Non-coding RNAs set the stage for RNA Polymerase II transcription

Cécilia Studniarek¹, Sylvain Egloff^{2*} and Shona Murphy^{1*}

¹ Sir William Dunn School of Pathology, University of Oxford, UK.

² Laboratoire de Biologie Moléculaire Eucaryote, CNRS, UMR5099, Centre de Biologie Intégrative, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 9, France.

*Corresponding author: Shona Murphy   e-mail: shona.murphy@path.ox.ac.uk
phone: (44)1865275500
fax: (44)1865275515

*Corresponding author: Sylvain Egloff   e-mail: sylvain.egloff@univ-tlse3.fr
phone: (33)561335991
fax: (33)561335886

Running title: Regulation of RNA polymerase II transcription by non-coding RNAs

Keywords: regulatory RNAs, gene expression, eRNA, B2 RNA, 7SK RNA, U1 snRNP

Abstract (120 words)

Effective synthesis of mammalian messenger (m)RNAs depends on many factors that together direct RNA polymerase II (pol II) through the different stages of the transcription cycle and ensure efficient co-transcriptional processing of mRNAs. In addition to the many proteins involved in transcription initiation, elongation and termination, several non-coding (nc)RNAs also function as global transcriptional regulators. Understanding the mode of action of these non-protein regulators has been an intense area of research in recent years. Here, we describe how these ncRNAs influence key regulatory steps of the transcription process, to affect large numbers of genes. Through direct association with pol II or by modulating the activity of transcription or RNA processing factors, these regulatory RNAs perform critical roles in gene expression.

Setting the stage (3947 words)

In mammals, **RNA polymerase II (pol II, see Glossary)** synthesizes all messenger RNAs and various non-coding (nc)RNAs. As in a play, transcription of protein-coding genes consists of several closely orchestrated 'acts' (Figure 1, Key Figure). Firstly, pol II is recruited to gene **promoters** to initiate transcription, which may require the function of an **enhancer**. Shortly after initiation, pol II must escape **promoter-proximal pausing** to continue transcription. Once the pause is overcome, the polymerase transcribes at a high rate until it reaches the 3' end of the gene. In the last act, recognition of a **polyadenylation signal (PAS)** triggers termination through dissociation of the polymerase from the DNA template. Each step of this transcription 'cycle' is regulated by numerous factors, and a growing number of ncRNAs are emerging as important global players in pol II transcription regulation. Recent reviews have covered the function of long non-coding (lnc)RNAs in modulating gene expression [1, 2], with XIST RNA serving as a paradigm for lncRNA-mediated recruitment of chromatin remodelling complexes that repress transcription [3]. The ncRNAs we focus on here are neither as long as XIST (17 kb), nor as short as the short non-coding microRNAs (~21 nts) [4] that mainly affect translation and/or RNA stability, but are rather medium sized (between ~50 bp and ~2 kb) RNAs that regulate transcription more globally by affecting the transcription machinery. Recent findings

provide important insights into how these ncRNAs influence the distinct ‘acts of the play’ (Key Table). These ‘actors’ influence transcription by facilitating **DNA looping** (enhancer (e)RNAs), modulating initiation in response to stress (SINEs RNAs), modulating promoter-proximal pausing (7SK snRNA) or preventing **premature transcription termination (PTT)** (U1 snRNA) (Figure 1).

eRNAs direct the actors

Recruitment of pol II minimally requires sequences close to the **transcription start site (TSS)** in core promoters. Additional proximal and distal regulatory sequences control the efficiency of pol II recruitment [5]. Enhancers are distal regulatory elements that play key roles in determining cell-type-specific transcriptional activity and the transcriptional response to stimuli [6, 7] (Figure 1). They comprise densely-clustered binding sites for **transcription factors (TFs)** and can regulate transcription independently of orientation and distance from target promoters. TFs associated with enhancers recruit **cofactors** to relay activation signals to target gene promoters, increasing the frequency of initiation events or **transcriptional bursts** [6]. Enhancers can also recruit pol II to produce short (50-2000 nt) non-coding and unstable RNAs referred to as enhancer RNAs (eRNAs) [6, 8]. Synthesis of eRNAs from enhancers is considered a good marker of enhancer activity, as it tends to correlate with the transcriptional activity of target genes [9, 10]. Enhancer transcription is the initial transcriptional event upon cell differentiation or activation, preceding induction of target gene expression [11-13]. Transcription of enhancers most likely results from the accumulation of TFs and the consequent recruitment of pol II. However, accumulating evidence over the last decade indicates that at least some eRNAs contribute to enhancer function [8, 14-17]. Whereas most eRNAs are unstable and rapidly subjected to **exosome**-dependent degradation, some are more stable and even polyadenylated [11, 18], which could underpin their function (Figure 2).

How enhancers act over long distances is not yet fully understood, partly due to the difficulty in mapping the complex interactions between enhancers and promoters [19]. Studies of the three-dimensional organization of the genome indicate that DNA looping between enhancers and promoters enables the integration of TF binding and cofactor recruitment from enhancers to promoters [20-22]. Some eRNAs functionally interact with factors that promote DNA-looping, including cohesin and **mediator complexes**

and the ubiquitously-expressed TF YY1 [12, 23-29] to stimulate interaction between enhancers and target promoters (Figure 2A).

eRNAs can also promote the recruitment of cofactors and stimulate their enzymatic activity (Figure 2B). For example, in human colorectal cancer cells, TNF- α signaling-induced eRNAs locally associate with the bromodomain-containing protein 4 (BRD4) to promote its binding to acetylated histones at enhancers, which in turn stimulates pol II recruitment to target proinflammatory-genes [30]. Similarly, a subset of eRNAs in mouse embryonic fibroblasts stimulate the histone-acetyltransferase activity of CREB-binding protein (CBP) by direct interaction, which in turn leads to increased histone acetylation of enhancer chromatin and increased transcriptional activation of target genes [31]. eRNAs synthesized from the human *ADAMDEC1* gene enhancer in response to LPS stimulation interact *in cis* with the CBP-related p300 histone-acetyltransferase, but whether this affects p300 recruitment or activity remains unclear [32]. In addition, eRNAs produced from the distal enhancer of the murine gonadotropin hormone α -subunit gene are required for an increase in activating histone marks at the enhancer and target promoter [33], indicating that eRNAs can play significant roles in remodelling the chromatin landscape. Stimulus-induced eRNAs can also enhance transcription elongation of nearby target genes by activating the positive elongation factor P-TEFb or acting as a decoy for the negative elongation factor, NELF in mouse and human cells [13, 34, 35] (Figure 2C).

As eRNA production at active enhancers is widespread, it is unclear how eRNAs can interact specifically with partners. Some factors, such as CBP, p300 and BRD4, might interact with eRNAs in a sequence-independent manner, simply because they accumulate at active enhancers [17, 28, 30, 31, 36, 37]. However, some eRNAs make sequence-specific interactions. For example, a CAAA tract in the MyoD-induced mouse seRNA-1 mediates its interaction with hnRNPL [38]. Likewise, the human *KLK3* eRNA carries an androgen-response element (ARE), for interaction with the androgen receptor (AR) and the mediator complex [12]. In addition, a specific motif present on a subset of AR-regulated human eRNAs allows binding of P-TEFb [35]. Additional features of eRNAs, such as secondary structure or recently identified post-transcriptional modifications, may also provide specificity [39].

