Functional impacts of non-coding RNA processing on enhancer activity and target gene expression

Ntini et al.

Evgenia Ntini[1,2,*] and Annalisa Marsico[1,2,3,*]

[1] Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany

[2] Free University Berlin, 14195, Berlin, Germany

[3] Institute of Computational Biology, Helmholtz Zentrum München, München, Germany

[*] Correspondence to: Evgenia Ntini, E-mail: ntini@molgen.mpg.de; Annalisa Marsico, E-mail: marsico@molgen.mpg.de

Abstract

Tight regulation of gene expression is orchestrated by enhancers. Through recent research advancements, it is becoming clear that enhancers are not solely distal regulatory elements harboring transcription factor binding sites and decorated with specific histone marks, but they rather display signatures of active transcription, showing distinct degrees of transcription unit organization. Thereby, a substantial fraction of enhancers give rise to different species of non-coding RNA transcripts with an unprecedented range of potential functions. In this review, we bring together data from recent studies indicating that non-coding RNA transcription from active enhancers, as well as enhancer-produced long non-coding RNA transcripts, may modulate or define the functional regulatory potential of the cognate enhancer. In addition, we summarize supporting evidence that RNA processing of the enhancer-associated long non-coding RNA transcripts may constitute an additional layer of regulation of enhancer activity, which contributes to the control and final outcome of enhancer-targeted gene expression.

Keywords: transcription, enhancer, RNA processing, co-transcriptional RNA splicing, chromatin, long non-coding RNA (lncRNA).

Introduction

Precise regulation of RNA polymerase II (Pol II) transcription is critical for accurate gene activity. Gene expression is to a large degree regulated by distal regulatory elements called enhancers, which positively drive the expression of their targets in an orientation-independent, but highly temporal- and cell type-specific manner (Shlyueva et al., 2014).

In the last years, several high-throughput and computational prediction methods have been developed to annotate enhancers genome-wide. Many of them rely on the accepted evidence that enhancer activity is reflected genome-wide in accessible chromatin regions flanked by nucleosomes carrying the H3K27Ac modification and

a higher proportion of H3K4me1 over H3K4me3 mark. Widely-used algorithms and established computational pipelines, such as chromHMM (Ernst and Kellis, 2012) (Table 1), can segment the genome in different chromatin states based on combinations of histone marks from ChIP-seq experiments and predict genomic regions corresponding to putative enhancers. Other methods rely on the overlap with transcription factor binding sites, such as p300, to define potential enhancers, or on signatures of nascent bidirectional RNA Pol II transcription (Hoffman et al., 2013; Andersson et al., 2014a; He et al., 2017). These methods are usually accompanied by experimental pipelines to validate the enhancer potential on the predicted associated target gene expression, since only the computational annotation of a region as an enhancer does not ensure that this region enhances gene activity *in vivo*.

Enhancers and gene promoters have been proposed to share similar organization, in terms of transcription initiation and transcription factor binding (Core et al., 2014; Andersson et al., 2015; Kim and Shiekhattar, 2015), and functional properties, such as the pausing of Pol II (Henriques et al., 2018; Tippens et al., 2018). In addition, highly transcribed enhancers can act as weak promoters *in vivo* (Nguyen et al., 2016; van Arensbergen et al., 2017; Mikhaylichenko et al., 2018) and a subset of promoters may intrinsically possess enhancer activity (Li et al., 2012; Arnold et al., 2013; Zabidi et al., 2015; Nguyen et al., 2016; Diao et al., 2017; Mikhaylichenko et al., 2018), suggesting that a dual enhancer-promoter activity has evolved at some regulatory sequence elements. In accordance, recent advancements of high-throughput genomics have set the ground to establish that enhancers are not just distal regulatory elements acting as binding platforms for proteins, but they are also actively transcribed, greatly contributing to the overall pervasive transcription observed in eukaryotic genomes (Berretta and Morillon, 2009; Dinger et al., 2009; Jensen et al., 2013). Large-scale sequencing studies have revealed that ~75 % of the human genome is transcribed, but only ~2 % of this output is protein-coding RNA (Kapranov et al., 2007; Djebali et al., 2012). The rest includes several species of non-coding RNA transcripts showing distinct degrees of stability and turnover, as well as different functions, although the specific roles of some non-coding RNA species are still not fully understood.

In this review, we will summarize data examining the role and functional implications of enhancer-associated non-coding RNA transcription and its products in transcription regulation of target gene expression, with a focus on stable long non-coding RNAs. We will discuss the findings of recent studies providing evidence that transcription-associated processes, including splicing, may have a functional impact on transcriptionally active enhancers producing long non-coding RNAs, shaping the regulatory potential of the cognate enhancer on target gene expression.

