DNA processing in the context of non-coding transcription

Uthra Gowthaman\textsuperscript{1,*}, Desiré García-Pichardo\textsuperscript{1,*}, Yu Jin\textsuperscript{1}, Isabel Schwarz\textsuperscript{1}, Sebastian Marquardt\textsuperscript{1,2}

\* These authors contributed equally

\textbf{AFFILIATION}

\textsuperscript{1} Copenhagen Plant Science Centre

Department of Plant and Environmental Sciences

University of Copenhagen

Bülowsvej 21, 1870 Frederiksberg C

\textbf{CORRESPONDING AUTHOR}

\textsuperscript{2} sebastian.marquardt@plen.ku.dk (Sebastian Marquardt)

\textbf{KEYWORDS}

Long non-coding RNA (IncRNA); RNA Polymerase II (RNAPII) transcription; gene regulation; tandem transcriptional interference (tTI); antisense transcription; DNA processing
ABSTRACT

RNA polymerase II (RNAPII) frequently transcribes non-protein coding DNA sequences in eukaryotic genomes into long non-coding RNA (lncRNA). Here, we focus on the impact of the act of lncRNA transcription on nearby functional DNA units. Distinct molecular mechanisms linked to the position of lncRNA relative to the coding gene illustrate how non-coding transcription controls gene expression. We review the biological significance of the act of lncRNA transcription on DNA processing, highlighting common themes, such as mediating cellular responses to environmental changes. This review presents the background in chromatin signaling to appreciate examples in different organisms where we can interpret functions of non-coding DNA through the act of RNAPII transcription.
The act of non-coding transcription as a source of regulation of nearby DNA elements

RNA polymerase II (RNAPII) transcription of non-coding DNA converts large parts of eukaryotic genomes into non-coding RNAs (ncRNAs) [1,2]. Many annotated genes are therefore in close proximity to non-coding transcription, raising the question of how this affects gene expression [3,4]. RNAPII transcribed long non-coding RNAs (lncRNAs) are distinguished from shorter ncRNAs based on their size (> 200 nt) [5]. lncRNA may regulate other transcripts either in cis (near the site of production) or in trans (at a distant genomic location) [6]. Thus far, we distinguish gene regulation through cis-acting lncRNA by two main mechanisms: (i) the function of the resulting lncRNA molecule or (ii) the act of non-coding transcription [7]. In general, lncRNAs are less abundant and less conserved at the sequence level than messenger RNAs (mRNAs) [8]. Gene regulation by the act of lncRNA transcription could reconcile where non-conserved lncRNA in equivalent positions near genes may have similar effects on gene expression in distant species [7]. This review focuses on different molecular mechanisms by which the act of ncRNA transcription affects DNA processing (see Glossary) in cis. We divided ncRNAs according to their genomic location and orientation concerning the target protein-coding genes to encapsulate common mechanistic characteristics. Our review complements other excellent reviews on the functions exerted by lncRNA molecules [9–11] and metazoan enhancers [12] by focusing on how the act of non-coding transcription affects DNA biology.
Chromatin signaling during RNAPII transcription

In eukaryotes, histone proteins package DNA into chromatin with profound effects on transcription. N-terminal tails of histone proteins are dynamically modified with post-translational modifications (PTMs) during transcription [13,14]. Histone PTMs contribute to the definition of RNAPII transcription stages, where different pre-RNA processing events execute RNA synthesis [15]. Importantly, chromatin-modifying enzymes are recruited by dynamic phosphorylation of RNAPII and contribute to transcription stage specification [16,17].

Amongst histone PTMs, acetylation (ac) and methylation (me) marks are most thoroughly characterized. Histone acetyltransferases catalyze the acetylation of histones including histone 3 (H3) and histone 4 (H4) [18]. In Saccharomyces cerevisiae, RNAPII carboxy-terminal domain-dependent signaling triggers the methylation at H3 lysine (K) 4 and 36 by histone methyltransferases Set1 and Set2 [19–21]. Expressed genes display a characteristic modification profile with H3/H4ac around the transcription start sites (TSSs), H3K4 trimethylation (H3K4me3) around the first nucleosome, and H3K36me3 towards the 3'-end [22,23]. Additional histone PTMs reinforce characteristic distribution profiles across genes [24]. The pattern of histone PTMs along genes shares many similarities between species, for example human [25], Drosophila [26], yeasts [23,27] and Arabidopsis [17]. H3K36me3 characterizes RNAPII elongation in multiple organisms [28,29]. One important function of H3K36me3 in budding yeast is the suppression of intragenic transcription initiation [30–32]. Repression of intragenic initiation mediated by H3K36me3 is linked to the Rpd3S histone deacetylase complex (HDAC) and local chromatin remodeling [14,32].