Recent findings also indicate that the interaction of eRNAs with cofactors can rely on the formation of membrane-less compartments at enhancers through **liquid-liquid phase separation** [40, 41]. For example, in human cells, oestrogen-induced eRNAs

embedded in larger ribonucleoprotein particles (RNPs) are required both for formation of phase-separated condensates and transcriptional activation [40]. As there are now several examples of phase-separation of active enhancers and transcription factors/cofactors [41-43], formation of phase-separated eRNAs-dependent RNPs could be a general mechanism for enhancer/eRNA activity (Figure 2D).

Stressed out, the lead actor waits for a SINE

The next stage of the transcription cycle of protein-coding genes is initiation (Figure 1) and RNAs transcribed by **RNA polymerase III (pol III)** can influence the frequency of initiation by pol II. A large fraction of mammalian genomes comprises poorly-conserved repetitive DNA [44]. Alu and B2 are **short interspersed elements (SINEs)** that propagate in the host genome using the reverse transcriptase and nuclease encoded by long interspersed elements (LINEs) [45]. There are ~350,000 copies of B2 in the mouse genome, and >1 000 000 Alu SINEs in the human genome (1 every 4 Kb) [46]. Long considered 'junk' DNA, they can influence expression in *cis* [47] and increase transcriptome diversity when embedded in mRNAs and by mediating formation of **circular (Circ)RNAs** [48]. B2 and Alu RNAs can also function as chromatin remodeler elements [49, 50], transcriptional elongation barriers [51], translational regulators [52], RNA localization signals [53, 54], mRNA alternative splicing switches [55] and regulators of mRNA stability [56, 57].

These SINEs also contain an internal pol III promoter and encode small ncRNAs (B2 RNAs are ~180 nt [46] and Alu RNAs are ~300 nt [47]) that can play key roles in transcription by pol II. Due to an inefficient pol III promoter, Alu/B2 transcripts accumulate at very low levels and in a highly cell-type specific manner in mammalian cells. In human cells, transcribed Alu elements may, in some instances, act as cell-type-specific enhancers [58, 59]. These enhancer-like pol III transcripts can interact with pol II and enhance transcription of neighboring protein-coding genes, as shown in neurons [60]. Production of Alu/B2 RNAs is modulated during development and is highly induced by cellular stress and diseases such as viral infection and cancer (Figure 3) [46, 47, 61, 62]. For example, when human or mouse cells undergo heat-shock, pol II transcription globally shuts down while expression of some key responsive genes is strongly up-regulated. Increased production of B2/Alu RNAs after heat-shock correlates with their co-localization with pol II at promoters of transcriptionally-repressed genes [63, 64], where they suppress transcription initiation [63, 65] (Figure

3). Mechanistically, they directly associate with pol II and bind the cleft normally occupied by DNA [63, 66-68]. Although contacts between pol II and the promoter are lost [69], SINE/pol II complexes remain at the promoter, possibly tethered by protein-protein interactions, but are unable to properly engage the DNA [66]. In addition, TFIIH incorporated into SINE-containing pre-initiation complexes (PICs) fails to efficiently phosphorylate the pol II **carboxy-terminal domain (CTD)** [70]. Stress-induced repression of transcription initiation by SINEs is reminiscent of the function of 6S RNA in bacteria during late-stationary-phase [71], and suggests that regulation of pol II activity during stress response by direct association of short RNAs with the transcription machinery is an evolutionarily-conserved mechanism.

Recent studies indicate that B2 RNA also participates in transcriptional activation of a subset of heat-shock-responsive genes. In unstressed NIH/3T3 mouse fibroblasts, B2 RNA broadly occupies the 5' regions, and, more surprisingly, the intronic regions of these genes, where it restricts the rate of pol II elongation [64]. After heat-shock, the **EZH2 Polycomb protein** is targeted to B2-dependent genes, where it stimulates the intrinsic ribozyme activity of B2 RNA, independently of EZH2 histone methyltransferase activity [64, 72]. B2-mediated inhibition is then rapidly relieved through endonucleolytic self-cleavage of the RNA, which releases the brake to elongation [64] (Figure 3). Intriguingly, the RNA-dependent RNA polymerase (RdRp) activity of pol II is able to lengthen the 3' end of B2 RNA by 18 nt, which turns B2 RNA into a less potent inhibitor and favours its degradation [73]. Pol II-directed extension of B2 RNA on this specific set of genes may convert B2 into a preferred substrate for EZH2 and favour ribozyme activation (Figure 3). The prolonged presence of EZH2 at heat shock-upregulated genes may also explain why these genes are not subjected to B2-mediated repression by heat-shock [64]. Although the sequences are different, Alu transcripts are also EZH2-dependent self-cleaving RNAs that could function as transcriptional switches during stress [72]. Thus, despite being a potential threat to genome integrity, SINEs prove beneficial to cells by acting as key players in the stress response.

7SK RNA is a versatile actor

After initiation, pol II in higher eukaryotes usually has to negotiate an **early-elongation checkpoint (EEC)** before it can enter productive elongation (Figure 1). The 7SK small nuclear (sn)RNA, which is an abundant pol III transcript of 331 nt in vertebrates, can

act at this step to regulate gene expression. 7SK snRNA folds into a complex secondary structure, recognised by constitutive binding partners 7SK γ -methylphosphate capping enzyme MePCE and the La-related protein family member Larp7 (Figure 4), which is essential for RNA stability.

The core 7SK snRNP, together with a HEXIM1/2 dimer, regulates pol II transcription by dynamically sequestering Positive Transcription Elongation Factor b (P-TEFb), which comprises Cdk9 kinase with a cyclin partner (Figure 4A) [74] and controls the release of pol II from the EEC [75]. Negative-elongation factors, NELF and DRB sensitivity-inducing factor DSIF, rapidly join pol II after transcription initiation and block elongation. P-TEFb recruited to promoter-proximally paused pol II phosphorylates the pol II CTD at serine 2 (Ser2), NELF, DSIF, thereby allowing the formation of an elongation-competent pol II complex [76]. Phospho-NELF dissociates from pol II, phospho-DSIF becomes a positive elongation factor and phospho-Ser2 participates in the recruitment of additional elongation-stimulatory factors (Figure 4A). Promoter-proximal pausing is crucial for regulation of gene expression under normal conditions and in response to a number of development cues and stimuli. In metazoans, P-TEFb associates with a large number of TFs, including MYC, NF- κ B, and cofactors, including BRD4 and Mediator, allowing specific recruitment to a wide range of promoters to control pause release [74]. 7SK/HEXIM sequesters P-TEFb in a 7SK/P-TEFb snRNP and blocks the catalytic site of Cdk9 [77] to regulate the amount of active P-TEFb available.