Different types of non-coding RNA transcripts encoded in the human genome show distinct degrees of processing and stability

Enhancers display extensive RNA Pol II transcription in a bidirectional mode, giving rise to enhancer-associated non-coding RNA molecules, termed eRNAs, which were first described in high-throughput experiments and large-scale transcriptome analyses (De Santa et al., 2010; Kim et al., 2010; Koch et al., 2011). eRNAs are typically short, unspliced and unstable, but based on their average detected length (>100 nt), sometimes they are also classified into 'long' non-coding RNAs (lncRNAs) (Nojima et al., 2018). By definition, long non-coding RNAs (lncRNAs) are molecules >100–200 nt, which lack coding potential (Jensen et al., 2013).

In the current human Gencode annotation (Derrien et al., 2012), lncRNAs consist of ten subclasses, including 'lincRNAs', 'sense overlapping', 'antisense' and 'sense intronic' transcripts. The last three subclasses refer to lncRNAs overlapping other annotated transcription units in the same (sense) or antisense orientation, or found within introns of protein coding genes, respectively. lincRNAs refer to 'intergenic' lncRNAs, meaning not overlapping any other annotated transcription unit. lincRNAs are usually spliced and polyadenylated (Ulitsky and Bartel, 2013) and thus rather stable. In fact, 84.5% of Gencode lincRNAs have at least one splicing junction.

lncRNAs can be identified starting by *de novo* transcript assembly in RNA-sequencing data using specialized algorithms like Cufflinks (Trapnell et al., 2010) and StringTie (Pertea et al., 2015). These tools also allow the identification of splicing junctions and putative alternative splicing isoforms. The assembled transcription unit can be supported by available CAGE data (Shiraki et al., 2003) (Table 1) indicating transcription initiation site. The coding potential of the assembled transcript is then assessed by tools like CPAT (Wang et al., 2013) and PhyloCSF (Lin et al., 2011), and transcripts with a protein-coding score above a given cutoff are filtered out. Intergenic lncRNAs (lincRNAs) are further identified by excluding overlaps with known protein-coding genes, while histone modification marks, like the promoter-associated H3K4me3, may also be employed to support the identification. For instance, for the automated annotation of lincRNAs included in the final Gencode datasets, Ensemble follows an established pipeline (Guttman et al., 2009) where firstly, regions of chromatin methylation (H3K4me3 and H3K36me3) outside known protein-coding loci are identified. Next, cDNAs that overlap with H3K4me3 or H3K36me3 features are identified as candidate lincRNAs. In the final evaluation step, candidate lincRNAs with substantial protein-coding potential (an open reading frame covering at least 35% of length and containing PFAM protein domains) are discarded.

eRNA as a term is not included in Gencode annotation; however, this does not preclude that many of the Gencode annotated lncRNA transcripts may actually overlap enhancer-transcribed eRNAs in specific cell lines. Because of their rapid turnover by nuclear decay pathways employing the exosome activity (Lubas et al., 2015), eRNAs are not readily detectable in steady-state RNA-seq data, and therefore for their annotation nascent RNA-sequencing techniques like GRO-seq (Table 1) or conditions of exosome depletion are employed (Core et al., 2008, 2014; Hah et al., 2011, 2013; Andersson et al., 2014a, b).

On the other hand, using actinomycin treatment and expression microarrays, Clark et al. (2012) showed that lncRNAs as a class cannot be characterized as unstable, since their median half-life (~3.5 h) is comparable to mRNAs (~5.1 h). They also showed that intergenic (lincRNAs) and antisense lncRNAs are more stable compared to intronic lncRNAs, and that spliced lncRNAs are more stable compared to mono-exonic transcripts, demonstrating that splicing as a process contributes to overall transcript stability (Hicks et al., 2006). Recently, Mele et al. (2017) showed that Gencode annotated mRNAs and lincRNAs with similar expression levels exhibit equal stability (half-lives), although lincRNAs as a class are overall less efficiently spliced compared to mRNAs. Yet, the authors do find a substantial fraction of lincRNAs showing efficient splicing and conserved splice junctions, which are enriched in functionally annotated lincRNAs (Amaral et al., 2011), like *FIRRE* (Hacisuleyman et al., 2014), *Xist* (Cerase et al., 2015), and *MIAT* (Liao et al., 2016). This might imply that efficient splicing and the process of maturation is important for a subset of lncRNAs in order to exert their functional roles. This possibility may apply not only to lincRNAs, but also to other lncRNA subclasses as well. For instance, by inhibiting the core spliceosome component PRP8, Marquardt et al. (2014) showed that splicing

of the *COOLAIR* lncRNA, overlapping and transcribed antisense to the *FLC* protein-coding gene, is functionally important for sense gene transcription.

