 doi:10.20944/preprints202007.0744.v1

TITLE

DNA processing in the context of non-coding transcription

AUTHOR NAMES

Uthra Gowthaman[1,*], Desiré García-Pichardo[1,*], Yu Jin[1], Isabel Schwarz[1], Sebastian Marquardt[1, 2]

* These authors contributed equally

AFFILIATION

[1] Copenhagen Plant Science Centre

Department of Plant and Environmental Sciences

University of Copenhagen

Bülowsvej 21, 1870 Frederiksberg C

Denmark

CORRESPONDING AUTHOR

[2] sebastian.marquardt@plen.ku.dk (Sebastian Marquardt)

KEYWORDS

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

MAIN TEXT

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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

instance, tTI by two lncRNAs, *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 lncRNA upstream of *IRT1*, in turn represses *IRT1* by tTI to promote *IME1* expression [45]. Likewise, a toggle-switch mechanism between two lncRNA 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 lncRNA *PWR1* alleviates the repression. In fission yeast, lncRNA 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 lncRNA 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 lncRNA *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*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

(Figure 1C). 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

associated with eukaryotic gene promoters [59,65]. Antisense lncRNA 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 lncRNA 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 lncRNA *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 lncRNAs for degradation in the cytoplasm [67]. In budding yeast, stabilization of antisense lncRNAs in an NMD pathway mutant correlates with lower expression of genes, arguing for repressive effects of antisense lncRNA [64]. In differentiated mouse embryonic stem cells, **genomic imprinting** of the *Igf2r* gene is mediated by transcription of the antisense lncRNA *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 lncRNA expression and epigenetic regulation. Collectively, these examples point to mutual inhibition of the antisense lncRNA generated from read-through transcription and mRNA transcription (Figure 2A).

Chromatin-based repression by antisense lncRNA

Regulation of expression of sense gene expression by the corresponding antisense lncRNA 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* lncRNA. The act of *GAL10* lncRNA transcription deposits H3K36me3 and H3K4me2/3 to recruit Rpd3S and **attenuates**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

expression of mRNA in the *GAL1-10* locus [69,70]. Reduced antisense lncRNA transcription increases H3K9ac at the *GAL1* gene, coinciding with increased *GAL1* expression [71]. Similarly, transcription of an antisense lncRNA at the phosphate-repressed *PHO84* gene induces Set1-dependent histone methylation that recruits HDACs to mediate repression of *PHO84* [72]. Chromatin-based repression of *PHO84* by antisense lncRNA may limit the threshold for gene activation to phosphate environments providing a sufficiently strong stimulus [72]. In *Drosophila melanogaster*, the **nascent** antisense lncRNA *ANRIL* recruits Polycomb repressive complex (PRC) to silence the nearby *Ink4b* gene by H3K27me3 [73]. Likewise, the mammalian *ATX7* locus harbors the antisense lncRNA *SCAANT1* whose transcription mediates increased deposition of H3K27me3, a PRC-linked repressive chromatin modification, to silence *ATX7* [74]. In *Arabidopsis,* cold-induced antisense lncRNA *COOLAIR* transcription at the *FLOWERING LOCUS C (FLC)* gene triggers *FLC* repression involving PRC and H3K27me3 [75,76]. Alternative *COOLAIR* co-transcriptional pre-RNA processing through varying poly-adenylation [77] and **splicing** [78] alters the efficiency of antisense mediated *FLC* repression. Single-cell analyses at the *FLC* locus suggest mutually exclusive cellular expression states of sense and antisense lncRNA transcription, arguing for mutual inhibition [79]. In yeast and human, genes with high natural antisense lncRNA 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 lncRNA 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

characteristics with tTI. In summary, the act of antisense lncRNA transcription affects chromatin signaling by diverse mechanisms to affect gene expression.