Histone chaperones such as FAcilitates Chromatin Transcription complex (FACT) and
Spt6 promote RNAPII elongation through displacement and re-assembly of nucleosomes [33]. During active transcription, FACT and Spt6 function selectively on H2A/H2B and H3 to maintain the dynamic chromatin state [33]. Histone chaperones suppress transcription initiation within transcription units in several organisms, including yeast [34], humans [35] and plants [36]. Generally, lncRNA transcription shares much of the chromatin-based signaling with mRNA transcription [37]. Chromatin signaling associated with RNAPII transcription provides a valuable model for effects of the act of non-coding transcription on DNA biology. Scenarios where lncRNA transcription overlaps functional DNA elements serve as paradigms to illustrate the effects by this model.

**Tandem Transcriptional Interference**

Transcriptional interference (TI) defines the direct suppressive effect of RNAPII transcription in cis from one transcription unit overlaying a second transcription unit [38]. Transcription from a non-coding transcript can trigger TI to regulate a protein-coding gene located downstream in tandem [39]. In this section, we review common molecular mechanisms of TI into downstream genes by defined non-coding transcription events upstream in tandem (referred to as tTI). While we note tantalizing connections to the inhibition of downstream TSSs by gene isoforms initiating further upstream, these interactions between transcripts with coding potential fall outside the scope of this review [40–43].

*Interplay between chromatin-modifying complexes and transcription factors in tTI*

The local chromatin environment influences RNAPII transcription. Chromatin-based signaling mechanisms during gene transcription appear central for tTI across eukaryotes [17]. A prevalent tTI mechanism in yeast involves the deposition of transcription-coupled H3K36me3 by non-coding transcription to repress downstream gene initiation [44–46]. For
instance, tTI by two IncRNAs, *IRT1* and *IRT2*, controls cell-type based regulation of *IME1* expression determining sexual differentiation in budding yeast [47]. *IRT1* tTI triggers Set2-dependent H3K36me3 and Set3-dependent HDAC activity to repress *IME1* initiation in haploids. *IRT2*, a second IncRNA upstream of *IRT1*, in turn represses *IRT1* by tTI to promote *IME1* expression [45]. Likewise, a toggle-switch mechanism between two IncRNA transcripts (*ICR1* and *PWR1*) influenced by competitive binding of transcription factors, regulates the budding yeast *FLO11* gene affecting cell adhesion and flocculation [48,49]. The act of *ICR1* transcription represses *FLO11* by tTI and the expression of antagonistic IncRNA *PWR1* alleviates the repression. In fission yeast, IncRNA transcription upstream of the *tgp1*+ and *pho1*+ metabolite transporter genes deposits H3K36me3 that recruits HDAC to repress RNAPII initiation [44]. FACT participates in tTI by mediating reduced transcription factor binding at the downstream gene promoter (Figure 1A). Together, these examples of tTI by upstream IncRNA illustrate how an interplay of RNAPII elongation-linked chromatin signaling represses initiation of downstream gene expression in various biological contexts.

**Physical displacement and changes in chromatin structure**

In addition to the recruitment of histone-modifying proteins to repress the downstream gene, the act of RNAPII transcription may physically prevent access to DNA elements (Figure 1B). Experiments such as *in vivo* footprinting [50,51] and nucleosome-scanning [52] suggest transcription increases nucleosome density over nucleosome depleted regions (NDRs) that otherwise characterize eukaryotic promoter regions. The displacement of transcription factors by chromatin-compaction of NDRs may represent an alternative mechanism for tTI.

In *S. cerevisiae*, tTI from the upstream IncRNA *SRG1* regulates the *SER3* gene involved in serine biosynthesis [53]. In serine-rich media, FACT mediates repression by promoting nucleosome assembly at the *SER3* promoter and facilitates elongation of *SRG1*
independently of the Set2-mediated H3K36me3 [39]. The increased nucleosome density compacts chromatin and repels transcription factor binding to regulatory sequences in the SER3 promoter [39]. More generally, the act of transcription from an upstream non-coding unit may trigger the displacement of transcription factors from binding to the downstream gene promoter. The zinc-responsive transcription of the ZRR1 lncRNA displaces the transcriptional activator of ADH1 to repress gene expression [50]. ZRR1-based tTI of ADH1 links the expression of ADH1 to the availability of zinc in the environment. In fission yeast, tTI mediates environmentally adjusted phosphate homeostasis. tTI displaces the transcription factor Pho7 by upstream lncRNA transcription events at the pho1\(^+\), pho84\(^+\) and tgp1\(^+\) loci [51]. Moreover, compound transcripts resulting from invading upstream transcription may contain the mRNA open reading frame (ORF). In such cases, failure to produce functional proteins may include elements of translational repression by upstream ORFs [54]. To summarize, we find support for two prominent mechanisms of tTI causing gene repression. A mechanism relying on (i) chromatin-based signaling and (ii) the repelling effects mediated by chromatin compaction. The mechanisms are not mutually exclusive and may act together in some contexts. Possibly, the function of FACT interlinks the two mechanisms by preventing transcription factor binding and promotion of nucleosome assembly [55]. However, this review distinguishes them to acknowledge the strong evidence for both mechanisms. These mechanisms underpin environment-induced gene regulation, arguing for roles of tTI in linking environmental sensing to gene expression.