On the other extreme, mono-exonic or inefficiently spliced lncRNAs are more likely to be byproducts of the act of transcription having itself a regulatory role *in cis*. In support of this, for one of the first identified inefficiently spliced lncRNAs, *Airn*, overlapping the *Igf2r* gene in the antisense direction, it was experimentally validated that the act of transcription exerts a regulatory effect *in cis*, whereas the RNA product has no important function (Latos et al., 2012). Adding to this, splicing may not be functionally necessary for those lncRNAs enriched in the chromatin fraction of the cell; lncRNAs tethered to chromatin may exert their function *in cis*, regulating gene expression through the act of their transcription and via mechanisms like transcriptional interference (Stojic et al., 2016) or by enhancing the binding of transcription factors to regulatory elements (Sigova et al., 2015).

Towards understanding the functional distinctions among different non-coding RNA transcripts produced at enhancers

Apart from the relatively short, unstable and unspliced bidirectional eRNAs (De Santa et al., 2010; Kim et al., 2010; Djebali et al., 2012), a subset of enhancers transcribe unidirectional lncRNAs showing different characteristics in terms of splicing and stability (Koch et al., 2011; Marques et al., 2013; Hon et al., 2017; Gil and Ulitsky, 2018; Tan et al., https://doi.org/10.1101/287706). In a previous study, using solely the H3K4me1 histone modification mark and the correlation of the H3K4me1 signal with cell type-specific expression of putative mRNA targets to predict enhancers (Corradin et al., 2014), it was estimated that about one third of Gencode annotated lincRNAs overlap cell type-specific enhancers (Vucicevic et al., 2015). The Marques lab refers to these intergenic enhancer-associated lncRNAs as 'elincRNAs' (Marques et al., 2013; Tan et al., https://doi.org/10.1101/287706). About half of the Marques elincRNAs (44%) are spliced, i.e. having at least one splice junction and more than one exon. Given the complexity of the transcriptional landscape at enhancers and that relatively little is known about the functional impact of the production of long, stable non-coding RNA transcripts on the enhancer activity, two recent studies have set out to unveil the functional distinctions of the different classes of lncRNA transcripts produced from active enhancers, using genome-wide computational approaches (Gil and Ulitsky, 2018; Tan et al., https://doi.org/10.1101/287706).

Bidirectional Pol II transcription producing eRNAs is tightly associated with enhancer activity and is therefore considered a hallmark of active enhancers. Hence, it is increasingly being employed in enhancer annotation, usually in combination with signal of DNase I hypersensitivity which reflects elevated chromatin accessibility due to a more open and active chromatin environment (Melgar et al., 2011; Andersson et al., 2014a; Nagari et al., 2017). Nascent RNA sequencing techniques, like GRO-seq (Core et al., 2008) (Table 1) map nascent RNA molecules produced from transcriptionally engaged RNA Pol II, and can therefore be used to detect the short-lived eRNAs (not captured in steady-state RNA-seq data). Using GRO-seq data, Gil and Ulitsky (2018) annotated enhancers as regions displaying substantial bidirectional transcription around DNase I hypersensitivity sites (DHSs), discarding any overlaps with protein coding genes (PCGs). These enhancers, termed eRNA-producing centers (EPCs), were validated to display canonical enhancer histone marks (high H3K4me1 and H3K27Ac signal) and were annotated in three ENCODE human cell lines and in mouse

embryonic stem cells (mESC). The authors computed the distance between each EPC and the closest Gencode lncRNA transcription start site (TSS), and found that ~5 % of all EPCs have an lncRNA TSS within a distance <0.8 kb. This implies that intrinsically bidirectional transcription originating at active enhancers (Andersson et al., 2014a, b) is sometimes preferentially elongated and strengthened in one direction to produce a stable lncRNA transcript, perhaps by the unidirectional accumulation of random mutations that bring together transcription factor binding sites, splice sites and poly(A) sites. They termed these lncRNA-associated enhancers 'la-EPCs', and the rest 'na-EPCs'; the latter correspond to non-lncRNA-associated enhancers, producing solely bidirectional eRNAs, and constitute the majority of active enhancers. Notably, >90 % of the enhancer-associated lncRNAs are spliced. Essentially, all of them could be classified as intergenic (lincRNAs) per definition, as any overlap with PCGs was excluded in the first place in their analysis.

