RNA, Genome Output and Input

Jörg Morf[1]*, Srinjan Basu[1] and Paulo P. Amaral[2]

[1] Jeffrey Cheah Biomedical Centre, Wellcome – Medical Research Council (MRC) Cambridge Stem Cell Institute, University of Cambridge, Cambridge, United Kingdom, [2] Jeffrey Cheah Biomedical Centre, The Milner Therapeutics Institute, University of Cambridge, Cambridge, United Kingdom

RNA, the transcriptional output of genomes, not only templates protein synthesis or directly engages in catalytic functions, but can feed back to the genome and serve as regulatory input for gene expression. Transcripts affecting the RNA abundance of other genes act by mechanisms similar to and in concert with protein factors that control transcription. Through recruitment or blocking of activating and silencing complexes to specific genomic loci, RNA and protein factors can favor transcription or lower the local gene expression potential. Most regulatory proteins enter nuclei from all directions to start the search for increased affinity to specific DNA sequences or to other proteins nearby genuine gene targets. In contrast, RNAs emerge from spatial point sources within nuclei, their encoding genes. A transcriptional burst can result in the local appearance of multiple nascent RNA copies at once, in turn increasing local nucleic acid density and RNA motif abundance before diffusion into the nuclear neighborhood. The confined initial localization of regulatory RNAs causing accumulation of protein co-factors raises the intriguing possibility that target specificity of non-coding, and probably coding, RNAs is achieved through gene/RNA positioning and spatial proximity to regulated genomic regions. Here we review examples of positional cis conservation of regulatory RNAs with respect to target genes, spatial proximity of enhancer RNAs to promoters through DNA looping and RNA-mediated formation of membrane-less structures to control chromatin structure and expression. We speculate that linear and spatial proximity between regulatory RNA-encoding genes and gene targets could possibly ease the evolutionary pressure on maintaining regulatory RNA sequence conservation.

Keywords: RNA, lncRNA, gene regulation, chromosome conformation, transcriptional bursting

INTRODUCTION

Many mechanisms for RNAs to regulate gene expression in the cell nucleus involve recruitment of regulatory protein factors, including chromatin modifiers and polymerase recruiters that affect the transcriptional output of genes (Ulitsky and Bartel, 2013). Non-coding RNAs above the length range of small RNAs exemplified by miRNAs and clearly distinguishable from transcription units of protein-coding loci, i.e., intergenic, have been studied extensively in the past. The definition of and focus on long non-coding RNAs (lncRNAs) facilitates the functional characterization of how RNA molecules affect the expression of genes whilst avoiding ambiguities arising from the bifunctionality of coding RNAs, i.e., an RNA with regulatory potential simultaneously encoding a protein with a certain function. However, it is unlikely that most nuclear complexes and machineries with affinity toward RNA distinguish transcripts primarily based on their coding potential, and roles

in regulatory relationships between coding and non-coding transcripts could therefore be assumed as interchangeable (Li and Liu, 2019).

Most regulatory proteins, such as transcription factors, enter nuclei after their synthesis through pores to eventually interact specifically, or broadly, with chromatin regions. RNAs, on the other hand, emerge from their encoding gene at a defined genomic and spatial position. Therefore, the position of origin for a regulatory RNA and the spatial genome neighborhood are arguably critical in defining target gene specificity for such transcripts. Particularly RNAs with shorter half-lives might exert roles in gene expression regulation restricted to nuclear regions in immediate vicinity to their gene locus encompassing neighboring genes in cis or loci brought into close proximity in the spatial genome structure through chromatin looping or DNA contacts in trans.

Once a polymerase engages in processive transcription, the synthesized, nascent transcript appears from the encoding gene locus. The RNA molecule grows in length with continuing transcription along the gene but stays tethered to chromatin by the polymerase until 3' RNA cleavage followed by polyadenylation releases the RNA molecule (Cramer, 2019). At a polymerase elongation rate usually between 1 and 4 kilobases per minute and a median gene length of around 24 kilobases in human cells, at least the 5' region of a nascent transcript is extruding chromatin while still tethered to it for a duration in the order of 10 min (Milo et al., 2010). Immediately after initiation of RNA synthesis, proteins with RNA-binding domains can interact co-transcriptionally with 5' ends of nascent transcripts. Interestingly, sequence conservation of non-coding RNAs, which is overall low compared to mRNAs, increases toward the 5' ends of the molecules, which raises the possibility that the longer half-life in chromatin association of 5' RNA regions compared to 3' ends has been co-opted to more efficiently recruit regulatory protein factors to chromatin through interactions with nascent, tethered RNA (Hezroni et al., 2015).