Emerging mechanisms of gene regulation by the act of antisense transcription

Antisense lncRNA transcription in budding yeast at the *CDC28* locus represents an example of regulation by gene looping that correlates positively with gene expression [81]. Antisense lncRNA 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 lncRNA 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 lncRNA *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 lncRNA 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 lncRNA-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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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 lncRNA transcription. In summary, the mechanistic diversity of gene regulation by antisense lncRNA 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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 poly-adenylation 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

297 stable association of the origin recognition complex (ORC) and the pre-replication complex
298 at ARSs promote RNAPII pausing by a roadblock mechanism and mediate transcriptional
299 termination at the edges of replication origins [99,100] (Figure 4A). ARS invasion by lncRNA
300 transcription correlates with low DNA replication efficiency and late origin firing, affecting the
301 temporal program of replication [99,100]. Mechanistically, RNAPII transcription through
302 ARSs affects replication in part by chromatin-mediated effects. Akin to tTI, RNAPII
303 elongation mediates a Set2-dependent increase of H3K36me3, a decrease of histone
304 acetylation and an increase in nucleosome occupancy around the ARSs [100]. This
305 chromatin landscape correlates with lower efficiency of ORC binding and inefficient ARS
306 firing [100] (Figure 4B). Future studies should investigate if tTI-like effects on DNA
307 replication, as detected in *S. cerevisiae,* represent a phenomenon preserved through
308 evolution. In summary, the act of lncRNA transcription over ARSs inhibits DNA replication
309 and thus cell cycle progression.
310 In human cells, the act of lncRNA transcription is essential for the repair of double-strand
311 breaks (DSBs), a harmful type of DNA lesion [101]. Transcription occurs in the first steps of
312 the DNA damage response (DDR). DDR couples a series of mechanisms responsible for
313 the recognition, signaling and repair of DNA damage. Detection of the lesion by the MRE11–
314 RAD50–NBS1 (MRN) complex allows the formation of functional promoters, where RNAPII
315 bidirectionally synthesizes damage-induced lncRNAs (dilncRNAs) [101,102]. dilncRNAs are
316 processed into small ncRNAs that base-pair with nascent dilncRNA to trigger the formation
317 of DDR foci, necessary for DSB repair [101]. Application of inhibitors of RNAPII transcription
318 revealed the requirement of dilncRNA transcription for the formation of DDR foci, and that
319 dilncRNA promote liquid-liquid phase-separation to concentrate repair factors around DDR

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

foci [101,102] (Figure 4C). These examples support functional roles for the act of lncRNA transcription in DNA biology beyond the regulation of gene expression.

Concluding remarks

Research focus is shifting from the discovery of lncRNAs 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 lncRNA transcription regardless of the DNA sequence. While most evidence connects the act of lncRNA 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 lncRNAs 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 lncRNA 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 lncRNA 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 lncRNA 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

1 Djebali, S. *et al.* (2012) Landscape of transcription in human cells. *Nature* 489, 101–108

2 Jensen, T. H. *et al.* (2013) Dealing with pervasive transcription. *Mol. Cell* 52, 473–484

3 Quinn, J. J. and Chang, H. Y. (2016) Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* 17, 47–62

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

5 Rinn, J. L. and Chang, H. Y. (2020) Long Noncoding RNAs: Molecular Modalities to Organismal Functions. *Annu. Rev. Biochem.* 89, 283-308

6 Kornienko, A. E. *et al.* (2013) Gene regulation by the act of long non-coding RNA transcription. *BMC Biology* 11, 59

7 Ard, R. *et al.* (2017) Emerging properties and functional consequences of noncoding transcription. *Genetics* 207, 357–367

8 Pang, K. C. *et al.* (2006) Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. *Trends Genet.* 22, 1–5

9 Geisler, S. and Coller, J. (2013) RNA in unexpected places: Long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* 14, 699–712

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

nuclear structure and gene expression. *Nat. Rev. Mol. Cell Biol.* 17, 756–770

11 Kopp, F. and Mendell, J. T. (2018) Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172, 393–407

12 Field, A. and Adelman, K. (2020) Evaluating Enhancer Function and Transcription. *Annu. Rev. Biochem.* 89, 213-234

13 Bannister, A. J. and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Research* 21, 381–395

14 Smolle, M. and Workman, J. L. (2013) Transcription-associated histone modifications and cryptic transcription. *Biochim. Biophys. Acta* 1829, 84–97