**Permissive effects of tandem lncRNA transcription on gene expression**

While tTI mediates repression, the act of lncRNA transcription near gene promoters may also activate gene expression. In fission yeast, lncRNA transcription upstream of the fbp1\(^+\) gene promotes expression by mediating an open configuration around the fbp1\(^+\) TSS [56]
In addition, efficient transcriptional termination of upstream lncRNA prevents tTI (see Box 1) and stimulates gene expression. The budding yeast lncRNA CUT60 strikingly illustrates the biological significance of this phenomenon. CUT60 termination promotes expression of the ATP16 gene, encoding a subunit of the mitochondrial ATP synthase complex. Inefficient termination of CUT60 represses ATP16 expression, resulting in mitochondrial genome loss [57]. CUT60 termination may also contribute to NDR maintenance to facilitate initiation. In summary, even though the act of RNAPII transcription is often associated with gene repression, several examples support roles in activation. Future characterization of examples illustrating roles in gene activation may help to discriminate molecular features connected to gene activation and repression.

**Gene regulation by the act of antisense transcription**

Transcription of lncRNAs from the antisense strand of protein-coding genes characterizes antisense lncRNA transcription [58,59]. A correlation between genes changing expression in response to new environmental conditions and the presence of non-coding antisense transcription suggests roles of antisense lncRNA in gene regulation [60,61]. Many case studies support a mutual inhibitory effect of the sense and antisense transcript pairs [60]. In addition to antisense lncRNA transcription, a range of synonymous terminologies exist, for example, natural antisense transcripts (NATs) [62], antisense lncRNAs (aslncRNAs) [63] or antisense RNAs (asRNAs) [64]. In this section, we review how the act of antisense non-coding transcription affects the expression of the corresponding sense gene.

**Transcriptional read-through to regulate gene expression**

Antisense lncRNA transcription can result from distinct promoters, read-through RNAPII transcription of neighboring genes facing each other, or from divergent lncRNA transcription
associated with eukaryotic gene promoters [59,65]. Antisense IncRNA transcription is abundant in gene-dense genomes like *S. cerevisiae* and associated with genes of varying expression levels [60]. For example, the *SUR7* gene is regulated by the *SUT719* antisense IncRNA transcript [60]. High expression levels of *SUR7* signals feedback repression by triggering increased *SUT719* expression, which in turn represses *SUR7*. Likewise, read-through transcription of the antisense IncRNA *RME2* into the 5’ untranslated region (UTR) of the *IME4* gene achieves cell-specific regulation of budding yeast meiosis [66]. *RME2* transcription opposes RNAPII elongation of *IME4* in haploid cells, thus repressing full-length *IME4* mRNA production. Nonsense-mediated RNA decay (NMD) targets antisense IncRNAs for degradation in the cytoplasm [67]. In budding yeast, stabilization of antisense IncRNAs in an NMD pathway mutant correlates with lower expression of genes, arguing for repressive effects of antisense IncRNA [64]. In differentiated mouse embryonic stem cells, genomic imprinting of the *Igf2r* gene is mediated by transcription of the antisense IncRNA *Airn* [68]. Repression of the paternal *Igf2r* allele by DNA methylation is alleviated by abolishing *Airn* transcription through the removal of the *Airn* promoter sequence, supporting a connection between antisense IncRNA expression and epigenetic regulation. Collectively, these examples point to mutual inhibition of the antisense IncRNA generated from read-through transcription and mRNA transcription (Figure 2A).