Nascent RNAs emerge as groups of multiple molecules in a short time window, so-called transcriptional bursts, which results in the amplification of available chromatin-tethered RNA binding sites and of protein recruitment to a given locus in that moment. Bursting, or discontinuous transcription, of active genes describes the temporal gating of transcription initiation into time windows of a few minutes and the interspersion of such "on" states with longer periods of promoter inactivity, or so-called "off" states (Rodriguez and Larson, 2020). First insights into discontinuous transcription have been gained by electron microscopy of chromosome spreading preparations (McKnight and Miller, 1979). More recently, bursting parameters such as frequency and burst size, the number of transcription initiations during a burst, have been measured by single-molecule RNA-FISH, short-lived protein reporters, MS2-RNA tagging and single-cell RNA-seq (McKnight and Miller, 1979; Raj et al., 2006; Suter et al., 2011; Tantale et al., 2016; Larsson et al., 2019). Typically, bursts measured in mammalian cells have frequencies in the order of one burst every 30 min up to several hours and last for a couple of minutes. RNA polymerases start transcribing in groups with inter-polymerase distances of a few hundred bases

during on states, which gives rise to up to hundreds of nascent transcripts emerging from and tethered to chromatin during the time required for polymerases to reach the 3' end of a gene (Dar et al., 2012; Tantale et al., 2016; Nicolas et al., 2017).

Intriguingly, a recent preprint applying RNA-FISH combined with expansion microscopy revealed that transcripts after completion of synthesis and chromatin dissociation remain locally restricted within sub-micron distances from gene loci for some time (Coté et al., 2020). The absence of gradients of decreasing RNA concentration from the encoding gene contrasts the notion of immediate free diffusion or transport away from genes after transcription termination and 3' RNA end processing. Such a delay in transcript re-localization after synthesis would further increase the chromatin residence time of transcripts.

Once regulatory transcripts escape localization to the vicinity of their encoding gene locus, the target gene search is expected to rely primarily on differential affinities to for example different DNA sequences, chromatin modifications and other chromatin-associated factors, comparable to regulatory proteins entering the nucleus through pores.

In summary, transcriptional bursts locally increase RNA concentration throughout the time of synthesis when RNA is tethered to chromatin and likely longer in an untethered state in the immediate vicinity of encoding genome loci. As a consequence, a high density of locally confined single-, double-stranded and structural RNA motifs presents itself at transcription units to concentrate and position nucleic acid-binding proteins with gene regulatory functions within the three-dimensional genome structure (**Figures 1A–C**). Indeed, many transcription factors have RNA-binding capacity and, vice versa, nuclear RNA-binding proteins are frequently found localized to chromatin (Cassiday and Maher, 2002; Hudson and Ortlund, 2014; Xiao et al., 2019).

The information of a single allelic genome motif can therefore be locally amplified in the orders of 10–100-fold when converted into RNA during a single transcriptional burst. However, whether the regulatory potential of nascent RNA is realized might depend on factors that, besides transcription and degradation kinetics, include the sequence and structural features of the RNA, RNA modifications and, importantly, linear or spatial proximity to potential target loci. In the following paragraphs we revisit a few illustrative examples of regulatory RNAs and their effects on the transcriptional output of genes encoded in close distance on the same chromosome or in proximity either through chromatin looping or gene positioning to nuclear bodies.

POSITIONAL CONSERVATION OF CIS-REGULATORY RNAs AND TARGET GENES

A case in point is illustrated by lncRNAs whose genomic position relative to neighboring genes is conserved (synteny) (Ulitsky, 2016). The classic lncRNA *Xist*, which forms RNA clouds covering exclusively the inactive X chromosome was the first regulatory lncRNA found to display genomic positional conservation across species (Brockdorff et al., 1991; Brown et al.,

FIGURE 1 | Localization and enrichment of regulatory or structural proteins by RNA. **(A)** The transcriptional output of a regulatory RNA-encoding gene can serve as regulatory input for genes in spatial proximity. Various parameters including transcription kinetics and chromatin association might define the spatial reach of the regulatory RNA (dotted gray circle). **(B)** Regulatory RNAs can affect expression of proximal genes either negatively (left) or positively (right). During a burst, the appearance of multiple, chromatin-associated regulatory RNAs attracts and localizes protein factors in proximity to the target gene to increase or decrease its transcriptional output. **(C)** Architectural RNAs with large burst sizes and multiple affinity sites for structural proteins, which form multivalent interactions when concentrated in close proximity, are envisaged to seed membrane-less structures co-transcriptionally. **(D)** The large burst sizes of known non-coding RNA genes (blue lines) are highlighted in comparison to burst sizes of all protein-coding genes (gray distribution, data from Larsson et al., 2019). Of note, despite overall lower burst frequencies for ncRNAs than protein-coding RNAs, distributions in burst size are similar (Kouno et al., 2019; Larsson et al., 2019).

1991, 1992). Thereafter, thousands of additional non-coding RNAs, whose genomic position, promoters and tissue expression profiles are conserved between human and mouse, have been identified and named positionally-conserved RNAs (pcRNAs) (Amaral et al., 2018). Most pcRNAs locate to chromatin loop anchor points and borders of topologically associating domains (TADs), occupied by the CTCF chromatin organizer and of

increased gene density in cis and 3D for lncRNAs to act upon (Kim et al., 2007). Their expression is correlated to the abundance of transcripts from neighboring genes, which are predominantly developmental genes. Experimental reduction of the RNA levels of several of these pcRNAs in different human and cancer cell lines leads to downregulation of the syntenic protein-coding gene, indicating positive regulatory roles in cis for pcRNAs.