The release of inactive P-TEFb occurs in response to various signals, including DNA damage, cardiac hypertrophy and HIV infection [74]. This induces a transient increase in the fraction of active P-TEFb free to interact with chromatin targeting factors to release paused pol II. Recent studies in mouse and human cells indicate that 7SK/P-TEFb snRNP anchored to a subset of promoters allows local release of P-TEFb for “on-site” transcriptional activation [78, 79]. To anchor the 7SK/P-TEFb snRNP to chromatin the 7SK snRNP can interact with the H4R3me2 histone mark at enhancers, which stimulate pause release from the target promoters in a BRD4-dependent manner [80]. MePCE can also directly interact with histone tails to target P-TEFb to chromatin in several breast cell lines, but whether it does so in the form of the 7SK/P-TEFb snRNP is unclear [81]. In addition, the Kruppel-associated box (KRAB)-associated protein 1 KAP1/TRIM28 tethers the 7SK/P-TEFb snRNP to a variety of promoters in

HEK293 cells, which helps to ensure efficient pause release on signal-induced genes upon stimulation [82].

7SK/P-TEFb snRNP dissociation occurs through several pathways in response to various cellular stimuli. HEXIM1 and Cdk9 phosphorylation, CycT1 acetylation and MePCE cleavage can all induce P-TEFb release from the 7SK snRNP [74, 83, 84]). 7SK snRNA itself undergoes base-modification, but whether this influences snRNP composition remains unclear [85, 86]. Mammalian P-TEFb can also be directly extracted from the 7SK snRNP by viral (Tat) and cellular proteins (BRD4, Rbm7) that interact with components of the 7SK snRNP [87-89]. P-TEFb release can occur on either the nucleoplasmic or chromatin-associated P-TEFb reservoirs. For example, at the promoters of human ribosomal protein genes, the DEAD-box RNA helicase DDX21 interacts with 7SK/P-TEFb snRNP and in cooperation with the WD40 domain-containing protein, WRD43, promotes P-TEFb release [90, 91]. Additionally, in mouse embryonic fibroblasts, the serine/arginine rich splicing factor SRSF2 binds to many promoter-associated nascent RNAs and stimulates transcription elongation by promoting the transfer of P-TEFb from nearby promoter-bound 7SK snRNP to paused pol II [92] (Figure 4A).

Importantly, 7SK snRNP is essential for various transcriptional reprogramming events [74]. Notably, the inactive reservoir of P-TEFb is essential for correct transcriptional induction of many immediate-early genes implicated in DNA-damage and inflammatory responses in metazoans [87, 93]. In addition, 7SK snRNA expression increases in neuronal cells during mouse development and is required for embryonic stem cell differentiation into neurons [94, 95]. It is therefore not surprising that abnormal 7SK snRNA levels are associated with neural disorders [96, 97]. In the human megakaryocyte lineage, dissociation of the 7SK/P-TEFb snRNP is blocked at neonatal stages and re-activated at later stages to drive megakaryocyte morphogenesis [98, 99]. Conversely, loss of 7SK-dependent P-TEFb control leads to hyperproliferative pathological conditions such as cardiac hypertrophy and cancers [100, 101]. It, however, remains to be fully elucidated if the 7SK/P-TEFb snRNP plays a role in transcription regulation under normal growth conditions beyond serving as a P-TEFb reservoir.

Several P-TEFb-independent functions for the 7SK snRNP in pol II transcription regulation have recently been reported [78] (Figure 4B). For example, in mice, 7SK snRNA binds to multiple enhancers together with the ATP-dependent nucleosome-

remodelling BAF complex to restrict enhancer transcription [102]. 7SK snRNA also associates with the human Little Elongation Complex (LEC) to stimulate transcription of pol II-dependent spliceosomal snRNA and small nucleolar (sno) RNA genes [103]. Finally, in human cells, 7SK snRNA regulates replication-dependent histone mRNA 3'end formation through interaction with the Ars2 protein [104].

U1 snRNA: the director's cut

Even after pol II has entered productive elongation, it can be derailed before it gets to the final polyadenylation site where the mature mRNA is completed. (Figure 1). U1 snRNP is primarily a critical component of the pre-mRNAs co-transcriptional **splicing machinery**. Recognition of the **5' splice site (5'SS)**, through RNA/RNA base-pairing between the 5'-end region of U1 snRNA and 5'SSs in introns, is the first step of spliceosome assembly [105]. However, U1 snRNA accumulates in eukaryotic cells at a much higher level than other spliceosomal RNAs, suggesting additional functions for this snRNA. Indeed, U1 snRNA, but not U2 snRNA, directly interacts with pol II in a splicing-independent manner [106]. *In vitro*, U1 snRNA can promote transcription re-initiation through association with the **general transcription factor (GTF)** TFIID [107, 108]. In addition, U1 snRNA can form a complex of unknown function with the TAF15 subunit of the GTF TFIID [109, 110]. More recently, binding of U1 snRNA to 5'SS motifs on lncRNAs was shown to favour their nuclear retention in both mouse and human cell lines [111]. During the last decade, the major role of metazoan U1 snRNA as a suppressor of premature transcription termination has also been uncovered.

Transcription termination usually occurs at the 3' end of genes, after pol II transcribes a PAS (Figure 1). Recognition of the PAS by the **cleavage and polyadenylation (CPA)** machinery triggers the final steps of co-transcriptional RNA processing and the release of pol II from the template DNA. However, there are usually multiple potential PASs within transcription units. At the 3' end, regulated recognition of different PASs gives rise to **alternative polyadenylation** that makes a major contribution to eukaryotic transcriptome diversity [112, 113]. However, potential PASs within the gene do not usually give rise to functional transcripts and must be ignored to make a full-length transcript. Inhibition of these gene-internal PASs is termed telescripting and requires the U1 snRNP. U1 snRNP bound to 5'SSs and related motifs throughout nascent RNAs prevents the activation of cryptic PASs even several hundred nucleotides away [114-116] (Figure 5). Consequently, loss of U1 snRNA induces