A different approach is utilized by Tan et al. (https://doi.org/10.1101/287706) to define a set of actively transcribed intergenic enhancers in mESCs, based on ENCODE chromHMM-annotated enhancers in the same cell line (Bogu et al., 2015) overlapping mESC-specific DHSs (Mouse et al., 2012) and a CAGE cluster indicating transcription initiation (Fraser et al., 2015). In agreement with Gil and Ulitsky (2018), they found that ~5 % of the analyzed mESC transcribed enhancers overlap previously annotated lincRNAs expressed in mESCs (Tan et al., 2015). They termed these intergenic enhancer-associated lncRNAs 'elincRNAs' to distinguish them from other mESC-expressed intergenic lncRNAs not overlapping enhancers. Because of the different initial methodologies employed to define the final working datasets of transcribed enhancers, the fraction of spliced (multi-exonic) enhancer-associated lncRNAs differs between the two studies: about half in Tan et al. (https://doi.org/10.1101/287706) versus > 90% in Gil and Ulitsky (2018). A schematic representation of the computational methodologies employed for the genome-wide annotation of transcribed enhancers is summarized in Figure 1.

Both groups then set out to characterize the properties and the functional distinctions between the different classes of enhancers and non-coding RNA transcripts produced at enhancer loci. In particular, Gil and Ulitsky (2018) conducted a detailed analysis to distinguish between bidirectionally transcribed enhancers associated with lncRNA production (la-EPCs) and the rest majority of bidirectionally transcribed enhancers not associated with lncRNAs (na-EPCs). Similarly, Tan et al. (https://doi.org/10.1101/287706) focused on characterizing the functional impact of transcribed elincRNAs on their cognate enhancer activity. They also compared the functional properties of elincRNAs to other expressed lincRNAs not associated with a transcribed enhancer.

A first interesting finding in Gil and Ulitsky (2018) is that lncRNA-associated enhancers (la-EPCs) exhibit significantly higher signals of marks associated with enhancer activity, compared to the rest of bidirectionally transcribed enhancers (na-EPCs). These are: increased transcription activity (GRO-seq), higher H3K27Ac—a hallmark of transcriptionally active enhancers, and higher ATAC-seq signal (Buenrostro et al., 2013) (Table 1) reflecting local chromatin accessibility. This result suggests that the production of lncRNAs may have a positive impact on enhancer activity. Intriguingly, this effect may be intrinsically encoded in the DNA sequences of the lncRNA-associated enhancers, as these also display higher STARR-seq signal (Arnold et al., 2013; Muerdter et al., 2015) (Table 1) compared to the rest bidirectionally transcribed enhancers. Notably, lncRNA-associated enhancers are preferentially found at the anchor points of chromosomal loops, the latter identified by Pol II ChIA-PET (Fullwood et al., 2009a, b; Fullwood and Ruan, 2009) (Table 1), suggesting that these

lncRNA-transcribing enhancers are more likely to form three-dimensional contacts with distal loci encompassing putative target genes.

Related to this, it was previously reported that some disease/trait-associated lncRNAs showing enrichment in enhancer-like chromatin signatures (elevated H3K4me1 to H3K4me3 ratio) are enriched at the boundaries of TADs (Tan et al., 2017) (Table 1). It was further suggested that these lncRNAs might promote intra-TAD chromosomal interactions, since their abundance (expression or RNA levels) correlates with the frequency of intra-TAD DNA–DNA contacts, the latter computed from Hi-C contact matrices (Lieberman-Aiden et al., 2009) (Table 1). Interestingly, TAD boundaries are the sites where chromosomal looping between enhancers and target gene promoters frequently occurs (Symmons et al., 2014; Lupianez et al., 2015). In their following study, Tan et al. (https://doi.org/10.1101/287706) unveiled that this preferential location of enhancer-associated lncRNAs at TAD boundaries is strictly restricted to spliced lncRNAs. In particular, they found that transcription initiation regions of multi-exonic (spliced) enhancer-associated lncRNAs are significantly enriched at the anchors of chromosomal loops, when compared to all other enhancer-derived non-coding RNA transcripts (i.e. eRNAs and mono-exonic enhancer-associated lncRNAs) or to other lincRNAs expressed in mESC. The authors therefore suggested that multi-exonic, spliced enhancer-associated lncRNAs might be involved in shaping (or modulating) the local chromosomal architecture (Tan et al., https://doi.org/10.1101/287706).

In support of a linkage between potentially functional lncRNAs and higher order chromatin organization, Amaral et al. (2018) found that the majority of a set of positionally conserved spliced lncRNAs, which are preserved in their genomic location relative to orthologous protein-coding genes between mouse and human, are located at loop anchor points and TAD boundaries, showing enrichment of CTCF binding sites. In the same line, we recently identified in breast cancer cells a subset of spliced lncRNAs transcribed from active enhancers at the anchors of chromosomal loops to putative target genes. These enhancer-associated lncRNAs show significant correlation in expression with their putative targets, stressing their role in cognate enhancer activity. Intriguingly, this group of enhancer-associated lncRNAs are less enriched in the chromatin-associated RNA fraction, indicating that these molecules are less tethered to chromatin, and that their chromatin-dissociation may be important for potential downstream functions (Ntini et al., 2018 and discussed in the last section).