15 Buratowski, S. (2009) Progression through the RNA polymerase II CTD cycle. *Mol. Cell* 36, 541–546

16 Li, B. *et al.* (2007) The Role of Chromatin during Transcription. *Cell* 128, 707–719

17 Leng, X. *et al.* (2020) A G(enomic)P(ositioning)S(ystem) for Plant RNAPII Transcription. *Trends Plant Sci.* 25, 744-764

18 Carrozza, M. J. *et al.* (2003) The diverse functions of histone acetyltransferase complexes. *Trends Genet.* 19, 321–329

19 Ng, H. H. *et al.* (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* 11, 709–719

20 Krogan, N. J. *et al.* (2003) Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. *Mol. Cell*

Biol. 23, 4207–4218

21 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

22 Bannister, A. J. *et al.* (2005) Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* 280, 17732–17736

23 Pokholok, D. K. *et al.* (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122, 517–527

24 Weiner, A. *et al.* (2015) High-resolution chromatin dynamics during a yeast stress response. *Mol. Cell* 58, 371–386

25 Barski, A. *et al.* (2007) High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 129, 823–837

26 Kharchenko, P. V. *et al.* (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* 471, 480–486

27 Sinha, I. *et al.* (2006) Genome-wide patterns of histone modifications in fission yeast. *Chromosom. Res.* 14, 95–105

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

29 Guenther, M. G. *et al.* (2007) A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* 130, 77–88

30 Carrozza, M. J. *et al.* (2005) Histone H3 methylation by Set2 directs deacetylation of

406 coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581–
407 592

408 31 Keogh, M. C. *et al.* (2005) Cotranscriptional set2 methylation of histone H3 lysine 36
409 recruits a repressive Rpd3 complex. *Cell* 123, 593–605

410 32 Venkatesh, S. *et al.* (2012) Set2 methylation of histone H3 lysine 36 suppresses
411 histone exchange on transcribed genes. *Nature* 489, 452–455

412 33 Reinberg, D. and Sims, R. J. (2006) De FACTo nucleosome dynamics. *J. Biol. Chem.*
413 281, 23297–23301

414 34 Kaplan, C. D. *et al.* (2003) Transcription elongation factors repress transcription
415 initiation from cryptic sites. *Science* 301, 1096–1099

416 35 Carvalho, S. *et al.* (2013) Histone methyltransferase SETD2 coordinates FACT
417 recruitment with nucleosome dynamics during transcription. *Nucleic Acids Res.* 41,
418 2881–2893

419 36 Nielsen, M. *et al.* (2019) Transcription-driven chromatin repression of Intragenic
420 transcription start sites. *PLoS Genet.* 15, e1007969

421 37 Schlackow, M. *et al.* (2017) Distinctive Patterns of Transcription and RNA Processing
422 for Human *lincRNAs. Mol. Cell 65, 25–38*

423 38 Shearwin, K. E. *et al. (2005)* Transcriptional interference - A crash course. *Trends*
424 *Genet. 21, 339–345*

425 39 Hainer, S. J. *et al.* (2011) Intergenic transcription causes repression by directing
426 nucleosome assembly. *Genes Dev.* 25, 29–40

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

41 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

42 Hollerer, I. *et al*. (2019) Evidence for an integrated gene repression mechanism based on mRNA isoform toggling in human cells. *G3* 9, 1045–1053

43 Jorgensen, V. *et al*. (2020) Tunable Transcriptional Interference at the Endogenous Alcohol Dehydrogenase Gene Locus in Drosophila melanogaster. *G3* 10, 1575–1583

44 Ard, R. and Allshire, R. C. (2016) Transcription-coupled changes to chromatin underpin gene silencing by transcriptional interference. *Nucleic Acids Res*. 44, 10619–10630

45 Moretto, F. *et al.* (2018) A regulatory circuit of two lncRNAs and a master regulator directs cell fate in yeast. *Nat. Commun.* 9, 780

46 Kim, J. *et al*. (2016) Modulation of mRNA and lncRNA expression dynamics by the Set2–Rpd3S pathway. *Nat. Commun.* 7, 13534

47 Van Werven, F. J. *et al.* (2012) Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* 150, 1170–1181

48 Bumgarner, S. L. *et al.* (2009) Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 106, 18321–18326

49 Bumgarner, S. L. *et al.* (2012) Single-Cell Analysis Reveals that Noncoding RNAs Contribute to Clonal Heterogeneity by Modulating Transcription Factor Recruitment.

