTD, NELF and DSIF by the CDK9 subunit of P-TEFb releases pol II from the EEC to allow productive elongation [68]. A complex named the super elongation complex (SEC) comprising the elongation factor ELL, AFF1, AFF4, AF9, ENL and P-TEFb is involved in productive elongation of transcription of protein-coding genes [41].

Instead of the SEC, another ELL-containing complex specifically regulates elongation of transcription of pol II-transcribed snRNA genes. This LEC comprises ELL, ICE1, ECE2, EAF and ZC3H8 [42]. It has been suggested that ICE1, ICE2, ELL and EAF are recruited to the promoter by the Mediator complex [29]. ICE1 has been shown to be the scaffold for LEC formation and is essential for recruitment of the other components. ELL interacts with the ICE1 N-terminal domain, whereas ICE2 and ZC3H8 interact with the ICE1 C-terminal domain. In addition, ICE1 is required for pol II recruitment [42]. ELL knockdown shifts the distribution of pol II on snRNA genes towards the 5′ end, as expected of a factor that facilitates transcription elongation [75]. It has been suggested that ELL is necessary to ensure a high level of transcription of snRNA genes [75]. The function of subunits ICE2 and ZC3H8 is currently unknown.

The LEC has been shown to co-localize with coilin [42], which is found in Cajal bodies, where small nuclear ribonucleoprotein (snRNP) maturation takes place [6]. ICE1 recruits another protein, USPL1, to Cajal bodies. USPL1 also interacts with U1 and U2 snRNPs and its knockdown causes a reduction in snRNA levels [43]. However, the molecular mechanism underlying the effect of USPL1 knockdown is unclear.

More recently, it has been shown that the 7SK snRNP, comprising 7SK RNA, MePCE and Larp7, plays a role in the recruitment of the LEC to snRNA genes. Knockdown of 7SK RNP components disrupts LEC integrity, affects pol II recruitment and reduces expression of snRNA genes [76]. This was a surprising finding as the 7SK RNP is a negative regulator of the CTD Ser2 kinase P-TEFb complex [68].

7. The Integrator complex and snRNA genes

The Integrator complex was discovered more recently than Mediator. It is present only in metazoans and is a large complex of 14 subunits and is required for 3′ end processing of pol II-transcribed pre-snRNAs in humans and flies [33,35,44].

Integrator is able to bind the pol II CTD phosphorylated on both Ser2 and Ser7 [64]. However, it is still unclear which subunits are involved in Ser2P and Ser7P recognition. Int11 is a paralogue of CPSF-73, which is responsible for RNA cleavage directed by poly(A) sites and the RNA processing signal in the replication-activated histone pre-mRNAs [77]. Depletion of Int11 causes accumulation of the pre-U1 and pre-U2 snRNAs [32,35], which is in agreement with the expected catalytic role of this subunit.

So far, Integrator subunits have not been found to interact directly with chromatin. However, subunit Int12 has a PHD (plant homeodomain) domain, which is a common chromatin-binding motif that recognizes methylated histones [78]. Subunits Int3 and Int6 are also part of another complex involved in detecting DNA double strand breaks, called sensor of single stranded DNA (SOSS) [79–83]. As snRNA genes are highly expressed and at least the U2 snRNA gene is depleted of canonical histones [45], it has been suggested that these Integrator subunits could help to prevent DNA damage that is facilitated by a constitutively open chromatin state [84].

Int5 is found associated with both the snRNA gene promoter region and the 3′ box, whereas Int11 is associated mainly with the 3′ box [62]. Integrator may therefore be a modular complex that assembles in a stepwise manner, with some subunits playing a role early in transcription, while others are active in RNA 3′ end processing.