291 global shortening of mRNAs, with most prematurely-terminated in the first intron.
 292 Notably, mammalian U1 snRNP helps to suppress proximal PASs enriched near the
 293 first stable nucleosome downstream of **CpG island** promoters, where pol II frequently
 294 stalls and can prematurely terminate [117]. At **divergent/bi-directional promoters**,
 295 U1 snRNP recognition motifs are largely depleted in the antisense direction [118, 119].
 296 This results in increased proximal PAS activation in the antisense direction, leading to
 297 the production of short and unstable transcripts targeted by the exosome [117-119].
 298 Thus, U1 snRNP helps to control both mRNA length and promoter directionality. While
 299 inhibiting splicing also favours PAS usage located in retained introns, telescripting
 300 appears specific to U1 and distinct from splicing, as inhibition of U2 snRNP does not
 301 increase gene-internal PAS usage [115, 120, 121].
 302 The exact mechanism of telescripting remains unclear, but likely relies on interplay
 303 between the U1 snRNP and the CPA machinery. Notably, U1 snRNP can suppress
 304 PAS usage through direct inhibition of Poly(A) polymerase, PAP α [105]. Several
 305 factors of the CPA machinery, including CSTF2, which is associated with proximal PAS
 306 usage, are part of the U1 snRNA interactome but not the U2 snRNA interactome [122,
 307 123]. U1 snRNP forms a complex with CPA factors (U1-CPAF) at cryptic PASs in
 308 introns, particularly in genes with long first introns [124]. Interestingly, binding of U1
 309 snRNA to 5'SS motifs is not required for U1-CPAF complex formation at cryptic PASs.
 310 Instead, inhibition of U1 snRNA binding to 5'SS motifs with **U1-antisense**
 311 **morpholinos** (AMOs) induces remodelling of the composition of the U1-CPAF
 312 complex. For example, PABPN1 and CFI68m association is increased after U1
 313 inhibition by AMOs. Although the role of PABPN1 and CFI68m in transcription
 314 termination remains controversial, they can stimulate CPA [124]. This suggests that
 315 U1 snRNP can impede the recruitment of factors critical for CPA. Alternatively, U1
 316 snRNP bound to nearby 5'SS motifs may interfere with the function of CPAFs. Other
 317 factors potentially involved include the FET family member FUS and Sam68, which
 318 were recently shown to interact with U1 snRNP in human and mouse cells [122, 125,
 319 126]. FUS and U1 snRNP cooperatively bind similar sites upstream of cryptic PAS on
 320 nascent RNAs [116]. Removal of either of the two disrupts the binding of the other and
 321 elicits cryptic PAS activation [116]. Similarly, a functional interaction between Sam68
 322 and U1 snRNP, both binding to nearby sites on specific pre-mRNAs, is required to
 323 suppress use of adjacent cryptic PASs in mouse male germ cells [127, 128].

While U1 snRNP-PAS suppression was discovered after artificial loss or inhibition of U1, several lines of evidence suggest that modulating PAS suppression is biologically relevant *in vivo*. For example, UV irradiation of human colon cancer cells causes a decrease in the level of U1 snRNA and strong induction of PTT, which can be reversed by U1 snRNA overexpression [129]. In a different context, activation of rat neurons elicits global transcription upregulation and mRNAs shortening, and this can be recapitulated by inhibition of U1 snRNA [130]. Additionally, maintenance of mouse embryonic stem cell pluripotency requires enhanced recruitment of U1 snRNP to the 5' region of genes by cap-binding complex (CBC) and Srrt/Ars2 to prevent PTT [131]. Interestingly, U1 snRNP binding inhibition by U1-AMOs reduces the expression of long genes through induction of PTT, while short genes, including many genes important for survival and response to stimuli, are instead transcriptionally induced [132], suggesting that differential U1-PAS suppression helps to sustain expression of essential genes while shutting down expression of non-essential genes. In line with this notion, vertebrate genes have evolved to harbour larger introns, with potentially more PASs. However, genes that respond rapidly to environmental conditions are generally shorter than average and therefore less susceptible to PTT caused by inappropriate recognition of PASs. Thus, the expansion of intron size as well as the differential distribution of PASs provide metazoans with an additional layer of gene regulation through U1-mediated PAS suppression.

Concluding remarks

Genome-wide analyses of nascent transcription has revealed that much of the genome is transcribed into various species of ncRNAs, most of which are unstable and/or expressed at a low level. While it is unlikely that pervasive transcription of the genome produces many more regulatory ncRNAs than we already know about, an important challenge will be to distinguish those which have a function in regulation of gene expression from those that are solely transcriptional “noise”. It is now clear that many unstable and/or poorly-expressed ncRNAs, such as eRNAs play major and general roles in regulating the transcription cycle and that ncRNAs can exert global control on gene expression in many different ways, depending on their stability, nuclear localisation, targeting to chromatin, or incorporation into RNPs. In addition to the ncRNAs discussed above that can widely regulate fundamental mechanisms of pol II transcription, a number of non-coding RNAs can modulate transcription in a more

gene-specific fashion. For instance, Gas5 (Growth arrest-specific 5) RNA mimics the DNA target site of the glucocorticoid receptor (GR), thereby inhibiting GR recruitment to its target promoters by reducing its cellular availability [133]. SRA (steroid receptor RNA activator) helps to activate transcription by steroid receptors [134] by providing an RNA scaffold for the recruitment of regulatory proteins to the target promoters. Some CircRNAs can be recruited to the promoter of their own gene to activate transcription through RNA-RNA interaction with U1 snRNA [135], although the underlying molecular mechanism is elusive (Table 1). Thus, ncRNAs have been co-opted to perform in a wide range of transcriptional control mechanisms in the cell. Notably, it is emerging that the stable and abundant 7SK and U1 snRNAs have a greater regulatory potential than initially thought, indicating that there is still much to learn even about the functions of ncRNAs that have been studied for decades. Several important questions remain regarding the functions and mechanisms of the ncRNAs discussed here (see Outstanding Questions) and future work will undoubtedly lead to new exciting discoveries about ncRNAs and transcription.

In conclusion, the importance of ubiquitous or widely-produced ncRNAs in the global regulation of transcription by pol II is becoming increasingly clear and recent research into the roles of these often-underestimated ncRNAs has led to important new insights into the plethora of mechanisms that are used to regulate expression of large and complex genomes.

Acknowledgements

We thank Tamás Kiss for attentively reading the manuscript and apologize to colleagues whose work has not been cited due to space restrictions. CS and SM are funded by Wellcome Trust Investigator Award 210641/Z/18/Z and SE is funded by INSERM Plan Cancer.

Table 1: Function of regulatory ncRNAs

Regulatory RNAs	Length (nt)	Established Function(s)	Refs
eRNAs	50-2000	<i>Facilitate DNA looping</i>	<i>[12, 23-29]</i>

		<i>Promote cofactor recruitment/activation to stimulate initiation</i> <i>Chromatin remodelling</i> <i>Stimulate pause release</i>	<i>[30-32]</i> <i>[31, 33]</i> <i>[13, 34, 35]</i>
B2 RNA	~180	<i>Restrict expression of heat shock genes under normal conditions</i>	<i>[64, 72]</i>
Alu RNAs	~300	<i>Repress transcription of protein-coding genes after stress</i> <i>Increase transcriptome diversity when embedded into mRNA</i>	<i>[63, 64, 66-70]</i> <i>[48, 55]</i>
7SK RNA	331	<i>Regulates P-TEFb activity</i> <i>Activates transcription of snRNA genes</i> <i>Regulates replication-dependent histone mRNA 3'end formation</i> <i>Controls enhancer transcription</i>	<i>[74, 77-79]</i> <i>[103]</i> <i>[104]</i> <i>[102]</i>
U1 snRNA	164	<i>Recognizes 5' splice sites to promote splicing</i> <i>Promotes transcription re-initiation</i> <i>Favours nuclear retention of lncRNAs</i> <i>Controls mRNA length through suppression of premature transcription termination</i> <i>Controls promoter directionality</i>	<i>[105]</i> <i>[107, 108]</i> <i>[111]</i> <i>[115, 116, 120, 124, 128, 130, 132]</i> <i>[117]</i>
Gas5	656	<i>Inhibits the recruitment of the glucocorticoid receptor</i>	<i>[133]</i>
SRA	687	<i>Co-activator of steroid receptors</i>	<i>[134]</i>
CircRNAs	600-65000	<i>Promote transcription through interaction with U1 snRNA</i>	<i>[135]</i>

References

1. Gil, N. and Ulitsky, I. (2020) Regulation of gene expression by cis-acting long non-coding RNAs. Nat Rev Genet 21 (2), 102-117.