Mol. Cell 45, 470–482

50 Bird, A. J. *et al.* (2006) Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. *EMBO J.* 25, 5726–5734

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

52 Hainer, S. J. *et al.* (2012) Identification of mutant versions of the Spt16 histone chaperone that are defective for transcription-coupled nucleosome occupancy in saccharomyces cerevisiae. *G3* 2, 555–567

53 Martens, J. A. *et al.* (2005) Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. *Genes Dev.* 19, 2695–2704

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

55 Feng, J. *et al.* (2016) Noncoding Transcription Is a Driving Force for Nucleosome Instability in spt16 Mutant Cells. *Mol. Cell Biol.* 36, 1856–1867

56 Hirota, K. *et al.* (2008) Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* 456, 130–134

57 Du Mee, D. J. M. *et al.* (2018) Efficient termination of nuclear lncrna transcription promotes mitochondrial genome maintenance. *Elife* 7, e31989

58 Pelechano, V. and Steinmetz, L. M. (2013) Gene regulation by antisense transcription. *Nat. Rev. Genet.* 14, 880–893

59 Mellor, J. *et al.* (2016) The Interleaved Genome. *Trends Genet.* 32, 57–71

60 Xu, Z. *et al.* (2011) Antisense expression increases gene expression variability and locus interdependency. *Mol. Syst. Biol.* 7, 468

61 Murray, S. C. *et al.* (2012) A pre-initiation complex at the 3′-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res.* 40, 2432–2444

62 Beltran, M. *et al.* (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev.* 22, 756–769

63 Wery, M. *et al.* (2018) Bases of antisense lncRNA-associated regulation of gene expression in fission yeast. *PLoS Genet.* 14, e1007465

64 Nevers, A. *et al.* (2018) Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. *Nucleic Acids Res*. 46, 6009–6025

65 Marquardt, S. *et al.* (2014) A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* 157, 1712–1723

66 Gelfand, B. *et al.* (2011) Regulated Antisense Transcription Controls Expression of Cell-Type-Specific Genes in Yeast. *Mol. Cell Biol.* 31, 1701–1709

67 Wery, M. *et al.* (2016) Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. *Mol. Cell* 61, 379–392

68 Latos, P. A. *et al.* (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338, 1469–1472

491 69 Houseley, J. *et al.* (2008) A ncRNA Modulates Histone Modification and mRNA
492 Induction in the Yeast GAL Gene Cluster. *Mol. Cell* 32, 685–695

493 70 Pinskaya, M. *et al.* (2009) H3 lysine 4 di- and tri-methylation deposited by cryptic
494 transcription attenuates promoter activation. *EMBO J*. 28, 1697–1707

495 71 Brown, T. *et al.* (2018) Antisense transcription-dependent chromatin signature
496 modulates sense transcript dynamics. *Mol. Syst. Biol.* 14, e8007

497 72 Castelnuovo, M. *et al.* (2013) Bimodal expression of PHO84 is modulated by early
498 termination of antisense transcription. *Nat. Struct. Mol. Biol.* 20, 851–858

499 73 Yap, K. L. *et al.* (2010) Molecular Interplay of the Noncoding RNA ANRIL and
500 Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of
501 INK4a. *Mol. Cell* 38, 662–674

502 74 Sopher, B. L. *et al.* (2011) CTCF Regulates Ataxin-7 Expression through Promotion
503 of a Convergently Transcribed, Antisense Noncoding RNA. *Neuron* 70, 1071–1084

504 75 Swiezewski, S. *et al.* (2009) Cold-induced silencing by long antisense transcripts of
505 an Arabidopsis Polycomb target. *Nature* 462, 799–802

506 76 Csorba, T. *et al.* (2014) Antisense COOLAIR mediates the coordinated switching of
507 chromatin states at FLC during vernalization. *Proc. Natl. Acad. Sci. U. S. A.* 111,
508 16160–16165