Gil and Ulitsky (2018) also found enrichment of CTCF binding at the lncRNA-associated enhancers (la-EPCs), which is considered a factor involved in the maintenance of chromosomal loops (Weintraub et al., 2017), as well as differential binding of several proteins preferentially found at PCG promoters. In addition, a higher fraction of the lncRNA-associated enhancers bear the H3K4me3 promoter histone mark (compared to the rest bidirectionally transcribed enhancers), overall underscoring that lncRNA-producing enhancers harbor promoter characteristics (Gil and Ulitsky, 2018). This is in agreement with previous studies reporting that lncRNA transcription initiation regions show both promoter and enhancer characteristics (Marques et al., 2013; Andersson et al., 2015). Notably, the promoter regions of the enhancer-associated lncRNAs show significant enrichment for DNA binding proteins which bind also RNA (i.e. containing an RNA recognition motif, RRM), whereas they are depleted from chromatin-remodeling forkhead-domain proteins acting as pioneer transcription factors (Cirillo et al., 2002; Lalmansingh et al., 2012; Soufi et al., 2015). Based on that, the authors suggested that the lncRNA-producing enhancers may employ specific mechanisms for opening the local chromatin structure and initiate transcription; thereby cotranscriptional splicing of the associated nascent lncRNA transcript may play a role in reinforcing this process, as discussed in the last section of this review.

Splicing has a positive impact on enhancer activity

A prominent distinguishing characteristic between the lncRNA-associated enhancers (la-EPCs) and the rest of the enhancers transcribing bidirectional transient eRNAs (na-EPCs), highlighted in the Gil and Ulitsky (2018) study, is splicing. In fact, among the DNA binding proteins with an RNA recognition motif enriched at the enhancer-associated lncRNA promoters several are RNA processing factors, including essential proteins involved in splicing, such as RBFOX2, SRSF1, and U2AF1. As a proxy for splicing activity, the authors used the exon density, defined as the number of exons of a transcript normalized to the locus length. They found a correlation between the exon density of the enhancer-transcribed lncRNAs and the density of DHSs and H3K27Ac signal at the same locus. This suggests that splicing is positively associated with the enhancer activity. Notably, the consequent effect of splicing on increasing chromatin accessibility is exerted on the broader enhancer region and not only on the lncRNA-body transcribed part. Based on that, the authors suggested that the observed splicing-coupled stimulating effect on enhancer activity is more likely to be due to the recruitment of activating factors—either by the produced lncRNA transcript itself or by its transcription and associated processing—rather than by some local effect of Pol II elongation along the lncRNA body (Gil and Ulitsky, 2018).

Similarly, Tan et al. (https://doi.org/10.1101/287706) also observe a positive impact of enhancer-associated lncRNA splicing on cognate enhancer activity. In particular, using CAGE data to measure transcription initiation during different stages of embryonic neurogenesis, they found that changes in enhancer-associated lncRNA transcription positively correlated with changes in transcription of the closest PCG, the latter considered as a putative *cis* target. This correlation was significantly stronger for the spliced (multi-exonic) enhancer-associated lncRNAs, compared to their mono-exonic counterparts. In agreement, there seems to be a positive correlation between the number of exons of the enhancer-associated lncRNAs and the expression changes of their putative *cis*-target genes, implying that the amount of splicing of the enhancer-associated lncRNAs may contribute to their *cis*-regulatory roles, impacting the regulatory potential of their cognate enhancer. In addition, in agreement with the findings by Gil and Ulitsky (2018), Tan et al. (https://doi.org/10.1101/287706) also find enrichment of H3K4me1, DHSs and H3K27Ac signals in enhancers transcribing multi-exonic lncRNAs versus mono-exonic ones, thereby strengthening the notion that splicing of lncRNAs contributes to enhancer activity.

In support of this, a significant enrichment of evolutionary conserved U1 splicing sites is found at the lncRNA-associated enhancers and only in the direction downstream of the lncRNA TSS, similarly to what is observed downstream of a PCG TSS (Gil and Ulitsky, 2018). Overall, this suggests that evolutionary selection acts on splicing signals to promote the formation of spliced lncRNA transcripts at some enhancers (~5 % of transcribed enhancers). In agreement, Tan et al. (https://doi.org/10.1101/287706) find that mutations are selectively suppressed at splice sites of enhancer-associated lncRNAs, and that their splice-site flanking regions are enriched in conserved exonic splicing enhancers and U1 snRNP recognition motifs, when compared to other intergenic lncRNAs. In general, there is a significant difference in the GC content between exonic and intronic sequences of all multi-exonic lincRNAs (either transcribed from enhancers or not) (Schuler et al., 2014; Haerty and Ponting, 2015; Tan et al., https://doi.org/10.1101/287706), a feature observed also for mRNAs and implicated in promoting splice site recognition and splicing efficiency (Amit et al., 2012).