Indeed, co-expression and cis-regulation of associated genes by neighboring lncRNAs is a common mode of regulation (Guil and Esteller, 2012; Gil and Ulitsky, 2020).

The example of pcRNA *Evx1-as* illustrates a possible sequence of steps that leads to the upregulation of *Evx1* target transcription in cis. *Evx1-as* pcRNA and *Evx1* coding transcripts are co-expressed in the primitive streak of mouse embryos (Bell et al., 2016). The lncRNA first recruits MLL chromatin modifiers, which deposit locally H3K4me3 histone marks upon differentiation of mouse embryonic stem cells toward mesoderm (Dinger et al., 2008). Increased H3K4me3 levels and subsequent recruitment of transcriptional activators, such as the Mediator complex, is followed by the induction of *Evx1* transcription (Luo et al., 2016). The importance of the proximity between the lncRNA and the target gene is highlighted by gain-of-function experiments. Consistent with a genome position-dependent role, ectopic over-expression and mis-localization of *Evx1-as* does not affect *Evx1* levels (Luo et al., 2016). Reduction in *Evx1-as* RNA levels phenocopies loss-of-function of *Evx1* (Bell et al., 2016; Luo et al., 2016). Similar modes of action have been identified for *HOTTIP, HoxBlinc*, and *HOTAIRM1* regulating multiple neighboring *HOX* genes in expression domains during cell differentiation (Zhang et al., 2009; Wang et al., 2011; Deng et al., 2016; Wang and Dostie, 2017) and for lncRNAs, such as *UMLILO*, priming the robust expression of proximal immune genes (Fanucchi et al., 2019).

Interestingly, in some cases the lncRNA and the neighboring gene act in concert in feedback loops. LncRNA *Deanr1*, encoded adjacent to *Foxa2*, recruits SMAD2/3 transcription factors to the *Foxa2* promoter to activate coding gene transcription (Luo et al., 2016). Reciprocally, reduction of *Foxa2* coding RNA results in decreased expression levels of its neighboring lncRNA. It is unclear to date whether *Foxa2* RNA mediates the effect on lncRNA expression or whether FOXA2 protein and its chromatin binding sites at its own promoter and that of the lncRNA reinforces the expression of both (Amaral et al., 2018).

As exemplified by *Xist*, cis-acting RNAs may also function as negative regulators of neighboring gene expression. Repression in *cis* commonly underlies regulation of imprinted loci associated with positionally conserved lncRNAs—as for the example of *Kcnq1ot1* and other lncRNAs, such as *Airn* and *H19* (Barlow and Bartolomei, 2014; Schertzer et al., 2019). The large transcriptional burst size of *Kcnq1ot1* and a transcript length of almost 100 kb, which requires an estimated half an hour to complete transcription, likely contribute to the generation of micrometer-large *Kcnq1ot1* RNA clouds (**Figure 1C**). Such tethered RNA sponges efficiently recruit polycomb repressive complexes to silence neighboring genes allele-specifically over megabase distances (Murakami et al., 2007; Mohammad et al., 2008; Redrup et al., 2009; Larsson et al., 2019; Schertzer et al., 2019). RNA has been found as a key determinant for the association of Polycomb repressive complex 2 (PRC2) with chromatin. PRC2 is recruited to chromatin through interactions with nascent RNAs as well as evicted from it upon interaction with G-tracts in RNAs (Beltran et al., 2019; Long et al., 2020).

Xist, on the other hand, provides an example of a lncRNA that covers a larger chromatin territory, a whole X chromosome.

Such a reach is unlikely to be achieved only through co-transcriptional tethering of the RNA to the *Xist* locus by RNA polymerases, and *Xist* interactions with hnRNP U/SAF-A and CIZ1 are believed to contribute to X chromosome association of the RNA beyond the immediate vicinity of the *Xist* transcription site (Hasegawa et al., 2010; Ridings-Figueroa et al., 2017; Sunwoo et al., 2017). Other post-transcriptional or polymerase-independent mechanisms to retain regulatory transcripts on chromatin include the hybridization of RNA to a complementary region in one strand of melted DNA to form R loops, Hoogsteen base-pairing resulting in RNA:DNA triplexes, or tethering of RNA to chromatin by U1 snRNP in a splicing-independent manner (Chédin, 2016; Li et al., 2016; Yin et al., 2020). *Xist*-directed dosage compensation in female mammals depends on the recruitment and eviction of regulatory protein complexes through repeated RNA motifs and structures in *Xist*, which show increasing evolutionary conservation toward the 5' end of the transcript (Brown et al., 1992; Colognori et al., 2020; Strehle and Guttman, 2020). Inactivation of the X chromosome initiates at the site of *Xist* transcription and then extends with *Xist* spreading to proximal chromosomal regions and subsequently to more distal sites according to a "first come first served" principle in three-dimensional space (Engreitz et al., 2013). After initial coating of the chromosome by *Xist*, RNA-binding proteins, while interacting simultaneously with a repeat region in *Xist*, are believed to form condensates to sustain anchoring of *Xist* to the inactivated X territory and X chromosome silencing (Pandya-Jones et al., 2020).