It has been demonstrated that Int4 interacts with NELF-A and Int6 interacts with the Spt5 subunit of DSIF [46]. In expression of snRNAs genes, DSIF has been proposed to bind pol II prior to transcription elongation [46]. Integrator could therefore be recruited both by pol II CTD phosphorylation and by association with DSIF. Recruitment of NELF through interaction with RNA and Integrator could contribute to specific 3′ box recognition by preventing recruitment of the cleavage stimulatory factor (CstF) polyadenylation factor [46]. These events would then allow Integrator to properly process the nascent snRNA transcript [46].

Until recently Integrator was thought to be an snRNA-specific processing complex, but it has now been shown to also regulate expression of protein-coding genes [47–49]. Expression of the immediate early genes (IEGs) activated by epidermal growth factor (EGF) is regulated by pausing prior to elongation and this pause is overcome by EGF induction. EGF induction regulates the recruitment of Integrator, which in turn recruits the SEC [47]. Integrator also plays a role in release of the NELF-mediated pause during transcription of

human genes and the HIV genome [48,49]. Interestingly, the human genes targeted by Integrator possess 3′ box-like sequences close to the end of the genes, and show a decrease in transcript levels upon Int11 knockdown, suggesting that Int11 is also involved in the production of mRNA [48]. Integrator therefore appears to play major roles in expression of protein-coding genes.

8. The 3′ box and termination of transcription of snRNA genes

The 3′ box is a signal for processing nascent snRNA transcripts and pol II continues transcribing for a few hundred base pairs past the 3′ box before transcription is terminated [39]. As already mentioned, Integrator binds DSIF, NELF and pol II, recognizes the 3′ box signal on the RNA and processes the nascent transcript. Pre-snRNA 3′ end formation and pol II transcription termination on snRNA genes is intrinsically linked, as demonstrated by a recent study in which knockdown of the catalytic subunits Int9 and Int11 caused termination defects [32]. NELF is also involved in termination. It accumulates downstream of the 3′ box of the U2 snRNA gene and its knockdown also causes a termination defect [45]. A CTCF binding site is located close to the region of NELF association, and delimits a nucleosome-depleted region spanning the transcription unit [45]. CTCF is best known as a transcription factor for protein-coding genes through its action as an insulator and organizer of chromatin structure [85]. CTCF has been shown to position nucleosomes and to impede pol II elongation [50,86]. By acting in this way, CTCF could control the chromatin structure and elicit termination of transcription. U2 snRNA genes may therefore have a 'late' elongation checkpoint where CTCF-positioned nucleosomes slow down pol II, allowing NELF to terminate transcription [45].

In agreement with these findings, a more recent study implicates CTCF in DSIF, NELF and P-TEFb recruitment to the U2 snRNA gene [51]. The findings suggest a model linking transcription termination to recognition of the 3′ box. In this model, NELF is required for termination of transcription, since its knockdown results in termination defects without affecting CTCF recruitment or RNA 3′ end formation. Additionally, P-TEFb phosphorylates the pol II CTD on Ser2, leading to the recruitment of Integrator subunits 9/11, which can elicit RNA processing driven by the 3′ box [51]. This model would explain how transcription termination and RNA processing are tightly linked during expression of snRNA genes (figure 4).

Intriguingly, mRNA 3′ end processing factors Pcf11, Ssu72 and Cstf64 also associate with snRNAs genes [32]. However, Pcf11 and Ssu72 were shown to affect termination of transcription of snRNA genes rather than 3′ end processing.