The notion that conserved splice sites guide the processing of lincRNAs may seem on a first sight to come in contrast with previous studies reporting that lincRNAs as a class are overall less efficiently spliced compared to mRNAs (Mele et al., 2017; Mukherjee et al., 2017). However, this does not apply to the enhancer-associated lincRNAs (elincRNAs), which show splicing efficiency comparable to mRNAs (Tan et al., 2018). Finally, using a statistical genetics approach, Tan et al. (https://doi.org/10.1101/287706) analyzed the effect of nucleotide variants on elincRNA transcription/splicing, cognate enhancer and putative target gene activity. They demonstrated that the decrease of elincRNA splicing due to mutations at splice junctions resulted in 90 % of the cases in significant downregulation of their putative protein-coding target gene expression. These inferred causal relationships further validate the hypothesis that it is the RNA splicing important in strengthening enhancer function, rather than the enhancer activity being a cause for splicing (Tan et al., https://doi.org/10.1101/287706).

Experimentally-derived data support a role of lncRNA splicing in regulation of gene expression and shaping enhancer functionality

Apart from the genome-wide approaches discussed in the previous sections (Gil and Ulitsky, 2018; Tan et al., https://doi.org/10.1101/287706), previous studies employing experimental strategies have provided significant support for the potential functional impact of lncRNA splicing in regulation of target gene expression and on enhancer activity.

Marquardt et al. (2014) analyzed the effect of antisense lncRNA splicing in the plant *Arabidopsis* flowering system: the *COOLAIR* lncRNA is transcribed antisense and overlapping the *FLC* gene, displaying a major promoter-proximal and an alternative distal polyadenylation site (the latter overlaps the sense *FLC* promoter). Analysis of a hypomorphic mutation in the core spliceosome component PRP8 (*prp8-6*) revealed elevated histone methylation in the *FLC* gene body (active histone mark H3K4me2) and upregulated *FLC* transcription. Interestingly, the *prp8-6* mutation did not affect the sense splicing events of *FLC* or two other control genes, but greatly inhibited the splicing of the unique short intron of the major (most abundant) antisense *COOLAIR* isoform. This splicing event is necessary to create the exon, which contains the proximal poly(A) site, hence in the *prp8-6* mutation, usage of the proximal poly(A) site is reduced, while usage of the distal alternative poly(A) site is relatively increased. These results could be recapitulated by mutating the proximal acceptor splice site of *COOLAIR*, which again led to reduced proximal poly(A) site usage, increased H3K4me2 and sense *FLC* transcriptional upregulation, confirming the role of antisense lncRNA splicing in sense gene transcriptional suppression. In particular, the essential splicing event is coupled with targeted 3' end processing of the antisense *COOLAIR* proximal poly(A) site, which triggers histone demethylase activity at the locus in an as yet mechanistically unclear step (Liu et al., 2010). Based on the findings by Liu et al. (2010) and Marquardt et al. (2014), the authors suggested that alternative splicing and targeted 3' end processing of antisense lncRNA transcripts may comprise a common mechanism in regulation of gene expression, by modulating the overlapping sense gene transcription through cotranscriptional coupling processes associated with repressive chromatin modifications. Functional investigation of other antisense lncRNAs in different systems is expected to further support this notion.

Through experimental dissection it is also possible to address the important question whether it is solely the act of transcription and/or cotranscriptional splicing, or rather the mature lncRNA product that constitutes the important functional aspect of enhancer activity, in the case of enhancer-associated lncRNAs. These two possible functional components are not mutually exclusive. The same question is also urging to be resolved in the case of antisense lncRNAs which may be involved in regulation of gene expression *in cis* through transcriptional interference. Stojic et al. (2016) uncoupled the act of transcription from the function of the lncRNA transcript by using siRNAs targeting different regions of the *GNG12-AS1* lncRNA. The authors demonstrated that transfection of siRNAs in human cell lines targeting the first exon of *GNG12-AS1*, near its transcription start site, causes downregulation of the lncRNA transcription, which is accompanied by a consequent transcriptional upregulation of the sense overlapping *DIRAS3* gene. The latter was profiled by an increase in Pol II binding and the active histone modifications H3K4me3 and H3K36me3. In contract, targeting other exons of *GNG12-AS1* (closer to its 3' end) with siRNAs led only to depletion of the lncRNA isoforms through post-transcriptional gene silencing, without affecting its nascent transcription, and importantly had no effect on the overlapping sense *DIRAS3* transcription. Hence, the authors established that it is the act of transcription of the antisense lncRNA that causes transcriptional interference of the sense overlapping target gene (Stojic et al., 2016).