SPATIAL PROXIMITY BETWEEN ENHANCER RNAS AND PROMOTERS

Enhancers are regulatory genomic elements, which modulate the expression of genes in linear and spatial proximity. Some key features of potent enhancers resemble those of active genes: open chromatin, certain shared chromatin modifications and promoter elements, RNA polymerase binding and the synthesis of RNA (Andersson and Sandelin, 2020). Enhancer-derived RNAs (eRNAs) are frequently short-lived, which coincides with their local restriction to corresponding, transcript-encoding enhancer regions. eRNAs have been implicated in the regulation of target genes by enhancers and different studies have shown that eRNA transcription precedes target mRNA transcription, a prerequisite to initiate first steps of target gene transcription (Arner et al., 2015; Kim et al., 2015; de Lara et al., 2019). It is believed that enhancer chromatin regions and associated eRNAs are placed into proximity of target promoters for gene activation, although different studies on the correlation of the time of DNA looping and gene activation reached different conclusions (Lai et al., 2013; Alexander et al., 2019; Benabdallah et al., 2019; Schoenfelder and Fraser, 2019).

Like other transcripts, eRNAs are synthesized discontinuously, and evidence suggests that eRNAs in turn can modulate transcriptional bursting of target genes. Single-cell sequencing analysis of enhancer expression revealed that estimates of burst size in eRNA transcription matches those of genes (Kouno

et al., 2019; Larsson et al., 2019). However, the frequency of bursts is lower for enhancers than for genes, contributing to an overall lower cell population-averaged RNA signal for enhancers than for genes. To gain insights into the regulatory relationship between eRNAs and target RNAs, Rahman et al. determined abundance and co-localization of both upon estrogen signaling and using fluorescence *in situ* hybridization (FISH). Estrogen treatment increases the number of cells, i.e., the burst frequency, that express eRNAs and the corresponding, estrogen-responsive target genes *Foxc1* and *P2ry2*. Co-localization of eRNA and nascent target gene RNA spots increases from less than 5% in unstimulated cells around 5-fold to 25% after estrogen treatment (Rahman et al., 2017). Importantly, the size of RNA-FISH spots of *Foxc1*, a proxy for burst size, was found increased in cells with co-localization of *Foxc1* and eRNA. However, the details of the underlying mechanism remain to be fully resolved.

The transcription factor Yin Yang 1 (YY1) and its interactions with eRNAs serves as one mechanistic example of how regulatory RNAs contribute to target gene regulation. YY1 binds to active enhancers and promoters and forms dimers to stabilize DNA looping (Weintraub et al., 2017). Different regions of the YY1 protein bind to DNA and RNA, respectively, at those regulatory elements. Upon experimentally decreasing the abundance of enhancer RNAs YY1—chromatin interactions are weakened, suggesting its affinity toward RNA assists in targeting YY1 to chromatin. Knock-down of exosome components results in an increased eRNA half-life and as a consequence a larger spatial reach and a less confined localization of eRNAs to chromatin. Indeed, upregulated and diffuse eRNA localization impairs YY1 binding to its chromatin binding sites. Therefore, RNA and DNA binding by YY1 act cooperatively if co-localized but compete when dispersed (Sigova et al., 2015).

CO-TRANSCRIPTIONAL, RNA-ASSISTED FORMATION OF MEMBRANE-LESS STRUCTURES

Similar to the estrogen-induced increase in transcriptional burst size of *Foxc1* upon co-localization with its enhancer and eRNA, genes close to large nuclear speckles containing the non-coding RNA *Malat1* are subject to an increase in burst size and amplify their transcriptional output. After heat shock, *Hsp* transgenes and endogenous genes are induced at the same time irrespective of whether the gene is speckle-associated or not. However, cells with the *Hsp* gene positioned in the vicinity of nuclear speckles surpass cells whose heat-inducible gene is apart from speckles in signal intensity and size of nascent RNA spots. The boosted transcriptional response at speckles correlates with lower exosome activity and larger foci of elongating RNA polymerase II (Khanna et al., 2014; Kim et al., 2020). Furthermore, association of endogenous *Hsp* genes with speckles showed a ripple effect. RNA-FISH of neighboring genes also revealed an increase in the size of transcriptional bursts, but not frequency, when associated with speckles (Kim et al., 2020). Not only are genes more efficiently expressed when in proximity to *Malat1*

speckles, but splicing rates are markedly elevated presumably due to an increase in the availability of splicing machinery (Ding and Elowitz, 2019).