9. Coupling of initiation and RNA 3′ end formation

The 3′ box is only recognized if transcription is initiated from a pol II-dependent snRNA gene promoter [15,20]. If the PSE is substituted by a protein-coding gene promoter, the transcripts become longer as the 3′ box signal is ignored and pol II transcribes until a functional polyadenylation site is
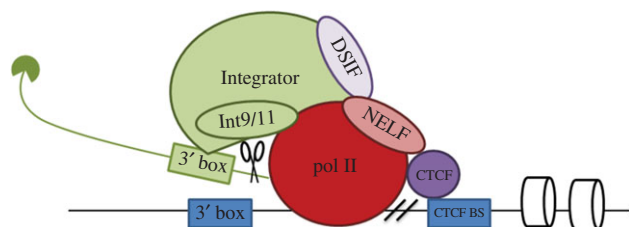
Figure 4. Model for transcription termination of the U2 snRNA gene. The snRNA transcript is represented in green with the cap in the 5′ end and nucleosomes are represented by barrels. Pol II continues to transcribe after the 3′ box and Integrator processes the nascent RNA after 3′ box recognition. CTCF recognizes the CTCF binding site downstream of the 3′ box and controls nucleosome occupancy. Negative elongation factor (NELF) is recruited by DRB-sensitivity-inducing factor (DSIF) and CTCF at the end of the transcription unit and causes transcription termination [51].

reached [87]. The mechanism underlying coupling of transcription initiation and RNA 3′ end formation is still not understood but may involve the snRNA gene-specific promoter factor PTF or the snRNA genes-specific components of the LEC. Alternatively, recognition of the 3′ box may only occur in the absence of SEC recruitment.

10. snRNA pseudogenes and variant snRNA genes

With the exception of teleosts and birds, all metazoans studied so far have pseudogenes derived from snRNA genes, with mammals possessing a large number [19]. Some sequences in the human genome that were previously thought to be U1 snRNA pseudogenes have been shown to be true genes that have recognizable DSE, PSE and 3′ box sequences and are actively transcribed [17]. These genes have been termed variant (v)U1 genes [17,88]. The DSE, PSE and 3′ box sequences of the vU1 snRNA genes are, however, sufficiently different from the U1 snRNA gene sequences to suggest that they are differentially regulated. Interestingly, the vU1 snRNAs appear to be most highly expressed in embryonic stem cells [17]. U1b is a mouse U1 variant gene only expressed in early embryonic stages, although the U1 and U1b snRNAs differ by only seven base changes [89,90]. However, the promoter of U1b differs significantly from that of U1 [90]. Although the function of U1b is still unclear, there are clear parallels between mouse U1b and the human vU1 snRNAs.

U5 snRNA variants in human and *Drosophila* have also been described [91,92]. The variants form different functional snRNP complexes from the canonical U5 snRNP and are thought to play roles in splicing [91]. Interestingly, the U5 variants are differentially expressed during development and may promote tissue-specific splicing [92].

11. Conclusion

The role of general transcription factors, CTD phosphorylation and the Integrator complex in expression of snRNA genes has been well characterized in the last two decades. The recent elucidation of the roles of DSIF, CTCF and NELF in expression of snRNA genes suggests how 3′ end processing of snRNA transcripts is linked to transcription termination. Also, it has been shown that mRNA 3′ end processing factors such as Pcf11 and Ssu72 are involved in termination of transcription of snRNA genes. We can now appreciate that there are many transcription factors shared by protein-coding genes and snRNA genes, although some play distinct roles in expression of the two gene types (figure 2).

Despite all this new information, there are still some important aspects of the expression of pol II-transcribed snRNA genes that remain to be elucidated. For example, how is initiation of transcription of these genes so tightly coupled to 3′ end formation of the transcripts? Why would U2 snRNA gene expression be regulated by different mechanisms than the U1, U4 and U5 snRNA genes? Why do snRNA genes require a specific elongation complex that is recruited by a non-coding RNA? Does the Mediator complex on the snRNA genes differ from that on protein-coding genes? And why does the herpesvirus transactivator protein VP16 only activate transcription of protein-coding genes and not snRNA genes [93] when activation occurs through interaction with TAF9 [94] and TFIIA [95]? However, VP16 also interacts with Med25 [96] and it is currently unclear whether this is recruited to snRNA genes.

Hopefully, future research will provide the answers necessary for a full understanding of the regulation of expression of this important class of pol II-transcribed genes.