509 77 Liu, F. *et al.* (2010) Targeted 3′ processing of antisense transcripts triggers
510 arabidopsis FLC chromatin silencing. *Science* 327, 94–97

511 78 Marquardt, S. *et al.* (2014) Functional Consequences of Splicing of the Antisense
512 Transcript COOLAIR on FLC Transcription. *Mol. Cell* 54, 156–165

513 79 Rosa, S. *et al.* (2016) Mutually exclusive sense-antisense transcription at FLC
514 facilitates environmentally induced gene repression. *Nat. Commun.* 7, 13031

515 80 Gill, J. K. *et al.* (2020) Fine Chromatin-Driven Mechanism of Transcription Interference
516 by Antisense Noncoding Transcription. *Cell Rep.* 31, 107612

517 81 Nadal-Ribelles, M. *et al.* (2014) Control of Cdc28 CDK1 by a Stress-Induced lncRNA.
518 *Mol. Cell* 53, 549–561

519 82 Hobson, D. J. *et al.* (2012) RNA Polymerase II Collision Interrupts Convergent
520 Transcription. *Mol. Cell* 48, 365–374

521 83 Kindgren, P. *et al.* (2018) Transcriptional read-through of the long non-coding RNA
522 SVALKA governs plant cold acclimation. *Nat. Commun.* 9, 4561

523 84 De Jesus, B. B. *et al.* (2018) Silencing of the lncRNA Zeb2-NAT facilitates
524 reprogramming of aged fibroblasts and safeguards stem cell pluripotency. *Nat.*
525 *Commun.* 9, 94

526 85 Yeganeh, M. *et al.* (2017) Transcriptional interference by RNA polymerase III affects
527 expression of the Polr3e gene. *Genes Dev.* 31, 413-421

528 86 Huang, Y. C. *et al.* (2010) Intragenic transcription of a noncoding RNA modulates
529 expression of ASP3 in budding yeast. *RNA* 16, 2085–2093

530 87 Cinghu, S. *et al.* (2017) Intragenic Enhancers Attenuate Host Gene Expression. *Mol.*
531 *Cell* 68, 104-117

532 88 Onodera, C. S. *et al.* (2012) Gene isoform specificity through enhancer-associated
533 antisense transcription. *PLoS One* 7, e43511

534 89 Andersson, R. and Sandelin, A. (2020) Determinants of enhancer and promoter
535 activities of regulatory elements. *Nat. Rev. Genet.* 21, 71–87

536 90 Dhir, A. *et al.* (2015) Microprocessor mediates transcriptional termination of long
537 noncoding RNA transcripts hosting microRNAs. *Nat. Struct. Mol. Biol.* 22, 319–327

538 91 Proudfoot, N. J. (2016) Transcriptional termination in mammals: Stopping the RNA
539 polymerase II juggernaut. *Science* 352, aad9926

540 92 Vilborg, A. *et al.* (2015) Widespread Inducible Transcription Downstream of Human
541 Genes. *Mol. Cell* 59, 449–461

542 93 Fukudome, A. *et al.* (2017) Salt stress and CTD PHOSPHATASE-LIKE4 mediate the
543 switch between production of small nuclear RNAs and mRNAs. *Plant Cell* 29, 3214–
544 3233

545 94 Bauer, D. L. V. *et al.* (2018) Influenza Virus Mounts a Two-Pronged Attack on Host
546 RNA Polymerase II Transcription. *Cell Rep*. 23, 2119-2129.e3

547 95 Wang, X. *et al.* (2020) Herpes simplex virus blocks host transcription termination via
548 the bimodal activities of ICP27. *Nat. Commun.* 11, 293

549 96 Hennig, T. *et al.* (2018) HSV-1-induced disruption of transcription termination
550 resembles a cellular stress response but selectively increases chromatin accessibility
551 downstream of genes. *PLOS Pathog.* 14, e1006954

552 97 Grosso, A. R. *et al.* (2015) Pervasive transcription read-through promotes aberrant
553 expression of oncogenes and RNA chimeras in renal carcinoma. *Elife* 4, e09214