Most examples of pc- and eRNAs, mentioned above, likely exert their regulatory roles before decay or diffusion away from their encoding loci. In contrast, in many cell types dozens or more *Malat1*-containing bodies are observed in individual nuclei, which outnumbers the *Malat1* alleles, suggesting these entities are positioned in nuclear space uncoupled from *Malat1*-encoding genes. However, there is evidence of co-transcriptional protein recruitment and body assembly at the gene locus of the architectural RNA. One strategy to assess the potential role of a RNA in co-transcriptional nuclear body formation comprises the artificial tethering of candidate RNAs to an ectopic genome location and monitoring protein recruitment and body formation. Chosen transcripts are tagged with MS2 and co-expressed in cells with a MS2-binding protein fused to LacI. The cells contain a LacO array as the ectopic genome site to concentrate RNA, mimicking endogenous RNA clusters of bursting transcription sites (Shevtsov and Dundr, 2011). Indeed, MS2-tagged *Malat1* efficiently recruits speckle protein markers, such as splicing factor SRSF1, to LacO arrays and forms nuclear puncta (Tripathi et al., 2012).

The following three features of *Malat1* emphasize its potential to assist in membrane-less body formation (Sanford et al., 2009; Tripathi et al., 2010; Larsson et al., 2019): (i) A *Malat1* RNA molecule contains around 50 potential SRSF1-binding sites to concentrate the splicing factor critical for speckle integrity. (ii) The locations of these sites are biased toward the 5' end of the *Malat1* molecule, the RNA part transcribed first and therefore with the most long-lived chromatin association. (iii) *Malat1* transcription is characterized by one of the largest transcriptional burst sizes (**Figure 1D**; Larsson et al., 2019). The temporally confined, quasi-synchronous emergence of RNA molecules, as opposed to a steady production of few transcripts at any given time point, likely amplifies the function of *Malat1* to act as a sponge during transcription.

Similarly to *Malat1*, non-coding RNA *Neat1* triggers paraspeckle formation co-transcriptionally (Mao et al., 2011; Shevtsov and Dundr, 2011). Increased RNA abundance at a gene locus as means to initiate sequestration of and multivalent interactions between proteins to form membrane-less structures is reminiscent of ribosomal RNA transcription in nucleolus formation (Hernandez-Verdun, 2011).

DISCUSSION

Parameters of single-cell and locus-specific transcription, transcript length, RNA decay and diffusion rates all influence the time of chromatin association for nascent, regulatory transcripts and are possibly of equal importance to the structural and motif content of the RNA in order to regulate proximal genes. We speculate that a large burst size combined with increased protein-binding motif occurrence toward the 5' ends of RNAs is one solution in the parameter space to locally concentrate protein factors with RNA affinity for subsequent,

site-specific regulation of nuclear processes. However, the act of transcription itself, independent of an increased abundance of specific RNA motifs, can affect the expression of proximal transcription units to some extent. Indication of regulatory RNA function with little sequence requirements is suggested by the correlation in expression of neighboring genes (Ebisuya et al., 2008). Furthermore, across rodents, expression levels of protein-coding transcripts co-evolve with expression of neighboring non-coding RNAs. Protein-coding RNAs show a narrower distribution of expression levels from different rodents than non-coding RNAs. Therefore, if expression of a non-coding RNA is gained or lost during evolution, the transcriptional output of neighboring protein-coding genes is accordingly found augmented or pruned (Kutter et al., 2012). Despite correlation in the expression levels of non-coding and coding RNAs, nucleotide substitution rates for non-coding transcripts are much faster in comparison to neighboring protein-coding genes, suggesting the contribution of general, in addition to RNA- and motif-specific, protein factors to the regulatory interplay between proximal transcription sites (Ponjavic et al., 2009; Orom et al., 2010). A mechanistic explanation is perhaps provided by interactions between low-complexity C-terminal domains (CTD) of multiple RNA polymerase II complexes and between the CTD and the transcription preinitiation complex (PIC), both interactions increase transcription efficiency and might take place between different but proximal genes or genes and transcribed super enhancers (Quintero-Cadena et al., 2020).

In a model in which co-transcriptional chromatin decoration with RNA overcomes barriers for transcription initiation or repression of proximal genes, protein-coding transcripts cannot be categorically excluded from regulatory roles commonly assigned to non-coding RNAs. Results from studies applying enhancer screening, followed by CRISPR-Cas9 manipulation, or analysis of gene expression levels associated with sequence variation in regulatory regions revealed protein-coding gene promoters as potent distal regulatory elements (Dao et al., 2017; Mitchelmore et al., 2020). The dual role of promoters

in the regulation of immediate downstream and distal gene expression is consistent with the notion that protein-coding RNAs, immediately downstream of promoters, might as well be involved in the regulatory process of other, proximal genes. Furthermore, protein-coding RNAs can seed larger, membrane-less structures. Using the MS2-tethering approach histone *H2b* RNA was found capable to induce subnuclear structures resembling histone locus bodies and RNA from a β-globin minigene to assemble splicing speckle components into nuclear puncta (Shevtsov and Dundr, 2011). Chromosome conformation techniques and proximity mapping of pairwise or multiple RNAs simultaneously (Lieberman-Aiden et al., 2009; Morf et al., 2019; Cai et al., 2020) in combination with measurements of transcription output and kinetics might be one way to comprehensively identify regulatory relationships between transcripts and genes. Furthermore, recent advances that allow monitoring of genome architecture and transcription at the single-cell level (Nagano et al., 2013; Stevens et al., 2017; Larsson et al., 2019) will provide further insights into how the interplay between genome structure, RNA, and characteristics of its synthesis, regulates proximal gene transcription with high spatial specificity. Altogether, these findings and new approaches progressively uncover a principle of genome physiology in which RNAs not only comprise its primary output, but simultaneously contribute to the regulatory input for genome expression.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