Chromatin-based repression by antisense IncRNA

Regulation of expression of sense gene expression by the corresponding antisense IncRNA transcription often relies on chromatin-based mechanisms (Figure 2B). The budding yeast *GAL10* locus represents a well-characterized example of this paradigm. Glucose-rich environments induce the transcription of the *GAL10* IncRNA. The act of *GAL10* IncRNA transcription deposits H3K36me3 and H3K4me2/3 to recruit Rpd3S and attenuates
expression of mRNA in the \textit{GAL1-10} locus [69,70]. Reduced antisense IncRNA transcription increases H3K9ac at the \textit{GAL1} gene, coinciding with increased \textit{GAL1} expression [71]. Similarly, transcription of an antisense IncRNA at the phosphate-repressed \textit{PHO84} gene induces Set1-dependent histone methylation that recruits HDACs to mediate repression of \textit{PHO84} [72]. Chromatin-based repression of \textit{PHO84} by antisense IncRNA may limit the threshold for gene activation to phosphate environments providing a sufficiently strong stimulus [72]. In \textit{Drosophila melanogaster}, the \textit{nascent} antisense IncRNA \textit{ANRIL} recruits Polycomb repressive complex (PRC) to silence the nearby \textit{Ink4b} gene by H3K27me3 [73]. Likewise, the mammalian \textit{ATX7} locus harbors the antisense IncRNA \textit{SCAANT1} whose transcription mediates increased deposition of H3K27me3, a PRC-linked repressive chromatin modification, to silence \textit{ATX7} [74]. In \textit{Arabidopsis}, cold-induced antisense IncRNA \textit{COOLAIR} transcription at the \textit{FLOWERING LOCUS C (FLC)} gene triggers \textit{FLC} repression involving PRC and H3K27me3 [75,76]. Alternative \textit{COOLAIR} co-transcriptional pre-RNA processing through varying poly-adenylation [77] and \textit{splicing} [78] alters the efficiency of antisense mediated \textit{FLC} repression. Single-cell analyses at the \textit{FLC} locus suggest mutually exclusive cellular expression states of sense and antisense IncRNA transcription, arguing for mutual inhibition [79]. In yeast and human, genes with high natural antisense IncRNA transcription are associated with low levels of H3K36me3 and H3K79me3 around the mRNA TSS [71]. These data indicate that transcription on both DNA strands of a gene coincides with impaired establishment of co-transcriptional chromatin signals. However, when antisense IncRNA transcription is induced, antisense transcription elevates levels of H3K36me3 in +1/-1 nucleosomes and triggers chromatin compaction to repress sense gene expression [80]. Here, antisense transcription appears to share some molecular
characteristics with tTI. In summary, the act of antisense IncRNA transcription affects chromatin signaling by diverse mechanisms to affect gene expression.

Emerging mechanisms of gene regulation by the act of antisense transcription

Antisense IncRNA transcription in budding yeast at the CDC28 locus represents an example of regulation by gene looping that correlates positively with gene expression [81]. Antisense IncRNA transcription recruits the Hog1 protein kinase that facilitates binding of the gene looping factor Ssu72. Gene looping mediates the transfer of Hog1 alongside transcription factors to the 5′-UTR of the CDC28 gene, promoting its expression (Figure 2C). The presence of antisense IncRNA transcription and mRNA poses a conundrum since biophysical considerations suggest that two RNAPII complexes transcribing towards each other should not be able to pass [82]. In Arabidopsis, the collision of RNAPII complexes corresponding to sense and antisense transcription of the same DNA locus represents a mechanism to regulate cold-tolerance [83]. Cold-induced read-through transcription of the antisense IncRNA SVALKA protrudes into the CBF1 gene, which triggers the collision of RNAPII complexes. At CBF1, RNAPII collision results in premature termination of CBF1 mRNA production in response to cold to uncouple fitness penalties associated with high-level CBF1 expression (Figure 2D). In mice, antisense IncRNA transcription at the Zeb2 gene (Zeb2-NAT) inhibits promoter-proximal splicing of Zeb2 pre-mRNA. Interestingly, the partially spliced transcript promotes Zeb2 protein expression through the inclusion of an internal ribosomal entry site required for the translation of mRNA in young mice [62]. Older mice display increased Zeb2-NAT and Zeb2 protein expression compared to younger animals, arguing for antisense IncRNA-mediated effects on ageing [84]. RNAPII also navigates non-coding antisense transcription by RNA polymerase III (RNAPIII). In mammals, the first intron of the RNAPII-transcribed Polr3e gene, encoding an RNAPIII
complex subunit, contains an interspersed repeat region. Simultaneous transcription of this
region by RNAPIII and transcription of sense Polr3e gene by RNAPII leads to a roadblock
and reduction in expression of Polr3e [85]. Future studies will clarify the scope of gene
expression regulation by mechanisms such as gene looping, polymerase collision and
splicing inhibition by the act of antisense IncRNA transcription. In summary, the mechanistic
diversity of gene regulation by antisense IncRNA transcription suggests context-dependent
interpretation of effects and calls for nuanced approaches to resolve common principles.