References

1. Grummt I. 2003 Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev.* **17**, 1691–1702. (doi:10.1101/gad.1098503R)

2. Jawdekar GW, Henry RW. 2008 Transcriptional regulation of human small nuclear RNA genes. *Biochim. Biophys. Acta* **1779**, 295–305. (doi:10.1016/j.bbagrm.2008.04.001)

3. White RJ. 2011 Transcription by RNA polymerase III: more complex than we thought. *Nat. Rev. Genet.* **12**, 459–463. (doi:10.1038/nrg3001)

4. Russell J, Zomerdijk JC. 2005 RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem. Sci.* **30**, 87–96. (doi:10.1016/j.tibs.2004.12.008)

5. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH *et al.* 2004 MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060.

6. Matera AG, Terns RM, Terns MP. 2007 Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.* **8**, 209–220. (doi:10.1038/nrm2124)

7. Tarn WY, Steitz JA. 1997 Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* **22**, 132–137. (doi:10.1016/S0968-0004(97)01018-9)

8. Marzluff WF, Wagner EJ, Duronio RJ. 2008 Metabolism and regulation of canonical histone

mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9**, 843–854. (doi:10.1038/nrg2438)

9. Culver GM. 2002 Sno-capped: 5′ ends of preribosomal RNAs are decorated with a U3 SnoRNP. *Chem. Biol.* **9**, 777–779. (doi:10.1016/S1074-5521(02)00171-0)

10. Egloff S, O'Reilly D, Murphy S. 2008 Expression of human snRNA genes from beginning to end. *Biochem. Soc. Trans.* **36**, 590–594. (doi:10.1042/BST0360590)

11. Hernandez N. 2001 Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J. Biol. Chem.* **276**, 26 733–26 736. (doi:10.1074/jbc.R100032200)

12. Murphy S, Yoon JB, Gerster T, Roeder RG. 1992 Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes. *Mol. Cell. Biol.* **12**, 3247–3261. (doi:10.1128/MCB.12.7.3247)

13. Janson L, Bark C, Pettersson U. 1987 Identification of proteins interacting with the enhancer of human U2 small nuclear RNA genes. *Nucleic Acids Res.* **15**, 4997–5016. (doi:10.1093/nar/15.13.4997)

14. Schaub M, Myslinski E, Schuster C, Krol A, Carbon P. 1997 Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. *EMBO J.* **16**, 173–181. (doi:10.1093/emboj/16.1.173)

15. de Vegvar HE, Lund E, Dahlberg JE. 1986 3′ end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell* **47**, 259–266. (doi:10.1016/0092-8674(86)90448-4)

16. Hung KH, Stumph WE. 2011 Regulation of snRNA gene expression by the *Drosophila melanogaster* small nuclear RNA activating protein complex (DmSNAPc). *Crit. Rev. Biochem. Mol. Biol.* **46**, 11–26. (doi:10.3109/10409238.2010.518136)

17. O'Reilly D, Dienstbier M, Cowley SA, Vazquez P, Drozdz M, Taylor S, James WS, Murphy S. 2013 Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. *Genome Res.* **23**, 281–291. (doi:10.1101/gr.142968.112)

18. Lindgren V, Ares Jr M, Weiner AM, Francke U. 1985 Human genes for U2 small nuclear RNA map to a major adenovirus 12 modification site on chromosome 17. *Nature* **314**, 115–116. (doi:10.1038/314115a0)

19. Marz M, Kirsten T, Stadler PF. 2008 Evolution of spliceosomal snRNA genes in metazoan animals. *J. Mol. Evol.* **67**, 594–607. (doi:10.1007/s00239-008-9149-6)

20. Hernandez N, Weiner AM. 1986 Formation of the 3′ end of U1 snRNA requires compatible snRNA promoter elements. *Cell* **47**, 249–258. (doi:10.1016/0092-8674(86)90447-2)

21. Janson L, Weller P, Pettersson U. 1989 Nuclear factor I can functionally replace transcription factor Sp1 in a U2 small nuclear RNA gene enhancer. *J. Mol. Biol.* **205**, 387–396. (doi:10.1016/0022-2836(89)90349-5)

22. Sadowski CL, Henry RW, Lobo SM, Hernandez N. 1993 Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters

through the PSE. *Genes Dev.* **7**, 1535–1548. (doi:10.1101/gad.7.8.1535)

23. Waldschmidt R, Wanandi I, Seifart KH. 1991 Identification of transcription factors required for the expression of mammalian U6 genes in vitro. *EMBO J.* **10**, 2595–2603.