- 391 2. Ransohoff, J.D. et al. (2018) The functions and unique features of long intergenic non-coding RNA.
392 Nat Rev Mol Cell Biol 19 (3), 143-157.
- 393 3. Monfort, A. and Wutz, A. (2020) The B-side of Xist. F1000Res 9.
- 394 4. Bartel, D.P. (2018) Metazoan MicroRNAs. Cell 173 (1), 20-51.
- 395 5. Haberle, V. and Stark, A. (2018) Eukaryotic core promoters and the functional basis of transcription
396 initiation. Nat Rev Mol Cell Biol 19 (10), 621-637.
- 397 6. Catarino, R.R. and Stark, A. (2018) Assessing sufficiency and necessity of enhancer activities for gene
398 expression and the mechanisms of transcription activation. Genes Dev 32 (3-4), 202-223.
- 399 7. Long, H.K. et al. (2016) Ever-Changing Landscapes: Transcriptional Enhancers in Development and
400 Evolution. Cell 167 (5), 1170-1187.
- 401 8. Sartorelli, V. and Lauberth, S.M. (2020) Enhancer RNAs are an important regulatory layer of the
402 epigenome. Nat Struct Mol Biol 27 (6), 521-528.
- 403 9. Andersson, R. et al. (2014) An atlas of active enhancers across human cell types and tissues. Nature
404 507 (7493), 455-461.
- 405 10. Mikhaylichenko, O. et al. (2018) The degree of enhancer or promoter activity is reflected by the
406 levels and directionality of eRNA transcription. Genes Dev 32 (1), 42-57.
- 407 11. Arner, E. et al. (2015) Transcribed enhancers lead waves of coordinated transcription in
408 transitioning mammalian cells. Science 347 (6225), 1010-4.
- 409 12. Hsieh, C.L. et al. (2014) Enhancer RNAs participate in androgen receptor-driven looping that
410 selectively enhances gene activation. Proc Natl Acad Sci U S A 111 (20), 7319-24.
- 411 13. Schaukowitch, K. et al. (2014) Enhancer RNA facilitates NELF release from immediate early genes.
412 Mol Cell 56 (1), 29-42.
- 413 14. de Lara, J.C. et al. (2019) Enhancer RNAs: Insights Into Their Biological Role. Epigenet Insights 12,
414 2516865719846093.
- 415 15. Arnold, P.R. et al. (2019) Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene
416 Expression and Cell Fate. Front Cell Dev Biol 7, 377.
- 417 16. Lewis, M.W. et al. (2019) Transcriptional control by enhancers and enhancer RNAs. Transcription
418 10 (4-5), 171-186.
- 419 17. Rahnamoun, H. et al. (2020) The role of enhancer RNAs in epigenetic regulation of gene expression.
420 Transcription 11 (1), 19-25.
- 421 18. Andersson, R. et al. (2014) Nuclear stability and transcriptional directionality separate functionally
422 distinct RNA species. Nat Commun 5, 5336.
- 423 19. Hariprakash, J.M. and Ferrari, F. (2019) Computational Biology Solutions to Identify Enhancers-
424 target Gene Pairs. Comput Struct Biotechnol J 17, 821-831.
- 425 20. Rao, S.S. et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of
426 chromatin looping. Cell 159 (7), 1665-80.
- 427 21. McGreal-Estrada, R.S. et al. (2018) Promoter-enhancer looping and shadow enhancers of the
428 mouse α A-crystallin locus. Biol Open 7 (12).
- 429 22. Rubin, A.J. et al. (2017) Lineage-specific dynamic and pre-established enhancer-promoter contacts
430 cooperate in terminal differentiation. Nat Genet 49 (10), 1522-1528.
- 431 23. Li, W. et al. (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional
432 activation. Nature 498 (7455), 516-20.
- 433 24. Tsai, P.F. et al. (2018) A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and
434 Regulates Transcription In trans. Mol Cell 71 (1), 129-141.e8.
- 435 25. Kim, Y.J. et al. (2018) Global transcriptional activity dynamics reveal functional enhancer RNAs.
436 Genome Res 28 (12), 1799-1811.
- 437 26. Sigova, A.A. et al. (2015) Transcription factor trapping by RNA in gene regulatory elements. Science
438 350 (6263), 978-81.
- 439 27. Weintraub, A.S. et al. (2017) YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell 171
440 (7), 1573-1588.e28.
- 441 28. Jiao, W. et al. (2018) HPSE enhancer RNA promotes cancer progression through driving chromatin
442 looping and regulating hnRNPU/p300/EGR1/HPSE axis. Oncogene 37 (20), 2728-2745.

29. Lai, F. et al. (2013) Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494 (7438), 497-501.
30. Rahnamoun, H. et al. (2018) RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. *Nat Struct Mol Biol* 25 (8), 687-697.
31. Bose, D.A. et al. (2017) RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell* 168 (1-2), 135-149.e22.
32. Shi, L. et al. (2018) Enhancer RNA and NFkappaB-dependent P300 regulation of ADAMDEC1. *Mol Immunol* 103, 312-321.
33. Pnueli, L. et al. (2015) RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin alpha-subunit gene. *Proc Natl Acad Sci U S A* 112 (14), 4369-74.
34. Shii, L. et al. (2017) SERPINB2 is regulated by dynamic interactions with pause-release proteins and enhancer RNAs. *Mol Immunol* 88, 20-31.
35. Zhao, Y. et al. (2016) Activation of P-TEFb by Androgen Receptor-Regulated Enhancer RNAs in Castration-Resistant Prostate Cancer. *Cell Rep* 15 (3), 599-610.
36. Shi, L. et al. (2018) Enhancer RNA and NFkB-dependent P300 regulation of ADAMDEC1. *Mol Immunol* 103, 312-321.
37. Lee, J.E. et al. (2017) Brd4 binds to active enhancers to control cell identity gene induction in adipogenesis and myogenesis. *Nat Commun* 8 (1), 2217.
38. Zhao, Y. et al. (2019) MyoD induced enhancer RNA interacts with hnRNPL to activate target gene transcription during myogenic differentiation. *Nat Commun* 10 (1), 5787.
39. Aguilo, F. et al. (2016) Deposition of 5-Methylcytosine on Enhancer RNAs Enables the Coactivator Function of PGC-1 α . *Cell Rep* 14 (3), 479-492.
40. Nair, S.J. et al. (2019) Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat Struct Mol Biol* 26 (3), 193-203.
41. Sabari, B.R. et al. (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361 (6400).
42. Boija, A. et al. (2018) Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175 (7), 1842-1855.e16.
43. Hnisz, D. et al. (2017) A Phase Separation Model for Transcriptional Control. *Cell* 169 (1), 13-23.
44. de Koning, A.P. et al. (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7 (12), e1002384.
45. Jurka, J. et al. (2007) Repetitive sequences in complex genomes: structure and evolution. *Annu Rev Genomics Hum Genet* 8, 241-59.
46. Kramerov, D.A. and Vassetzky, N.S. (2011) SINEs. *Wiley Interdiscip Rev RNA* 2 (6), 772-86.
47. Chen, L.L. and Yang, L. (2017) ALUalternative Regulation for Gene Expression. *Trends Cell Biol* 27 (7), 480-490.
48. Zhang, X.O. et al. (2014) Complementary sequence-mediated exon circularization. *Cell* 159 (1), 134-147.
49. Ferrari, R. et al. (2020) TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. *Mol Cell* 77 (3), 475-487 e11.
50. Lunyak, V.V. et al. (2007) Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* 317 (5835), 248-51.
51. Tajaddod, M. et al. (2016) Transcriptome-wide effects of inverted SINEs on gene expression and their impact on RNA polymerase II activity. *Genome Biol* 17 (1), 220.
52. Schein, A. et al. (2016) Identification of antisense long noncoding RNAs that function as SINEUPs in human cells. *Sci Rep* 6, 33605.
53. Chen, L.L. et al. (2008) Alu element-mediated gene silencing. *EMBO J* 27 (12), 1694-705.
54. Lubelsky, Y. and Ulitsky, I. (2018) Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* 555 (7694), 107-111.
55. Sorek, R. et al. (2002) Alu-containing exons are alternatively spliced. *Genome Res* 12 (7), 1060-7.