554 98 Leng, X. *et al.* (2020) Organismal benefits of transcription speed control at gene
555 boundaries. *EMBO Rep.* 21, e49315

556 99 Candelli, T. *et al.* (2018) Perv*asi*ve transcription fine-tunes replication origin activity.
557 *Elife* 7, e40802

558 100 Soudet, J. *et al.* (2018) Noncoding transcription influences the replication initiation
559 program through chromatin regulation. *Genome Res.* 28, 1867–1881

560 101 Michelini, F. *et al.* (2017) Damage-induced lncRNAs control the DNA damage
561 response through interaction with DDRNAs at individual double-strand breaks. *Nat.*
562 *Cell Biol.* 19, 1400–1411

563 102 Pessina, F. *et al.* (2019) Functional transcription promoters at DNA double-strand
564 breaks mediate RNA-driven phase separation of damage-response factors. *Nat. Cell*
565 *Biol.* 21, 1286–1299

566 103 Mischo, H. E. and Proudfoot, N. J. (2013) Disengaging polymerase: Terminating RNA
567 polymerase II transcription in budding yeast. *Biochim. Biophys.* Acta 1829, 174–185

568 104 LaCava, J. *et al.* (2005) RNA degradation by the exosome is promoted by a nuclear
569 polyadenylation complex. *Cell* 121, 713–724

570 105 Wyers, F. *et al.* (2005) Cryptic Pol II transcripts are degraded by a nuclear quality
571 control pathway involving a new poly(A) polymerase. *Cell* 121, 725–737

572 106 Wery, M. *et al.* (2009) The nuclear poly(A) polymerase and Exosome cofactor Trf5 is
573 recruited cotranscriptionally to nucleolar surveillance. *RNA* 15, 406–419

574 107 Ogami, K. *et al.* (2018) RNA surveillance by the nuclear RNA exosome: Mechanisms
575 and significance. *Noncoding RNA* 4, 8

576 108 Steinmetz, E. J. *et al.* (2001) RNA-binding protein Nrd1 directs poly(A)-independent
577 3′-end formation of RNA polymerase II transcripts. *Nature* 413, 327–331

578 109 Vasiljeva, L. and Buratowski, S. (2006) Nrd1 interacts with the nuclear exosome for
579 3′ processing of RNA polymerase II transcripts. *Mol. Cell* 21, 239–248

580 110 Porrua, O. and Libri, D. (2015) Transcription termination and the control of the
581 transcriptome: Why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* 16, 190–202

582 111 Thiebaut, M. *et al.* (2006) Transcription Termination and Nuclear Degradation of
583 Cryptic Unstable Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome
584 Surveillance. *Mol. Cell* 23, 853–864

585 112 Tudek, A. *et al.* (2014) Molecular basis for coordinating transcription termination with
586 noncoding RNA degradation. *Mol. Cell* 55, 467–481

587 113 Fasken, M. B. *et al.* (2015) Nab3 Facilitates the Function of the TRAMP Complex in
588 RNA Processing via Recruitment of Rrp6 Independent of Nrd1. *PLoS Genet*. 11,
589 e1005044

590 114 Tatomer, D. C. *et al.* (2019) The Integrator complex cleaves nascent mRNAs to
591 attenuate transcription. *Genes Dev.* 33, 1525–1538

592 115 Thomas, Q. A. *et al.* (2020) Transcript isoform sequencing reveals widespread
593 promoter-proximal transcriptional termination in Arabidopsis. *Nat. Commun.* 11, 2589

594 116 Paralkar, V. R. *et al.* (2016) Unlinking an lncRNA from Its Associated cis Element. *Mol.*
595 *Cell* 62, 104–110

596 117 Qi, L. S. *et al*. (2013) Repurposing CRISPR as an RNA-Guided Platform for
597 Sequence-Specific Control of Gene Expression. *Cell* 152, 1173–1183

598 118 Howe, F. S. *et al.* (2017) CRISPRi is not strand-specific at all loci and redefines the
599 transcriptional landscape. *Elife* 6, e29878