Emerging functions of non-coding transcription

Transcriptional control by intragenic ncRNAs

Although intergenic elements such as gene promoters are key for gene expression
regulation, intragenic ncRNA transcription may superimpose transcriptional control. For
instance, nitrogen deprivation causes rapid induction of the ASP3 gene in S. cerevisiae,
aided by an intragenic ncRNA transcribed in the same orientation as the mRNA [86]. Even
when ASP3 mRNA is repressed, the ncRNA transcription persists and mediates high
residual levels of H3K4me3 at the ASP3 promoter to facilitate rapid transcriptional
adaptation to changing nitrogen availability (Figure 3A). In mammals, transcription of
intragenic enhancer RNA may affect host gene expression [87,88]. We refer readers to
excellent recent reviews on enhancer function [12,89]. In brief, enhancer RNA transcription
can attenuate host gene expression by promoting premature termination of mRNA
transcription [87] (Figure 3B). Moreover, transcription of an intragenic antisense enhancer
can facilitate the expression of a shorter isoform of the host gene [88]. In summary, the act
of transcription of intragenic ncRNAs influences host gene expression. In this regard,
functions of mammalian intragenic enhancers can be interpreted through regulatory
mechanisms by the act of RNAPII transcription.
Defective termination as a gene expression regulator

An alternative transcriptional termination pathway in human cells, specific to most lncRNAs that contain microRNAs (miRNAs), prevents read-through into downstream genes [90]. The microprocessor complex responsible for recognizing and generating pre-miRNA is key in this termination mechanism. Generally, non-coding transcription of RNAPII beyond the polyadenylation site is part of the canonical RNAPII transcription termination mechanism and thus represents an integral part of gene expression [91]. Diverse environmental cellular stresses in mammals and plants, besides virus infections (e.g. herpes simplex (HSV-1) or influenza), interfere with RNAPII termination, leading to defective termination and widespread generation of read-through non-coding transcription into downstream protein-coding genes [92–96] (Figure 3C). The molecular mechanisms underlying read-through transcription in HSV-1 infection point to a blockage of mammalian RNAPII termination by preventing the mRNA 3'-cleavage [95]. Malignant reprogramming of cancer cells associates with widespread read-through transcription [97], further supporting the correlation between excessive non-coding read-through transcription and cellular stresses. Moreover, acceleration of RNAPII transcription in Arabidopsis revealed extensive transcriptional read-through linked to an autoimmune phenotype [98]. In conclusion, these data support links between non-coding read-through transcription and cellular fitness. It remains to be investigated whether widespread non-coding transcription events reflect cellular adaptations to new environments, or whether they indicate a breakdown of cellular functions.

Non-coding transcription and DNA processing

The act of lncRNA transcription also impacts DNA replication and genome maintenance. Autonomously replicating sequences (ARSs) instruct DNA replication in budding yeast. The
stable association of the origin recognition complex (ORC) and the pre-replication complex at ARSs promote RNAPII pausing by a roadblock mechanism and mediate transcriptional termination at the edges of replication origins [99,100] (Figure 4A). ARS invasion by lncRNA transcription correlates with low DNA replication efficiency and late origin firing, affecting the temporal program of replication [99,100]. Mechanistically, RNAPII transcription through ARSs affects replication in part by chromatin-mediated effects. Akin to tTI, RNAPII elongation mediates a Set2-dependent increase of H3K36me3, a decrease of histone acetylation and an increase in nucleosome occupancy around the ARSs [100]. This chromatin landscape correlates with lower efficiency of ORC binding and inefficient ARS firing [100] (Figure 4B). Future studies should investigate if tTI-like effects on DNA replication, as detected in S. cerevisiae, represent a phenomenon preserved through evolution. In summary, the act of lncRNA transcription over ARSs inhibits DNA replication and thus cell cycle progression.

In human cells, the act of lncRNA transcription is essential for the repair of double-strand breaks (DSBs), a harmful type of DNA lesion [101]. Transcription occurs in the first steps of the DNA damage response (DDR). DDR couples a series of mechanisms responsible for the recognition, signaling and repair of DNA damage. Detection of the lesion by the MRE11–RAD50–NBS1 (MRN) complex allows the formation of functional promoters, where RNAPII bidirectionally synthesizes damage-induced lncRNAs (dilncRNAs) [101,102]. dilncRNAs are processed into small ncRNAs that base-pair with nascent dilncRNA to trigger the formation of DDR foci, necessary for DSB repair [101]. Application of inhibitors of RNAPII transcription revealed the requirement of dilncRNA transcription for the formation of DDR foci, and that dilncRNA promote liquid-liquid phase-separation to concentrate repair factors around DDR.
foci [101,102] (Figure 4C). These examples support functional roles for the act of IncRNA transcription in DNA biology beyond the regulation of gene expression.