56. Gong, C. and Maquat, L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 470 (7333), 284-8.

57. Gong, C. et al. (2013) mRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay. *Nat Struct Mol Biol* 20 (10), 1214-20.

58. Su, M. et al. (2014) Evolution of Alu elements toward enhancers. *Cell Rep* 7 (2), 376-385.

59. Zhang, X.O. et al. (2019) Genome-wide analysis of polymerase III-transcribed Alu elements suggests cell-type-specific enhancer function. *Genome Res* 29 (9), 1402-1414.

60. Policarpi, C. et al. (2017) Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. *Cell Rep* 21 (10), 2879-2894.

61. Kaczkowski, B. et al. (2016) Transcriptome Analysis of Recurrently Deregulated Genes across Multiple Cancers Identifies New Pan-Cancer Biomarkers. *Cancer Res* 76 (2), 216-26.

62. Moolhuijzen, P. et al. (2010) The transcript repeat element: the human Alu sequence as a component of gene networks influencing cancer. *Funct Integr Genomics* 10 (3), 307-19.

63. Mariner, P.D. et al. (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 29 (4), 499-509.

64. Zovoilis, A. et al. (2016) Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell* 167 (7), 1788-1802 e13.

65. Allen, T.A. et al. (2004) The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* 11 (9), 816-21.

66. Espinoza, C.A. et al. (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol* 11 (9), 822-9.

67. Kassube, S.A. et al. (2013) Structural insights into transcriptional repression by noncoding RNAs that bind to human Pol II. *J Mol Biol* 425 (19), 3639-48.

68. Ponicsan, S.L. et al. (2013) The non-coding B2 RNA binds to the DNA cleft and active-site region of RNA polymerase II. *J Mol Biol* 425 (19), 3625-38.

69. Yakovchuk, P. et al. (2009) B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. *Proc Natl Acad Sci U S A* 106 (14), 5569-74.

70. Yakovchuk, P. et al. (2011) B2 RNA represses TFIIF phosphorylation of RNA polymerase II. *Transcription* 2 (1), 45-9.

71. Wassarman, K.M. (2018) 6S RNA, a Global Regulator of Transcription. *Microbiol Spectr* 6 (3).

72. Hernandez, A.J. et al. (2020) B2 and ALU retrotransposons are self-cleaving ribozymes whose activity is enhanced by EZH2. *Proc Natl Acad Sci U S A* 117 (1), 415-425.

73. Wagner, S.D. et al. (2013) RNA polymerase II acts as an RNA-dependent RNA polymerase to extend and destabilize a non-coding RNA. *EMBO J* 32 (6), 781-90.

74. C Quaresma, A.J. et al. (2016) Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. *Nucleic Acids Res* 44 (16), 7527-39.

75. Core, L. and Adelman, K. (2019) Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* 33 (15-16), 960-982.

76. Zaborowska, J. et al. (2016) The pol II CTD: new twists in the tail. *Nat Struct Mol Biol* 23 (9), 771-7.

77. Kobbi, L. et al. (2016) An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb. *Proc Natl Acad Sci U S A* 113 (45), 12721-12726.

78. Egloff, S. et al. (2018) 7SK small nuclear RNA, a multifunctional transcriptional regulatory RNA with gene-specific features. *Transcription* 9 (2), 95-101.

79. McNamara, R.P. et al. (2016) Transcription elongation control by the 7SK snRNP complex: Releasing the pause. *Cell Cycle* 15 (16), 2115-2123.

80. Liu, W. et al. (2013) Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. *Cell* 155 (7), 1581-1595.

81. Shelton, S.B. et al. (2018) Crosstalk between the RNA Methylation and Histone-Binding Activities of MePCE Regulates P-TEFb Activation on Chromatin. *Cell Rep* 22 (6), 1374-1383.

82. McNamara, R.P. et al. (2016) KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. *Mol Cell* 61 (1), 39-53.

546 83. Lee, S. et al. (2020) JMJD6 cleaves MePCE to release positive transcription elongation factor b (P-
547 TEFb) in higher eukaryotes. *Elife* 9.

548 84. Sun, Y. et al. (2019) Activation of P-TEFb by cAMP-PKA signaling in autosomal dominant polycystic
549 kidney disease. *Sci Adv* 5 (6), eaaw3593.

550 85. Warda, A.S. et al. (2017) Human METTL16 is a N(6)-methyladenosine (m(6)A) methyltransferase
551 that targets pre-mRNAs and various non-coding RNAs. *EMBO Rep* 18 (11), 2004-2014.

552 86. Zhao, Y. et al. (2016) Pseudouridylation of 7SK snRNA promotes 7SK snRNP formation to suppress
553 HIV-1 transcription and escape from latency. *EMBO Rep* 17 (10), 1441-1451.

554 87. Bugai, A. et al. (2019) P-TEFb Activation by RBM7 Shapes a Pro-survival Transcriptional Response
555 to Genotoxic Stress. *Mol Cell* 74 (2), 254-267.e10.

556 88. Schröder, S. et al. (2012) Two-pronged binding with bromodomain-containing protein 4 liberates
557 positive transcription elongation factor b from inactive ribonucleoprotein complexes. *J Biol Chem* 287
558 (2), 1090-9.

559 89. Muniz, L. et al. (2010) Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator
560 Tat. *PLoS Pathog* 6 (10), e1001152.

561 90. Calo, E. et al. (2015) RNA helicase DDX21 coordinates transcription and ribosomal RNA processing.
562 *Nature* 518 (7538), 249-53.