600 119 Gilbert, L.A. *et al.* (2014) Genome-Scale CRISPR-Mediated Control of Gene
601 Repression and Activation. *Cell* 159, 647–661

602 120 Takahashi, H. *et al.* (2012) 5′ end-centered expression profiling using cap-analysis
603 gene expression and next-generation sequencing. *Nat. Protoc.* 7, 542–561

604 121 Ntini, E. *et al.* (2013) Polyadenylation site-induced decay of upstream transcripts
605 enforces promoter directionality. *Nat. Struct. Mol. Biol.* 20, 923–928

606 122 Pelechano, V. *et al.* (2014) Genome-wide identification of transcript start and end sites
607 by transcript isoform sequencing. *Nat. Protoc.* 9, 1740–1759

608 123 Churchman, L. S. and Weissman, J. S. (2011) Nascent transcript sequencing
609 visualizes transcription at nucleotide resolution. *Nature* 469, 368–373

610 124 Nojima, T. *et al.* (2015) Mammalian NET-seq reveals genome-wide nascent
611 transcription coupled to RNA processing. *Cell* 161, 526–540

612 125 Kindgren, P. *et al.* (2020) Native elongation transcript sequencing reveals temperature
613 dependent dynamics of nascent RNAPII transcription in Arabidopsis. *Nucleic Acids*
614 *Res.* 48, 2332–2347

615 126 Core, L. J. *et al.* (2008) Nascent RNA sequencing reveals widespread pausing and
616 divergent initiation at human promoters. *Science* 322, 1845–1848

617 127 Kwak, H. *et al.* (2013) Precise maps of RNA polymerase reveal how promoters direct
618 initiation and pausing. *Science* 339, 950–953

619 128 Zhu, J. *et al.* (2018) RNA polymerase II activity revealed by GRO-seq and pNET-seq
620 in Arabidopsis. *Nat. Plants* 4, 1112–1123

129 Tome, J. M. *et al.* (2018) Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat. Genet*. 50, 1533–1541

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

131 Herzog, V. A. *et al.* (2017) Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* 14, 1198–1204

132 Schofield, J. A. *et al.* (2018) TimeLapse-seq: Adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat. Methods* 15, 221–225

133 Szabo, E. X. *et al.* (2020) Metabolic Labeling of RNAs Uncovers Hidden Features and Dynamics of the Arabidopsis Transcriptome. *Plant Cell* 32, 871-887

134 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

BOX 1: RNAPII transcription termination and tTI

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 lncRNA 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 lncRNA 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 lncRNA

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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 *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 *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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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 lncRNA 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 lncRNA 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 lncRNA 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 *GAL1* locus, *S. cerevisiae)*, Rpd3S (at *PHO84* locus, *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 *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 *ASP3* host gene promoter. The open **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.

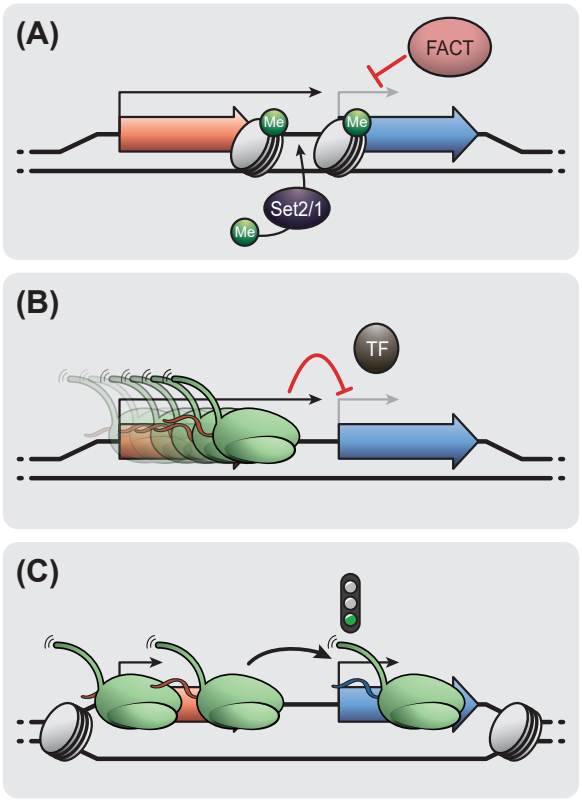
Figure 4. The act of lncRNA transcription impacts DNA replication and repair. Illustration of the non-coding and coding gene is in red and blue respectively. (A) The binding of the origin recognition complex (ORC) to the autonomously replicating sequence (ARS, in yellow)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