24. Henry RW, Sadowski CL, Kobayashi R, Hernandez N. 1995 A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerase II and III. *Nature* **374**, 653–656. (doi:10.1038/374653a0)

25. Murphy S. 1997 Differential *in vivo* activation of the class II and class III snRNA genes by the POU-specific domain of Oct-1. *Nucleic Acids Res.* **25**, 2068–2076. (doi:10.1093/nar/25.11.2068)

26. Christova R, Oelgeschlager T. 2002 Association of human TFIID-promoter complexes with silenced mitotic chromatin *in vivo*. *Nat. Cell Biol.* **4**, 79–82. (doi:10.1038/ncb733)

27. Zaborowska J, Taylor A, Roeder RG, Murphy S. 2012 A novel TBP-TAF complex on RNA polymerase II-transcribed snRNA genes. *Transcription* **3**, 92–104. (doi:10.4161/trns.19783)

28. Bieniossek C *et al.* 2013 The architecture of human general transcription factor TFIID core complex. *Nature* **493**, 699–702. (doi:10.1038/nature11791)

29. Takahashi H *et al.* 2015 MED26 regulates the transcription of snRNA genes through the recruitment of little elongation complex. *Nat. Commun.* **6**, 5941. (doi:10.1038/ncomms6941)

30. Kuhlman TC, Cho H, Reinberg D, Hernandez N. 1999 The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. *Mol. Cell. Biol.* **19**, 2130–2141. (doi:10.1128/MCB.19.3.2130)

31. Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, Bentley DL. 2009 TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol. Cell. Biol.* **29**, 5455–5464. (doi:10.1128/MCB.00637-09)

32. O'Reilly D, Kuznetsova OV, Laitem C, Zaborowska J, Dienstbier M, Murphy S. 2014 Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res.* **42**, 264–275. (doi:10.1093/nar/gkt892)

33. Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S. 2007 Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**, 1777–1779. (doi:10.1126/science.1145989)

34. Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ. 2009 TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **34**, 387–393. (doi:10.1016/j.molcel.2009.04.016)

35. Baillat D, Hakimi MA, Naar AM, Shilatifard A, Cooch N, Shiekhattar R. 2005 Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276. (doi:10.1016/j.cell.2005.08.019)

36. Medlin J, Scurry A, Taylor A, Zhang F, Peterlin BM, Murphy S. 2005 P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *EMBO J.* **24**, 4154–4165. (doi:10.1038/sj.emboj.7600876)

37. Wada T, Takagi T, Yamaguchi Y, Watanabe D, Handa H. 1998 Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO J.* **17**, 7395–7403. (doi:10.1093/emboj/17.24.7395)

38. Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H. 1999 NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51. (doi:10.1016/S0092-8674(00)80713-8)

39. Cuello P, Boyd DC, Dye MJ, Proudfoot NJ, Murphy S. 1999 Transcription of the human U2 snRNA genes continues beyond the 3′ box *in vivo*. *EMBO J.* **18**, 2867–2877. (doi:10.1093/emboj/18.10.2867)

40. Allen BL, Taatjes DJ. 2015 The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155–166. (doi:10.1038/nrm3951)

41. Smith E, Lin C, Shilatifard A. 2011 The super elongation complex (SEC) and MLL in development and disease. *Genes Dev.* **25**, 661–672. (doi:10.1101/gad.2015411)

42. Smith ER *et al.* 2011 The little elongation complex regulates small nuclear RNA transcription. *Mol. Cell* **44**, 954–965. (doi:10.1016/j.molcel.2011.12.008)