563 91. Bi, X. et al. (2019) RNA Targets Ribogenesis Factor WDR43 to Chromatin for Transcription and
564 Pluripotency Control. *Mol Cell* 75 (1), 102-116.e9.

565 92. Ji, X. et al. (2013) SR proteins collaborate with 7SK and promoter-associated nascent RNA to release
566 paused polymerase. *Cell* 153 (4), 855-68.

567 93. McNamara, R.P. et al. (2013) Transcription factors mediate the enzymatic disassembly of promoter-
568 bound 7SK snRNP to locally recruit P-TEFb for transcription elongation. *Cell Rep* 5 (5), 1256-68.

569 94. Bazi, Z. et al. (2018) Rn7SK small nuclear RNA is involved in neuronal differentiation. *J Cell Biochem*
570 119 (4), 3174-3182.

571 95. Skreka, K. et al. (2012) Identification of differentially expressed non-coding RNAs in embryonic stem
572 cell neural differentiation. *Nucleic Acids Res* 40 (13), 6001-15.

573 96. Santoro, M. et al. (2016) Expression Profile of Long Non-Coding RNAs in Serum of Patients with
574 Multiple Sclerosis. *J Mol Neurosci* 59 (1), 18-23.

575 97. Alazami, A.M. et al. (2012) Loss of function mutation in LARP7, chaperone of 7SK ncRNA, causes a
576 syndrome of facial dysmorphism, intellectual disability, and primordial dwarfism. *Hum Mutat* 33 (10),
577 1429-34.

578 98. Elagib, K.E. et al. (2013) Calpain 2 activation of P-TEFb drives megakaryocyte morphogenesis and is
579 disrupted by leukemogenic GATA1 mutation. *Dev Cell* 27 (6), 607-20.

580 99. Elagib, K.E. et al. (2017) Neonatal expression of RNA-binding protein IGF2BP3 regulates the human
581 fetal-adult megakaryocyte transition. *J Clin Invest* 127 (6), 2365-2377.

582 100. Rice, A.P. (2009) Dysregulation of positive transcription elongation factor B and myocardial
583 hypertrophy. *Circ Res* 104 (12), 1327-9.

584 101. Ji, X. et al. (2014) LARP7 suppresses P-TEFb activity to inhibit breast cancer progression and
585 metastasis. *Elife* 3, e02907.

586 102. Flynn, R.A. et al. (2016) 7SK-BAF axis controls pervasive transcription at enhancers. *Nat Struct Mol*
587 *Biol* 23 (3), 231-8.

588 103. Egloff, S. et al. (2017) The 7SK snRNP associates with the little elongation complex to promote
589 snRNA gene expression. *EMBO J* 36 (7), 934-948.

590 104. Gruber, J.J. et al. (2012) Ars2 promotes proper replication-dependent histone mRNA 3' end
591 formation. *Mol Cell* 45 (1), 87-98.

592 105. Guirio, J. and O'Reilly, D. (2015) Insights into the U1 small nuclear ribonucleoprotein complex
593 superfamily. *Wiley Interdiscip Rev RNA* 6 (1), 79-92.

594 106. Spiluttini, B. et al. (2010) Splicing-independent recruitment of U1 snRNP to a transcription unit in
595 living cells. *J Cell Sci* 123 (Pt 12), 2085-93.

596 107. Kwek, K.Y. et al. (2002) U1 snRNA associates with TFIIF and regulates transcriptional initiation.
597 *Nat Struct Biol* 9 (11), 800-5.

598 108. O'Gorman, W. et al. (2005) Analysis of U1 small nuclear RNA interaction with cyclin H. *J Biol Chem*
599 280 (44), 36920-5.

600 109. Jobert, L. et al. (2009) Human U1 snRNA forms a new chromatin-associated snRNP with TAF15.
601 *EMBO Rep* 10 (5), 494-500.

602 110. Leichter, M. et al. (2011) A fraction of the transcription factor TAF15 participates in interactions
603 with a subset of the spliceosomal U1 snRNP complex. *Biochim Biophys Acta* 1814 (12), 1812-24.

604 111. Yin, Y. et al. (2020) U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* 580
605 (7801), 147-150.

606 112. Kamieniarz-Gdula, K. and Proudfoot, N.J. (2019) Transcriptional Control by Premature
607 Termination: A Forgotten Mechanism. *Trends Genet* 35 (8), 553-564.

608 113. Tian, B. and Manley, J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell*
609 *Biol* 18 (1), 18-30.

610 114. Engreitz, J.M. et al. (2014) RNA-RNA interactions enable specific targeting of noncoding RNAs to
611 nascent Pre-mRNAs and chromatin sites. *Cell* 159 (1), 188-199.

612 115. Kaida, D. et al. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and
613 polyadenylation. *Nature* 468 (7324), 664-8.

614 116. Masuda, A. et al. (2020) tRIP-seq reveals repression of premature polyadenylation by co-
615 transcriptional FUS-U1 snRNP assembly. *EMBO Rep* 21 (5), e49890.

616 117. Chiu, A.C. et al. (2018) Transcriptional Pause Sites Delineate Stable Nucleosome-Associated
617 Premature Polyadenylation Suppressed by U1 snRNP. *Mol Cell* 69 (4), 648-663.e7.

618 118. Almada, A.E. et al. (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation
619 signals. *Nature* 499 (7458), 360-3.

620 119. Ntini, E. et al. (2013) Polyadenylation site-induced decay of upstream transcripts enforces
621 promoter directionality. *Nat Struct Mol Biol* 20 (8), 923-8.

622 120. Li, W. et al. (2015) Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing
623 and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet*
624 11 (4), e1005166.

625 121. Lee, S.H. et al. (2018) Widespread intronic polyadenylation inactivates tumour suppressor genes
626 in leukaemia. *Nature* 561 (7721), 127-131.

627 122. Chi, B. et al. (2018) Interactome analyses revealed that the U1 snRNP machinery overlaps
628 extensively with the RNAP II machinery and contains multiple ALS/SMA-causative proteins. *Sci Rep* 8
629 (1), 8755.

630 123. Chu, C. et al. (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161 (2), 404-16.

631 124. So, B.R. et al. (2019) A Complex of U1 snRNP with Cleavage and Polyadenylation Factors Controls
632 Telescripting, Regulating mRNA Transcription in Human Cells. *Mol Cell* 76 (4), 590-599.e4.

633 125. Chi, B. et al. (2018) The neurodegenerative diseases ALS and SMA are linked at the molecular level
634 via the ASC-1 complex. *Nucleic Acids Res* 46 (22), 11939-11951.

635 126. Yu, Y. and Reed, R. (2015) FUS functions in coupling transcription to splicing by mediating an
636 interaction between RNAP II and U1 snRNP. *Proc Natl Acad Sci U S A* 112 (28), 8608-13.

637 127. Subramania, S. et al. (2019) SAM68 interaction with U1A modulates U1 snRNP recruitment and
638 regulates mTor pre-mRNA splicing. *Nucleic Acids Res* 47 (8), 4181-4197.

639 128. Naro, C. et al. (2019) Functional Interaction between U1snRNP and Sam68 Insures Proper 3' End
640 Pre-mRNA Processing during Germ Cell Differentiation. *Cell Rep* 26 (11), 2929-2941.e5.