REFERENCES

Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., et al. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* 8:e41769. doi: 10.7554/eLife.41769.054

Amaral, P. P., Leonardi, T., Han, N., Vire, E., Gascoigne, D. K., Arias-Carrasco, R., et al. (2018). Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biol.* 19:32. doi: 10.1186/s13059-018-1405-5

Andersson, R., and Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* 21, 71–87. doi: 10.1038/s41576-019-0173-8

Arner, E., Daub, C. O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., et al. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* 347, 1010–1014. doi: 10.1126/science.1259418

Barlow, D. P., and Bartolomei, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* 6:a018382. doi: 10.1101/cshperspect.a018382

Bell, C. C., Amaral, P. P., Kalsbeek, A., Magor, G. W., Gillinder, K. R., Tangermann, P., et al. (2016). The Evx1/Evx1as gene locus regulates anterior-posterior patterning during gastrulation. *Sci. Rep.* 6:26657. doi: 10.1038/srep26657

Beltran, M., Tavares, M., Justin, N., Khandelwal, G., Ambrose, J., Foster, B. M., et al. (2019). G-tract RNA removes Polycomb repressive complex 2 from genes. *Nat. Struct. Mol. Biol.* 26, 899–909. doi: 10.1038/s41594-019-0293-z

Benabdallah, N. S., Williamson, I., Illingworth, R. S., Kane, L., Boyle, S., Sengupta, D., et al. (2019). Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol. Cell.* 76, 473.e7–484.e7. doi: 10.1016/j.molcel.2019.07.038

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., et al. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* 351, 329–331. doi: 10.1038/351329a0

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., et al. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527–542. doi: 10.1016/0092-8674(92)90520-m

Brown, C. J., Lafreniere, R. G., Powers, V. E., Sebastio, G., Ballabio, A., Pettigrew, A. L., et al. (1991). Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* 349, 82–84. doi: 10.1038/349082a0

Cai, Z., Cao, C., Ji, L., Ye, R., Wang, D., Xia, C., et al. (2020). RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature* 582, 432–437. doi: 10.1038/s41586-020-2249-1

Cassiday, L. A., and Maher, L. J. III (2002). Having it both ways: transcription factors that bind DNA and RNA. *Nucleic Acids Res.* 30, 4118–4126. doi: 10.1093/nar/gkf512

Chédin, F. (2016). Nascent connections: R-loops and chromatin patterning. *Trends Genet.* 32, 828–838. doi: 10.1016/j.tig.2016.10.002

Colognori, D., Sunwoo, H., Wang, D., Wang, C. Y., and Lee, J. T. (2020). Xist repeats A and B account for two distinct phases of X inactivation establishment. *Dev. Cell* 54, 21.e5–32.e5. doi: 10.1016/j.devcel.2020.05.021

Coté, A., Coté, C., Bayatpour, S., Drexler, H. L., Alexander, K. A., Chen, F., et al. (2020). The spatial distributions of pre-mRNAs suggest post-transcriptional splicing of specific introns within endogenous genes. *bioraxiv* [Preprint]. Available online at: https://www.bioraxiv.org/content/10.1101/2020.04.06.028092v1 (Accessed July 21, 2020). doi: 10.1101/2020.04.06.028092

Cramer, P. (2019). Organization and regulation of gene transcription. *Nature* 573, 45–54. doi: 10.1038/s41586-019-1517-4

Dao, L. T. M., Galindo-Albarran, A. O., Castro-Mondragon, J. A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., et al. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* 49, 1073–1081. doi: 10.1038/ng.3884

Dar, R. D., Razooky, B. S., Singh, A., Trimeloni, T. V., McCollum, J. M., Cox, C. D., et al. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17454–17459. doi: 10.1073/pnas.1213530109

de Lara, J. C., Arzate-Mejia, R. G., and Recillas-Targa, F. (2019). Enhancer RNAs: insights into their biological role. *Epigenet. Insights* 12:2516865719846093. doi: 10.1177/2516865719846093

Deng, C., Li, Y., Zhou, L., Cho, J., Patel, B., Terada, N., et al. (2016). HoxBlinc RNA recruits Set1/MLL complexes to activate hox gene expression patterns and mesoderm lineage development. *Cell Rep.* 14, 103–114. doi: 10.1016/j.celrep.2015.12.007

Ding, F., and Elowitz, M. B. (2019). Constitutive splicing and economies of scale in gene expression. *Nat. Struct. Mol. Biol.* 26, 424–432. doi: 10.1038/s41594-019-0226-x

Dinger, M. E., Amaral, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., et al. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18, 1433–1445. doi: 10.1101/gr.078378.108

Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. *Nat. Cell Biol.* 10, 1106–1113. doi: 10.1038/ncb1771

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341, 1237973. doi: 10.1126/science.1237973