43. Hutten S, Chachami G, Winter U, Melchior F, Lamond AI. 2014 A role for the Cajal-body-associated SUMO isopeptidase USPL1 in snRNA transcription mediated by RNA polymerase II. *J. Cell Sci.* **127**, 1065–1078. (doi:10.1242/jcs.141788)

44. Chen J, Ezzeddine N, Waltenspiel B, Albrecht TR, Warren WD, Marzluff WF, Wagner EJ. 2012 An RNAi screen identifies additional members of the *Drosophila* Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3′-end formation. *RNA* **18**, 2148–2156. (doi:10.1261/rna.035725.112)

45. Egloff S, Al-Rawaf H, O'Reilly D, Murphy S. 2009 Chromatin structure is implicated in 'late' elongation checkpoints on the U2 snRNA and beta-actin genes. *Mol. Cell. Biol.* **29**, 4002–4013. (doi:10.1128/MCB.00189-09)

46. Yamamoto J, Hagiwara Y, Chiba K, Isobe T, Narita T, Handa H, Yamaguchi T. 2014 DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat. Commun.* **5**, 4263. (doi:10.1038/ncomms5263)

47. Gardini A, Baillat D, Cesaroni M, Hu D, Marinis JM, Wagner EJ, Lazar MA, Shilatifard A, Shiekhattar R. 2014 Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* **56**, 128–139. (doi:10.1016/j.molcel.2014.08.004)

48. Stadelmayer B *et al.* 2014 Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* **5**, 5531. (doi:10.1038/ncomms6531)

49. Skaar JR, Ferris AL, Wu X, Saraf A, Khanna KK, Florens L, Washburn MP, Hughes SH, Pagano M. 2015 The Integrator complex controls the

termination of transcription at diverse classes of gene targets. *Cell Res.* **25**, 288–305. (doi:10.1038/cr.2015.19)

50. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011 CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79. (doi:10.1038/nature10442)

51. Laitem C, Zaborowska J, Tellier M, Yamaguchi Y, Cao Q, Egloff S, Handa H, Murphy S. 2015 CTCF regulates NELF, DSIF and P-TEFb recruitment during transcription. *Transcription* **6**, 79–90. (doi:10.1080/21541264.2015.1095269)

52. Nguyen VT, Kiss T, Michels AA, Bensaude O. 2001 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414**, 322–325. (doi:10.1038/35104581)

53. Yang F, Hsu P, Lee SD, Yang W, Hoskinson D, Xu W, Moore C, Varani G. 2017 The C terminus of Pcf11 forms a novel zinc-finger structure that plays an essential role in mRNA 3′-end processing. *RNA* **23**, 98–107. (doi:10.1261/rna.058354.116)

54. Lee TI, Young RA. 2000 Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137. (doi:10.1146/annurev.genet.34.1.77)

55. Marshall NF, Peng J, Xie Z, Price DH. 1996 Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* **271**, 27 176–27 183. (doi:10.1074/jbc.271.43.27176)

56. Zhao X, Pendergrast PS, Hernandez N. 2001 A positioned nucleosome on the human U6 promoter allows recruitment of SNAPc by the Oct-1 POU domain. *Mol. Cell* **7**, 539–549.

57. Boyd DC, Greger IH, Murphy S. 2000 *In vivo* footprinting studies suggest a role for chromatin in transcription of the human 7SK gene. *Gene* **247**, 33–44. (doi:10.1016/S0378-1119(00)00134-7)

58. Boyd DC, Pombo A, Murphy S. 2003 Interaction of proteins with promoter elements of the human U2 snRNA genes in vivo. *Gene* **315**, 103–112. (doi:10.1016/S0378-1119(03)00717-0)

59. Stunkel W, Kober I, Seifart KH. 1997 A nucleosome positioned in the distal promoter region activates transcription of the human U6 gene. *Mol. Cell. Biol.* **17**, 4397–4405. (doi:10.1128/MCB.17.8.4397)