In a similar direction, aiming to distinguish between the *cis* and *trans* implicated functions of transcribed lncRNAs, Engreitz et al. (2016a) employed genetic manipulation of twelve lncRNA loci, using CRISPR technology (Cong et al., 2013; Mali et al., 2013), to perform heterozygous knockouts in *129/castaneus* F1 hybrid mESC. This line contains a polymorphic site every ~140 bp, which allowed to distinguish between the two alleles and classify the lncRNA effects as *cis* or *trans*, based on whether the observed associated expression changes of nearby genes (within 1Mb of the lncRNA locus) are on the same modified (*cis*) or on the unmodified (*trans*) homologous chromosome allele. When the observed target expression changes are restricted to the *cis* allele, this most probably reflects a direct local *cis*-regulatory effect from the lncRNA locus; on the other hand, when the observed expression changes involve both alleles, this most probably means that the lncRNA transcript itself (dissociated from the *cis* modified allele) exerts some downstream *trans* regulatory role(s). Depletion of five lncRNA promoters specifically affected the expression of nearby genes solely on the same allele (*in cis*); intriguingly, that was also the case by depleting six PCG promoters, suggesting that both non-coding and coding transcription units may contribute to shaping neighboring gene expression. This is well related to the general correlation observed in neighboring gene expression (Purmann et al., 2007; Ebisuya et al., 2008). The observed perturbation of *cis* regulation upon promoter depletion could be due to abolishment of essential promoter-associated DNA regulatory elements, due to impeding the process of transcription, or due to abolishment of the RNA transcript itself. To distinguish among these three possibilities, the authors introduced polyadenylation signals (PAS) 0.5–3 kb downstream of the TSS to terminate transcription early and abolish most of the RNA transcript, without affecting the promoter. Interestingly, insertion of the early PAS turned down completely transcription at the whole lncRNA locus, and consequently abolished the lncRNA product; most probably, an early PAS inserted within the first intron prevents splicing, which in turn substantially reduces transcription. Importantly, despite abrogating both transcription and the lncRNA transcript product itself, insertion of the early PAS did not have an effect on the *cis* target gene, suggesting that essential DNA regulatory elements in the depleted lncRNA promoter-proximal region (~750 bp) exert the *cis* regulatory effect on the

nearby gene. These broad lncRNA promoter regions did not show an enhancer-like H3K4me1 to H3K4me3 ratio, hence they would not be classified as enhancers based solely on this measure (Heintzman et al., 2007; Koch and Andrau, 2011; Djebali et al., 2012; Calo and Wysocka, 2013). Still, they seem to possess enhancer activity since they are regulating neighboring gene expression *in cis*, perhaps by containing binding sites for proteins recruiting histone modifiers and chromatin remodelers, which results in increased chromatin accessibility. This observation is also in line with additional reports supporting that promoters and enhancers share organizational and functional characteristics (Li et al., 2012; Marques et al., 2013; Andersson et al., 2015; Kim and Shiekhattar, 2015; Sahlen et al., 2015), and that promoters may function as enhancers (Li et al., 2012; Sanyal et al., 2012; Arnold et al., 2013, 2017; Zabidi et al., 2015; Nguyen et al., 2016; Paralkar et al., 2016; Rajagopal et al., 2016; Dao et al., 2017; Diao et al., 2017; Groff et al., 2018; Mikhaylichenko et al., 2018) and *vice versa* (Kowalczyk et al., 2012; Nguyen et al., 2016; van Arensbergen et al., 2017; Mikhaylichenko et al., 2018).

In order to assess the effect of splicing in shaping the lncRNA functional potential, Engreitz et al. (2016a) deleted the first 5' splice site of the lncRNA *Blustr* using CRISPR in the same system. The functional coupling between splicing and transcription was well established in previous studies, demonstrating that promoter-proximal splice sites and the process of splicing can significantly enhance transcription (Brinster et al., 1988; Fong and Zhou, 2001; Le Hir et al., 2003). Components of the spliceosome can directly enhance Pol II transcription initiation (Kwek et al., 2002) and elongation (Fong and Zhou, 2001), mechanistically explaining why and how splicing enhances gene expression (Schuler et al., 2014). In accordance, by depleting the first 5' splice site of *Blustr*, Engreitz et al. (2016a) reported a total reduction of the lncRNA transcription (using GRO-seq) and a consequent downregulation in neighboring gene expression (*Sfmbt2*). This demonstrates that the process of transcription at the lncRNA locus coupled with cotranscriptional splicing is functionally important for transcription regulation of the target gene. Since the process of splicing depends on direct interactions between the spliceosome and the nascent RNA transcript, the nascent lncRNA itself is required for target gene activation. However, this mechanism does not seem to involve specific primary sequences of the lncRNA, as progressive depletion of individual exons with CRISPR had no significant effect on target gene expression.