Fanucchi, S., Fok, E. T., Dalla, E., Shibayama, Y., Borner, K., Chang, E. Y., et al. (2019). Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat. Genet.* 51, 138–150. doi: 10.1038/s41588-018-0298-2

Gil, N., and Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* 21, 102–117. doi: 10.1038/s41576-019-0184-5

Guil, S., and Esteller, M. (2012). Cis-acting noncoding RNAs: friends and foes. *Nat. Struct. Mol. Biol.* 19, 1068–1075. doi: 10.1038/nsmb.2428

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K., and Nakagawa, S. (2010). The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell* 19, 469–476. doi: 10.1016/j.devcel.2010.08.006

Hernandez-Verdun, D. (2011). Assembly and disassembly of the nucleolus during the cell cycle. *Nucleus* 2, 189–194. doi: 10.4161/nucl.2.3.16246

Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., and Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* 11, 1110–1122. doi: 10.1016/j.celrep.2015.04.023

Hudson, W. H., and Ortlund, E. A. (2014). The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* 15, 749–760. doi: 10.1038/nrm3884

Khanna, N., Hu, Y., and Belmont, A. S. (2014). HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol.* 24, 1138–1144. doi: 10.1016/j.cub.2014.03.053

Kim, J., Venkata, N. C., Hernandez Gonzalez, G. A., Khanna, N., and Belmont, A. S. (2020). Gene expression amplifcation by nuclear speckle association. *J. Cell Biol.* 219:ie201904046.

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., et al. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 1231–1245. doi: 10.1016/j.cell.2006.12.048

Kim, Y. W., Lee, S., Yun, J., and Kim, A. (2015). Chromatin looping and eRNA transcription precede the transcriptional activation of gene in the beta-globin locus. *Biosci. Rep.* 35:e00179.

Kouno, T., Moody, J., Kwon, A. T., Shibayama, Y., Kato, S., Huang, Y., et al. (2019). C1 CAGE detects transcription start sites and enhancer activity at single-cell resolution. *Nat. Commun.* 10:360. doi: 10.1038/s41467-018-08126-5

Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., et al. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* 8:e1002841. doi: 10.1371/journal.pgen.1002841

Lai, F., Orom, U. A., Cesaroni, M., Beringer, M., Taatjes, D. J., Blobel, G. A., et al. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494, 497–501. doi: 10.1038/nature11884

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., et al. (2019). Genomic encoding of transcriptional burst kinetics. *Nature* 565, 251–254. doi: 10.1038/s41586-018-0836-1

Li, J., and Liu, C. (2019). Coding or noncoding, the converging concepts of RNAs. *Front. Genet.* 10:496. doi: 10.3389/fgene.2019.00496

Li, Y., Syed, J., and Sugiyama, H. (2016). RNA-DNA triplex formation by long noncoding RNAs. *Cell. Chem. Biol.* 23, 1325–1333. doi: 10.1016/j.chembiol.2016.09.011

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Long, Y., Hwang, T., Gooding, A. R., Goodrich, K. J., Rinn, J. L., and Cech, T. R. (2020). RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. *Nat. Genet.* 52, 931–938. doi: 10.1038/s41588-020-0662-x

Luo, S., Lu, J. Y., Liu, L., Yin, Y., Chen, C., Han, X., et al. (2016). Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell* 18, 637–652. doi: 10.1016/j.stem.2016.01.024

Mao, Y. S., Sunwoo, H., Zhang, B., and Spector, D. L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* 13, 95–101. doi: 10.1038/ncb2140

McKnight, S. L., and Miller, O. L. Jr. (1979). Post-replicative nonribosomal transcription units in *D. melanogaster* embryos. *Cell* 17, 551–563. doi: 10.1016/0092-8674(79)90263-0

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* 38, D750–D753.

Mitchelmore, J., Grinberg, N. F., Wallace, C., and Spivakov, M. (2020). Functional effects of variation in transcription factor binding highlight long-range gene regulation by epromoters. *Nucleic Acids Res.* 48, 2866–2879. doi: 10.1093/nar/gkaa123

Mohammad, F., Pandey, R. R., Nagano, T., Chakalova, L., Mondal, T., Fraser, P., et al. (2008). Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Mol. Cell. Biol.* 28, 3713–3728. doi: 10.1128/mcb.02263-07

Morf, J., Wingett, S. W., Farabella, I., Cairns, J., Furlan-Magaril, M., Jimenez-Garcia, L. F., et al. (2019). RNA proximity sequencing reveals the spatial organization of the transcriptome in the nucleus. *Nat. Biotechnol.* 37, 793–802. doi: 10.1038/s41587-019-0166-3

Murakami, K., Oshimura, M., and Kugoh, H. (2007). Suggestive evidence for chromosomal localization of non-coding RNA from imprinted LIT1. *J. Hum. Genet.* 52, 926–933. doi: 10.1007/s10038-007-0196-4

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., et al. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64. doi: 10.1038/nature12593

Nicolas, D., Phillips, N. E., and Naef, F. (2017). What shapes eukaryotic transcriptional bursting? *Mol. bioSyst.* 13, 1280–1290. doi: 10.1039/c7mb00154a