60. Yuan CC, Zhao X, Florens L, Swanson SK, Washburn MP, Hernandez N. 2007 CHD8 associates with human Staf and contributes to efficient U6 RNA polymerase III transcription. *Mol. Cell. Biol.* **27**, 8729–8738. (doi:10.1128/MCB.00846-07)

61. Zaborowska J, Egloff S, Murphy S. 2016 The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.* **23**, 771–777. (doi:10.1038/nsmb.3285)

62. Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. 2012 Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol. Cell* **45**, 111–122. (doi:10.1016/j.molcel.2011.11.006)

63. Albrecht TR, Wagner EJ. 2012 snRNA 3′ end formation requires heterodimeric association of integrator subunits. *Mol. Cell. Biol.* **32**, 1112–1123. (doi:10.1128/MCB.06511-11)

64. Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S. 2010 The Integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J. Biol. Chem.* **285**, 20 564–20 569. (doi:10.1074/jbc.M110.132530)

65. Bartkowiak B, Liu P, Phatnani HP, Fuda NJ, Cooper JJ, Price DH, Adelman K, Lis JT, Greenleaf AL. 2010 CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* **24**, 2303–2316. (doi:10.1101/gad.1968210)

66. Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, Cimermancic P, Ule J, Peterlin BM. 2011 The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.* **25**, 2158–2172. (doi:10.1101/gad.16962311)

67. Liang K, Gao X, Gilmore JM, Florens L, Washburn MP, Smith E, Shilatifard A. 2015 Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Mol. Cell. Biol.* **35**, 928–938. (doi:10.1128/MCB.01426-14)

68. Jonkers I, Lis JT. 2015 Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177. (doi:10.1038/nrm3953)

69. Ho CK, Shuman S. 1999 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. (doi:10.1016/S1097-2765(00)80468-2)

70. Mattaj IW. 1986 Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46**, 905–911. (doi:10.1016/0092-8674(86)90072-3)

71. Lariviere L, Plaschka C, Seizl M, Wenzeck L, Kurth F, Cramer P. 2012 Structure of the Mediator head module. *Nature* **492**, 448–451. (doi:10.1038/nature11670)

72. Soutourina J, Wydau S, Ambroise Y, Boschiero C, Werner M. 2011 Direct interaction of RNA polymerase II and Mediator required for transcription *in vivo*. *Science* **331**, 1451–1454. (doi:10.1126/science.1200188)

73. Seizl M, Lariviere L, Pfaffeneder T, Wenzeck L, Cramer P. 2011 Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization. *Nucleic Acids Res.* **39**, 6291–6304. (doi:10.1093/nar/gkr229)

74. Lariviere L, Geiger S, Hoeppner S, Rother S, Strasser K, Cramer P. 2006 Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat. Struct. Mol. Biol.* **13**, 895–901. (doi:10.1038/nsmb1143)

75. Hu D *et al.* 2013 The little elongation complex functions at initiation and elongation phases of snRNA gene transcription. *Mol. Cell* **51**, 493–505. (doi:10.1016/j.molcel.2013.07.003)

76. Egloff S, Vitali P, Tellier M, Raffel R, Murphy S, Kiss T. 2017 The 7SK snRNP associates with the little elongation complex to promote snRNA gene expression. *EMBO J.* **36**, 934–948. (doi:10.15252/embj.201695740)

77. Weiner AM. 2005 E Pluribus Unum: 3′ end formation of polyadenylated mRNAs, histone mRNAs, and U snRNAs. *Mol. Cell* **20**, 168–170. (doi:10.1016/j.molcel.2005.10.009)

78. Burglin TR, Affolter M. 2016 Homeodomain proteins: an update. *Chromosoma* **125**, 497–521. (doi:10.1007/s00412-015-0543-8)

79. Skaar JR, Richard DJ, Saraf A, Toschi A, Bolderson E, Florens L, Washburn MP, Khanna KK, Pagano M. 2009 INTS3 controls the hSSB1-mediated DNA damage response. *J. Cell Biol.* **187**, 25–32. (doi:10.1083/jcb.200907026)