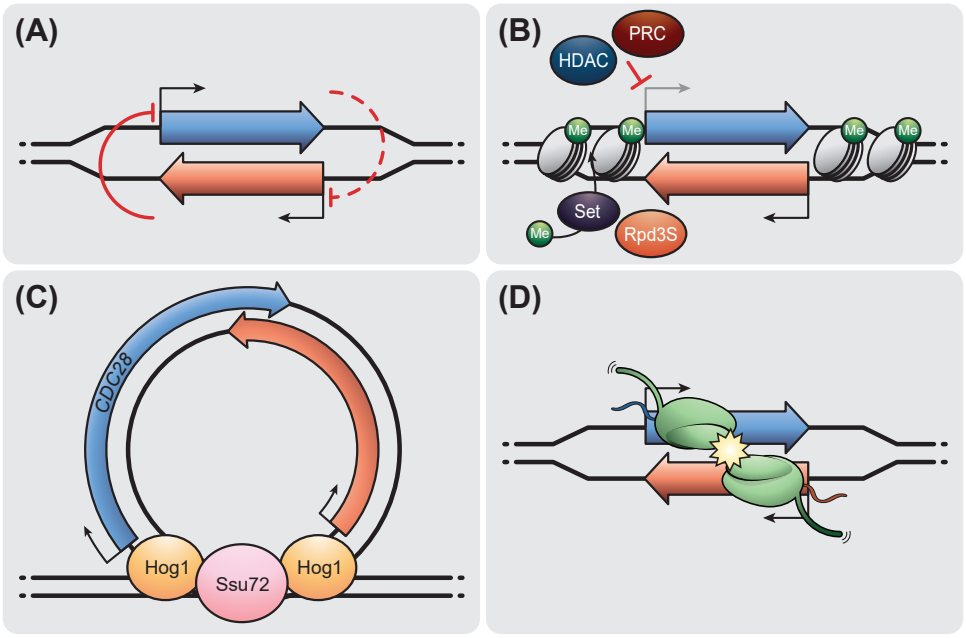
Figure 1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

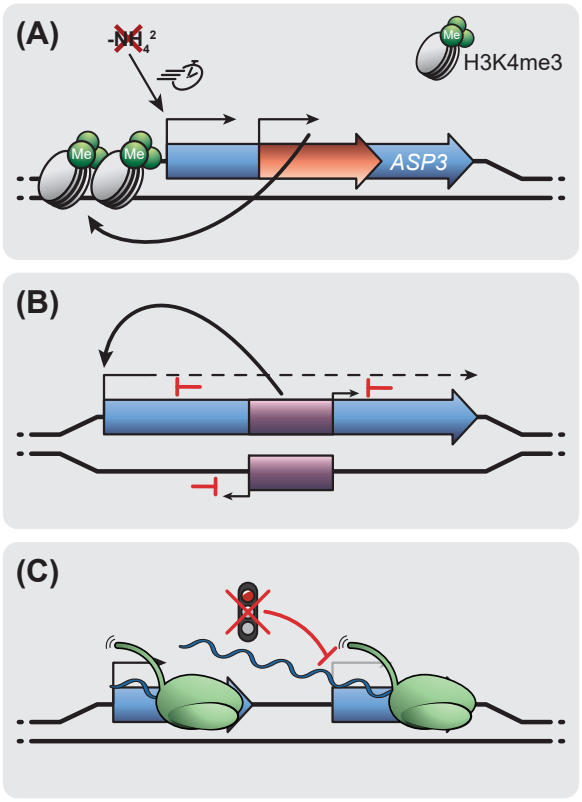
Figure 2

Gowthaman et al.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

Figure 3



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 July 2020 doi:10.20944/preprints202007.0744.v1

Figure 4