Orom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143, 46–58. doi: 10.1016/j.cell.2010.09.001

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W. R., Damianov, A., et al. (2020). A protein assembly mediates Xist localization and gene silencing. *Nature* [Epub ahead of print]. doi: 10.1038/s41586-020-2703-0

Ponjavic, J., Oliver, P. L., Lunter, G., and Ponting, C. P. (2009). Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet.* 5:e1000617. doi: 10.1371/journal.pgen.1000617

Quintero-Cadena, P., Lenstra, T. L., and Sternberg, P. W. (2020). RNA Pol II length and disorder enable cooperative scaling of transcriptional bursting. *Mol. Cell.* 79, 207.e8–220.e8.

Rahman, S., Zorca, C. E., Traboulsi, T., Noutahi, E., Krause, M. R., Mader, S., et al. (2017). Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription. *Nucleic Acids Res.* 45, 3017–3030.

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4:e309. doi: 10.1371/journal.pbio.0040309

Redrup, L., Branco, M. R., Perdeaux, E. R., Krueger, C., Lewis, A., Santos, F., et al. (2009). The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development* 136, 525–530. doi: 10.1242/dev.031328

Ridings-Figueroa, R., Stewart, E. R., Nesterova, T. B., Coker, H., Pintacuda, G., Godwin, J., et al. (2017). The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* 31, 876–888. doi: 10.1101/gad.295907.117

Rodriguez, J., and Larson, D. R. (2020). Transcription in living cells: molecular mechanisms of bursting. *Annu. Rev. Biochem.* 89, 189–212. doi: 10.1146/annurev-biochem-011520-105250

Sanford, J. R., Wang, X., Mort, M., Vanduyn, N., Cooper, D. N., Mooney, S. D., et al. (2009). Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* 19, 381–394. doi: 10.1101/gr.082503.108

Schertzer, M. D., Braceros, K. C. A., Starmer, J., Cherney, R. E., Lee, D. M., Salazar, G., et al. (2019). lncRNA-induced spread of polycomb controlled by genome architecture, RNA Abundance, and CpG Island DNA. *Mol. Cell.* 75, 523.e10–537.e10.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20, 437–455. doi: 10.1038/s41576-019-0128-0

Shevtsov, S. P., and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. *Nat. Cell Biol.* 13, 167–173. doi: 10.1038/ncb2157

Sigova, A. A., Abraham, B. J., Ji, X., Molinie, B., Hannett, N. M., Guo, Y. E., et al. (2015). Transcription factor trapping by RNA in gene regulatory elements. *Science* 350, 978–981. doi: 10.1126/science.aad3346

Stevens, T. J., Lando, D., Basu, S., Atkinson, L. P., Cao, Y., Lee, S. F., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544, 59–64. doi: 10.1038/nature21429

Strehle, M., and Guttman, M. (2020). Xist drives spatial compartmentalization of DNA and protein to orchestrate initiation and maintenance of X inactivation. *Curr. Opin. Cell Biol.* 64, 139–147. doi: 10.1016/j.ceb.2020.04.009

Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y., and Lee, J. T. (2017). Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl. Acad. Sci. U.S.A.* 114, 10654–10659. doi: 10.1073/pnas.1711206114

Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332, 472–474. doi: 10.1126/science.1198817

Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J. M., Robert, M. C., et al. (2016). A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* 7:12248.

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell.* 39, 925–938. doi: 10.1016/j.molcel.2010.08.011

Tripathi, V., Song, D. Y., Zong, X., Shevtsov, S. P., Hearn, S., Fu, X. D., et al. (2012). SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Mol. Biol. Cell.* 23, 3694–3706. doi: 10.1091/mbc.e12-03-0206

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* 17, 601–614. doi: 10.1038/nrg.2016.85

Ulitsky, I., and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46. doi: 10.1016/j.cell.2013.06.020

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120–124. doi: 10.1038/nature09819

Wang, X. Q., and Dostie, J. (2017). Reciprocal regulation of chromatin state and architecture by HOTAIRM1 contributes to temporal collinear HOXA gene activation. *Nucleic Acids Res.* 45, 1091–1104. doi: 10.1093/nar/gkw966

Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. *Cell* 171, 1573.e28–1588.e28. doi: 10.1016/j.cell.2017.11.008

Xiao, R., Chen, J. Y., Liang, Z., Luo, D., Chen, G., Lu, Z. J., et al. (2019). Pervasive chromatin-RNA binding protein interactions enable RNA-based regulation of transcription. *Cell* 178, 107.e18–121.e18. doi: 10.1016/j.cell.2019.06.001

Yin, Y., Lu, J. Y., Zhang, X., Shao, W., Xu, Y., Li, P., et al. (2020). U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* 580, 147–150. doi: 10.1038/s41586-020-2105-3

Zhang, X., Lian, Z., Padden, C., Gerstein, M. B., Rozowsky, J., Snyder, M., et al. (2009). A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* 113, 2526–2534. doi: 10.1182/blood-2008-06-162164