**Concluding remarks**

Research focus is shifting from the discovery of IncRNAs to clarifying their functional roles and molecular mechanisms of action. A general purpose of non-coding transcription remains elusive and most insight stems from case studies. Given the extent of spurious RNAPII transcription in genomes, it is attractive to reflect on functional implications of the act of IncRNA transcription regardless of the DNA sequence. While most evidence connects the act of IncRNA transcription to the regulation of gene expression, recent advances link this mechanism to DNA processing more broadly. An increasing repertoire of molecular techniques (see Box 2) promises to facilitate functional dissection of IncRNAs at their site of production to resolve the scope of these mechanisms. While superficially similar, the regulation of neighboring gene expression by the act of IncRNA transcription in cis may involve a range of mechanisms distinct in molecular detail. An important aspect of future studies will be the identification of molecular characteristics common to regulatory roles by the act of IncRNA transcription. Such information would inform predictions on which and how many non-coding DNA elements may function through the act of transcription (see Outstanding Questions), perhaps bypassing the need for detailed experimental dissection of IncRNA transcription events.

**ACKNOWLEDGEMENTS**

We thank the Marquardt lab members for their feedback and suggestions on the manuscript. We acknowledge funding from the Novo Nordisk Foundation NNF15OC0014202 and
Copenhagen Plant Science Centre Young Investigator Starting grant. The Marquardt Lab acknowledges support from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme StG2017-757411 (S.M.).

REFERENCES


10. Engreitz, J. M. et al. (2016) Long non-coding RNAs: Spatial amplifiers that control...


Biol. 23, 4207–4218

Li, B. et al. (2003) The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. J. Biol. Chem. 278, 8897–8903


Bell, O. et al. (2007) Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. EMBO J. 26, 4974–4984


Carrozza, M. J. et al. (2005) Histone H3 methylation by Set2 directs deacetylation of
coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581–592


Lin, D. et al. (2018) Intragenic transcriptional interference regulates the human immune ligand MICA. *EMBO J.* 37, e97138

Chia, M. et al. (2017) Transcription of a 5′ extended mRNA isoform directs dynamic chromatin changes and interference of a downstream promoter. *Elife* 6, e27420


Garg, A. et al. (2018) A long noncoding (Inc)RNA governs expression of the phosphate transporter Pho84 in fission yeast and has cascading effects on the flanking prtlncRNA and pho1 genes. J. Biol. Chem. 293, 4456–4467


Johnstone, T. G. et al. (2016) Upstream ORFs are prevalent translational repressors in vertebrates. EMBO J. 35, 706–723


Sopher, B. L. *et al.* (2011) CTCF Regulates Ataxin-7 Expression through Promotion of a Convergently Transcribed, Antisense Noncoding RNA. *Neuron* 70, 1071–1084


LaCava, J. et al. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* 121, 713–724

Wyers, F. et al. (2005) Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121, 725–737

Wery, M. et al. (2009) The nuclear poly(A) polymerase and Exosome cofactor Trf5 is recruited cotranscriptionally to nucleolar surveillance. *RNA* 15, 406–419

Ogami, K. et al. (2018) RNA surveillance by the nuclear RNA exosome: Mechanisms and significance. *Noncoding RNA* 4, 8


Howe, F. S. *et al.* (2017) CRISPRI is not strand-specific at all loci and redefines the transcriptional landscape. *Elife* 6, e29878


Pelechano, V. et al. (2014) Genome-wide identification of transcript start and end sites by transcript isoform sequencing. *Nat. Protoc.* 9, 1740–1759


Schwalb, B. et al. (2016) TT-seq maps the human transient transcriptome. Science 352, 1225–1228


Drexler, H. L. et al. (2020) Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Mol. Cell 77, 985-998.e8

Glossary

**Attenuation**: The regulatory mechanism operating by premature termination of transcription to cause a reduction in gene expression.

**Chromatin conformation**: Spatial organization of chromatin in a nucleus.

**Compound transcript**: A fused transcript resulting from intervening transcription from an upstream transcript into the downstream ORF.
Cryptic transcript: The uncharacterized RNA transcripts of RNAPII subjected to rapid nuclear degradation.

DNA processing: All the molecular and biochemical events that act on DNA.

Enhancer: Short genomic regions that facilitate the binding of regulatory proteins to increase the transcription of an associated gene. Enhancers can be located far away upstream or downstream (intergenic) or on the same associated gene (intragenic).

Genomic imprinting: A biochemical process that marks the parental origin of epigenetic information.

Histone chaperone: A protein complex that interacts with histone proteins and regulates nucleosome assembly.

Intragenic transcription: Independent transcription occurring within the annotated region of a gene producing intragenic RNAs.

In vivo footprinting: A technique to analyze the DNA-protein interactions in a cell at a given time point.

Nascent transcript: The RNA transcript attached to RNA polymerase during the act of transcription.

Splicing: A molecular process involving the spliceosome complex in the removal of intragenic and non-coding sequences (introns) from primary RNA transcripts.