80. Li Y *et al.* 2009 HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J. Biol. Chem.* **284**, 23 525–23 531. (doi:10.1074/jbc.C109.039586)

81. Zhang F, Ma T, Yu X. 2013 A core hSSB1-INTS complex participates in the DNA damage response. *J. Cell Sci.* **126**, 4850–4855. (doi:10.1242/jcs.132514)

82. Huang J, Gong Z, Ghosal G, Chen J. 2009 SOSS complexes participate in the maintenance of genomic stability. *Mol. Cell* **35**, 384–393. (doi:10.1016/j.molcel.2009.06.011)

83. Gu P, Deng W, Lei M, Chang S. 2013 Single strand DNA binding proteins 1 and 2 protect newly replicated telomeres. *Cell Res.* **23**, 705–719. (doi:10.1038/cr.2013.31)

84. Baillat D, Wagner EJ. 2015 Integrator: surprisingly diverse functions in gene expression. *Trends Biochem. Sci.* **40**, 257–264. (doi:10.1016/j.tibs.2015.03.005)

85. Ong CT, Corces VG. 2014 CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246. (doi:10.1038/nrg3663)

86. Fu Y, Sinha M, Peterson CL, Weng Z. 2008 The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4**, e1000138. (doi:10.1371/journal.pgen.1000138)

87. Neuman de Vegvar HE, Dahlberg JE. 1989 Initiation and termination of human U1 RNA transcription requires the concerted action of multiple flanking elements. *Nucleic Acids Res.* **17**, 9305–9318. (doi:10.1093/nar/17.22.9305)

88. Guiro J, O'Reilly D. 2015 Insights into the U1 small nuclear ribonucleoprotein complex superfamily. *Wiley Interdiscip. Rev. RNA* **6**, 79–92. (doi:10.1002/wrna.1257)

89. Marzluff WF, Brown DT, Lobo S, Wang SS. 1983 Isolation and characterization of two linked mouse U1b small nuclear RNA genes. *Nucleic Acids Res.* **11**, 6255–6270. (doi:10.1093/nar/11.18.6255)

90. Lund E, Kahan B, Dahlberg JE. 1985 Differential control of U1 small nuclear RNA expression during mouse development. *Science* **229**, 1271–1274. (doi:10.1126/science.2412294)

91. Sontheimer EJ, Steitz JA. 1992 Three novel functional variants of human U5 small nuclear RNA.

Mol. Cell. Biol. **12**, 734 – 746. (doi:10.1128/MCB.12.2.734)

92. Chen L, Lullo DJ, Ma E, Celniker SE, Rio DC, Doudna JA. 2005 Identification and analysis of U5 snRNA variants in *Drosophila*. *RNA* **11**, 1473 – 1477. (doi:10.1261/rna.2141505)

93. Das G, Hinkley CS, Herr W. 1995 Basal promoter elements as a selective determinant of transcriptional activator function. *Nature* **374**, 657 – 660. (doi:10.1038/374657a0)

94. Uesugi M, Nyanguile O, Lu H, Levine AJ, Verdine GL. 1997 Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science* **277**, 1310 – 1313. (doi:10.1126/science.277.5330.1310)

95. Kobayashi N, Horn PJ, Sullivan SM, Triezenberg SJ, Boyer TG, Berk AJ. 1998 DA-complex assembly activity required for VP16C transcriptional activation. *Mol. Cell. Biol.* **18**, 4023 – 4031. (doi:10.1128/MCB.18.7.4023)

96. Milbradt AG, Kulkarni M, Yi T, Takeuchi K, Sun ZY, Luna RE, Selenko P, Näär AM, Wagner G. 2011 Structure of the VP16 transactivator target in the Mediator. *Nat. Struct. Mol. Biol.* **18**, 410 – 415. (doi:10.1038/nsmb.1999)