641 129. Devany, E. et al. (2016) Intronic cleavage and polyadenylation regulates gene expression during
642 DNA damage response through U1 snRNA. *Cell Discov* 2, 16013.

643 130. Berg, M.G. et al. (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell*
644 150 (1), 53-64.

645 131. Kainov, Y.A. and Makeyev, E.V. (2020) A transcriptome-wide antitermination mechanism
646 sustaining identity of embryonic stem cells. *Nat Commun* 11 (1), 361.

647 132. Oh, J.M. et al. (2017) U1 snRNP telescripting regulates a size-function-stratified human genome.
648 *Nat Struct Mol Biol* 24 (11), 993-999.

133. Kino, T. et al. (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 3 (107), ra8.

134. Sheng, L. et al. (2018) New Insights Into the Long Non-coding RNA SRA: Physiological Functions and Mechanisms of Action. *Front Med (Lausanne)* 5, 244.

135. Li, Z. et al. (2015) Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* 22 (3), 256-64.

Glossary (max 450 words) 440

5' splice site: 5' end of an intron where first splicing step occurs.

Alternative polyadenylation: recognition of either one or another PAS.

Antisense morpholinos (AMOs): short chemically-modified oligonucleotides that base-pair with target RNAs.

Carboxyl-terminal domain (CTD): carboxyl-terminal domain of the pol II Rpb1 subunit comprising repeats of YSPTSPS reversibly post-translationally modified during transcription.

Circular RNAs: non-coding RNAs from aberrant backsplicing events that join a 5' splice site with an upstream 3' splice site.

Cleavage and polyadenylation (CPA): RNA cleavage directed by a PAS that generates a 3'OH to which a poly(A) tail is added.

Cofactor: protein without DNA-binding ability that modulates transcription.

CpG island: region of DNA enriched in CG dinucleotides frequently found at promoters. Methylation of cytosines in a CpG island can modulate transcription.

Divergent/bi-directional promoters: adjacent promoters that direct transcription in opposite directions.

DNA looping: folding of the DNA that brings distant regions, eg promoters and enhancers, closer.

Early elongation checkpoint: Transcriptional rate-limiting step just after initiation, when pol II pauses until P-TEFb activity allows productive elongation.

Enhancer: genomic region containing binding sites for transcription factors that stimulates transcription from target promoters over long and variable distances.

Exosome: protein complex that catalyses 3' to 5' degradation of RNA.

General Transcription Factor (GTF): component of the basic transcriptional apparatus that forms the preinitiation complex close to the TSS.

Liquid-liquid phase separation: cooperative interaction between multivalent and intrinsically disordered proteins resulting in condensates with liquid droplet-like properties.

684 **Mediator complex:** multiprotein complex that interacts with transcription factors and
685 pol II.

686 **Polyadenylation signal (PAS):** sequence on nascent RNAs recognized by the
687 cleavage and polyadenylation machinery.

688 **Polycomb proteins:** group of proteins implicated in epigenetic silencing of genes
689 through chromatin remodelling.

690 **Premature transcription termination (PTT):** transcription termination occurring
691 before the normal endpoint of transcription.

692 **Promoter:** the core promoter region, to which general transcription factors bind and
693 recruit pol II, and near upstream regulatory elements, containing binding sites for
694 transcription factors.

695 **Promoter-proximal pausing:** pausing of pol II just after transcription initiation,
696 caused by negative elongation factors.

697 **RNA polymerase II (pol II):** multi-subunit protein complex that synthesises RNA
698 from protein-coding and various ncRNA genes.

699 **RNA polymerase III (pol III):** multiprotein complex that transcribes the genes for
700 various ncRNAs, including tRNAs.

701 **Short interspersed elements (SINEs):** short repetitive DNA that can be transcribed
702 into ncRNAs.

703 **Splicing machinery:** proteins and snRNAs that together catalyse the removal of
704 introns from nascent RNAs.

705 **Telescripting:** suppression of premature transcription termination.

706 **Transcriptional bursting:** discontinuous production of transcripts in bursts or
707 pulses.

708 **Transcription factors (TF):** proteins with a DNA-binding domain and a trans-
709 activating domain that interact with pol II or cofactors to control transcription.

710 **Transcription start site (TSS):** 5'end of a gene where transcription starts.

711

712

713 **Figure legends**

714

Figure 1: Non-coding RNAs are important players in Pol II transcription regulation. eRNAs, B2 RNA, 7SK RNA and U1 snRNA acts at different steps of the transcription cycle and widely influence pol II transcription.

Figure 2: How eRNAs affect transcription by pol II. **A.** eRNAs interact with DNA looping factors to stabilize enhancer-promoter interactions and stimulate pol II activity. **B.** The enzymatic activity of enhancer-bound cofactors is enhanced by eRNA. **C.** eRNAs stimulate pol II pause release by activating the positive elongation factor P-TEFb and by acting as a decoy for the negative elongation factor NELF. **D.** At enhancers, eRNAs and cofactors can form ribonucleoprotein particles with phase separation properties.

Figure 3: B2 RNA functions as a transcriptional switch during the response to heat shock. Prior to heat shock, B2 RNA accumulates at low levels in cells and is targeted to heat shock genes to hinder transcription elongation. After heat shock, Pol III-directed transcription of B2 SINEs is induced, and B2 RNA is incorporated into Pol II pre-initiation complexes to block further initiation on most protein-encoding genes. Heat shock genes bypass transcription inhibition through EZH2-mediated degradation of B2 RNA, which releases the brakes on transcription elongation. Extension of B2 RNA by Pol II may facilitate EZH2-dependent degradation.

Figure 4: The function of 7SK snRNA in pause release and beyond. **A.** 7SK snRNA controls P-TEFb availability to release paused pol II. Stimulus-induced nucleoplasmic 7SK/P-TEFb snRNP dissociation releases P-TEFb from inhibition. Active P-TEFb targets paused Pol II complexes for phosphorylation of several substrates, including NELF, DSIF and CTD-Ser2, to promote entry into productive elongation. DDX21/WDR40 or SRSF2 proteins promote chromatin-anchored 7SK/P-TEFb snRNP dissociation to locally stimulate pause release. **B.** P-TEFb-independent functions of 7SK snRNA. 7SK snRNA forms complexes with the Little Elongation Complex (LEC) to promote snRNA gene expression, the BAF complex to restrict enhancer transcription and the Ars2 protein to prevent 3'-end processing of histone mRNAs.

746

747 **Figure 5: U1 snRNA suppress cryptic poly(A) site (PAS) usage and controls**
748 **promoter directionality.** In the sense direction, U1 snRNA binds to 5'SS and
749 functionally interacts with FUS or Sam68 proteins to inhibit the activity of the cleavage
750 and polyadenylation factors bound to a nearby PAS. U1-dependent PAS suppression
751 ensures production of full-length mRNAs. In the antisense direction, depletion of 5'SSs
752 favours the use of proximal PASs, giving rise to short and unstable transcripts
753 degraded by the exosome.

754