Transcription factor: Proteins involved in activating or repressing the transcription machinery for gene expression.
In eukaryotes, poly-(A) site (PAS) dependent transcription termination and 3’-end processing of transcripts are tightly coupled [103]. Eukaryotic PAS-dependent RNAPII termination is mechanistically linked to extensive non-coding transcription beyond the 3’-cleavage site [91]. RNAPII reading beyond the 3’-cleavage site generates unstable RNA targeted for co-transcriptional degradation linked to the termination process [103]. Nevertheless, the effects of termination-linked read-through RNAPII transcription in gene regulation await clarification. In yeast and humans, the co-transcriptionally recruited multi-protein Trf4p/Air2p/Mtr4p polyadenylation complex (TRAMP), mediates efficient turnover and processing of a wide range of RNA transcripts. TRAMP interacts with and polyadenylates the RNA substrate to mark it for degradation by the exosome complex [104–107]. The interaction of TRAMP and nuclear exosome functions as an RNA surveillance mechanism of the cell by eliminating the cryptic transcripts arising from pervasive transcription [2,104].

In *S. cerevisiae*, RNAPII termination of small nuclear RNA and IncRNA depends on Nrd1, Nab3, and Sen1 (NNS) proteins [108–110] that terminate specific short sequences in the nascent RNA. It interacts with the cap-binding complex and recruits nuclear exosome for rapid degradation of ncRNA [109–113]. Sequence level conservation of key players in the NNS-dependent termination is limited. However, the Integrator complex appears to fulfil the role of these termination pathways in mammals and plants [114,115]. Elucidation of the mechanisms specifying transcription termination pathways informs experimental approaches to terminate IncRNA transcription. Targeted termination of the upstream transcript is expected to alleviate the downstream gene from repression by tTI mechanisms. In mammalian cells, insertion of a poly-(A) signal near the 5’-end of the IncRNA
may trigger lncRNA transcription termination [68,116]. In plants, insertion of a T-DNA cassette between non-coding and coding DNA also terminates interfering transcription [83]. However, these approaches disrupt the underlying DNA sequence elements. Using targeted roadblocks for transcription may offer an alternative strategy to trigger RNAPII termination without extensive DNA sequence modification. Catalytically dead Cas9 (dCas9) proteins targeted to a specific locus of interest may facilitate RNAPII termination of upstream transcription by acting as a roadblock [117]. However, CRISPR-dCas9-based approaches applied to selectively terminate transcription may have to be evaluated cautiously. dCas9 recruitment may result in the generation of new transcripts [118], and may act most efficiently near TSSs prior to the elongation phase [119]. Application of experimental approaches aiming to terminate RNAPII transcription promise targeted tests of gene repression by tTI.

**BOX 2: Novel methods for RNA detection**

We may still underestimate lncRNA transcription since the detection of unstable RNA molecules represents a formidable challenge. Transcriptome analyses in mutants of RNA degradation factors offer one strategy to address this issue. Genome-wide sequencing strategies capturing the 5’-end of RNA enable mapping of TSSs (i.e., TSS-seq, CAGE). Total RNA extract from cells deficient in nuclear exosome RNA degradation activity reveals TSSs of unstable (i.e. cryptic) transcripts [83,120,121]. To distinguish full-length transcripts and different isoforms, Transcript IsoForm sequencing (TIF-seq) circularizes cDNA of the whole RNA molecule and identifies the 5’ and 3’ ends [115,122].

The specific purification of nascent RNA attached to RNAPII from cells reveals the act of nascent transcription in genomes (NET-seq, mNET-seq, plaNET-seq) [123–125]. Since RNAPII-bound RNA is protected from RNA degradation pathways, nascent RNA mapping
helps to reveal RNAPII transcription at genomic locations without prior transcript annotations, likely representing uncharacterized lncRNA.

Run-on sequencing methods reveal transcripts by allowing incorporation of labeled nucleotides following purification of stalled transcription complexes (GRO-Seq, PRO-seq) [126–128]. An advantage of continuing transcription \textit{in vitro} is the ability to capture short transcripts that are otherwise challenging to map. However, co-transcriptional RNA degradation pathways may be co-purified. Additional refinements permit the selection of capped transcripts, nucleotide resolution, paired-end sequencing, and reactivation of paused polymerases (CoPRO) [129].

Metabolic pulse-labelling of cells by nucleotide base conversion approaches, such as TT-seq, TimeLapse-seq and SLAM-seq, allow us to compare recently generated (i.e. labeled) RNA to total RNA, the latter two even avoiding experimental RNA enrichment [130–132]. Base conversion rates provide information on RNA half-lives, and short labelling times inform on newly synthesized RNA. In whole \textit{Arabidopsis} plants, NEU-seq captures transcriptome dynamics of nuclear and organellar RNA [133].

The combined information of several methods interrogates the presence of lncRNA transcription comprehensively. nano-COP combines 4sU-labelled chromatin-associated RNA for direct RNA nanopore long-read sequencing to analyze co-transcriptional processing [134]. Long-read sequencing methods are continuously improved and offer reduced bias by avoiding reverse transcription and PCR amplification that can limit short-read sequencing. However, short-read sequencing methods will remain influential in the near future to inform on RNAPII transcription.
**FIGURE LEGEND**

**Figure 1.** Tandem transcriptional interference (tTI) affects gene expression. Illustration of the non-coding and coding gene is in red and blue respectively. (A) Productive elongation from upstream non-coding transcription represses the transcription of the downstream coding gene. In budding yeast, downstream gene repression often includes deposition of elongation-specific chromatin marks by the histone methyltransferases Set1 and Set2. Recruitment of elongation-promoting factors such as FACT maintains repression of RNAPII initiation at the gene promoter by promoting productive elongation. (B) Physical displacement of protein complexes by tTI from upstream non-coding transcription. Elongating RNAPII from upstream non-coding transcription triggers the displacement of transcription factors (TFs) from the downstream gene promoter. This mechanism causes repression of the downstream coding gene by allowing productive elongation of the non-coding transcript. (C) The act of transcription from the upstream non-coding unit maintains an open chromatin state. This allows the translocation of RNAPII through the TSS of the downstream coding gene and enables the binding of transcription machinery to activate the downstream gene expression.

**Figure 2.** Antisense IncRNA transcription in the regulation of gene expression. Illustration of the non-coding and coding gene is in red and blue respectively. (A) The on and off-model of sense and antisense transcription. Transcription of antisense IncRNA extends into the promoter of the coding gene and prevents its expression. However, with changes in the environmental conditions, the induction of the coding gene suppresses the expression of the antisense transcript. (B) Antisense IncRNA transcription mediates recruitment of histone methyltransferase, the Set protein, to alter the histone methylation states. The change in
local chromatin state recruits repressive complexes like HDAC (at \textit{GAL1} locus, \textit{S. cerevisiae}), Rpd3S (at \textit{PHO84} locus, \textit{S. cerevisiae}), and PRC (in higher eukaryotes) to inhibit the expression of the coding gene. (C) Promotion of sense gene transcription by induction of looping at \textit{CDC28} locus. Transcription of antisense lncRNA recruits the transcription factor Hog1 and the gene looping factor Ssu72. A gene loop transfers the transcription factor to the promoter and activates sense gene transcription. (D) Collision of RNAPII complexes transcribing the same DNA region in opposing directions. Collision may result in premature termination of the sense transcript or act as a roadblock and reduce gene expression.

Figure 3. Various mechanisms of gene regulation by lncRNA transcription. Illustration of the coding gene is in blue. (A) The act of transcription of the intragenic lncRNA (red) maintains high levels of H3K4me3 in the \textit{ASP3} host gene promoter. The open \textbf{chromatin conformation} through continuous low-level lncRNA transcription promotes rapid induction of gene expression in nitrogen starvation. (B) The transcription of an intragenic enhancer (purple) can activate or attenuate host gene expression. Transcription of intragenic enhancers may result in prematurely terminated transcripts through a mechanism that bears similarity to transcriptional interference. The balance between RNAPII initiation and attenuation of gene expression mediated by enhancer transcription defines the level of fine-tuning. (C) Cellular stress and increased transcription speed trigger defective mRNA termination, resulting in 3'-extended transcripts. These transcripts may extend into neighbouring genes, potentially causing a widespread tTI.

Figure 4. The act of lncRNA transcription impacts DNA replication and repair. Illustration of the non-coding and coding gene is in red and blue respectively. (A) The binding of the origin recognition complex (ORC) to the autonomously replicating sequence (ARS, in yellow)
terminates transcription of adjacent coding or non-coding sequences. (B) Pervasive lncRNA transcription through ARSs causes changes in chromatin, such as an increase of H3K36me3 by Set2, increased nucleosome density and reduced levels of histone acetylation around ARSs. These chromatin-mediated effects reduce ORC binding and the efficiency of ARS firing. (C) After a DNA double-strand break (DSB) event, the MRE11–RAD50–NBS1 (MRN) complex recognizes the lesion and recruits RNAPII. RNAPII synthesizes divergent damage-induced long non-coding RNAs (dilncRNAs) from the DSBs. Small ncRNAs base-pair with dilncRNAs to trigger the formation of DNA damage response foci (blue bubble), stimulating the recruitment of factors through liquid-liquid phase separation.
Figure 2

(A) [Diagram showing molecular interactions]

(B) [Diagram showing protein complexes and interactions]

(C) [Diagram showing circular molecular interactions]

(D) [Diagram showing molecular interactions]
Figure 4

(A) ORC

(B) Set2

(C) MRN