In a previous study, removing the splicing signal from the lncRNA *Haunt* by replacing the endogenous lncRNA locus with its cDNA sequence could not rescue its *cis*-regulatory function (Yin et al., 2015). However, in this case, the possibility that essential DNA *cis*-regulatory elements in the omitted intronic sequences are depleted from the genetically modified lncRNA locus cannot be excluded. Therefore, CRISPR genetic manipulation employing depletion of genomic intervals (like promoter regions, exonic or intronic sequences) in independent experiments, complemented by carefully designed short deletions around individual splice sites and isolated sequence motifs, allows for a more thorough dissection and conclusive examination of an lncRNA locus of interest. Consequently, this enables to evaluate the functional importance of individual lncRNA elements and characterize mechanistic aspects of transcription regulation of target gene expression.

Possible mechanisms underlying the functional impact of lncRNA splicing on cognate enhancer and target gene activity

Despite the genome-wide results and the experimental supportive data summarized in this review, which strongly implicate a role of enhancer-transcribed lncRNA splicing in shaping enhancer activity, the molecular mechanism(s) and underlying mechanistic details of this process are unknown. So far, some possible implicated scenarios exist that we will discuss here.

First of all, the lack of primary sequence conservation in the exonic lncRNA sequences of the enhancer-associated lncRNAs (Marques et al., 2013; Tan et al., https://doi.org/10.1101/287706), or even the lack of conservation in their specific exon-intron structure (Ulitsky, 2016), may on a first sight shadow the evidence for a functional role of the lncRNA transcript itself in enhancer activity and *cis* regulation of target gene expression. The latter is established through detailed examination and experimental dissection of individual lncRNA loci (Nagano et al., 2008; Wang et al., 2008, 2009; Khalil et al., 2009; Orom et al., 2010; Cabianca et al., 2012; Guil and Esteller, 2012; Engreitz et al., 2016a; Ntini et al., 2018). In the case of functionally characterized lncRNAs with virtually conserved functions such as *Xist*, a minimum primary sequence conservation is observed (Nesterova et al., 2001; Kirk et al., 2018), although the overall lncRNA gene structure may be well conserved (Brockdorff, 2002; Hoki et al., 2009; Senner et al., 2011). Similarly, the sequence motifs associated with splicing in enhancer-associated lncRNAs, including splice sites and exonic splicing enhancers, are evidently conserved (Ponjavic et al., 2007; Schuler et al., 2014; Haerty and Ponting, 2015; Nitsche et al., 2015; Mele et al., 2017; Tan et al., https://doi.org/10.1101/287706), supporting that splicing is an important functional aspect of enhancer-associated lncRNAs. But how is this functional impact mechanistically exerted?

One possible scenario is that splice factors binding to the conserved lncRNA splice sites recruit in turn protein factors which remodel the chromatin structure in the broader locus, for instance through histone modifications (Kim et al., 2011) (Figure 2A). Data from Schüler et al. (2014) suggest a coupling mechanism between splicing and effected changes in chromatin organization. In this model, splicing affects chromatin through the recruitment of splice-coupled activating factors, such as the CHD1 and SWI/SNF chromatin remodelers, which in turn may modulate neighboring gene activity. A second possibility is that cotranscriptional splicing of the nascent lncRNA transcript has an indirect effect on the transcription activity of the locus, through a functional coupling between splicing and transcription, as described above (Brinster et al., 1988; Fong and Zhou, 2001; Kwek et al., 2002; Le Hir et al., 2003; Schuler et al., 2014). This could involve interactions between the spliceosome and components of the transcriptional machinery (McCracken et al., 1997). Increased transcription and the process of transcription itself may then contribute to *cis* gene regulation by recruiting activating factors like histone modifiers and nucleosome remodelers, resulting in an overall chromatin opening of the locus and increased chromatin accessibility at *cis* target genes (Ebisuya et al., 2008). In support of this, impeding either transcription or splicing at the lncRNA gene via genetic manipulation led to reduction in the H3K4me3 promoter histone mark and a spreading of the repressive mark H3K27me3 at the neighboring gene locus (Engreitz et al., 2016a).

Co-transcriptional splicing as well as the 3' end formation (cleavage and polyadenylation) are processes tightly coupled to transcription and required for the efficient dissociation of the nascent RNA transcript from the chromatin-associated site of transcription (Perales and Bentley, 2009; Rigo and Martinson, 2009; Pandya-Jones et al., 2013; Proudfoot, 2016). Using splicing inhibiting morpholinos we recently showed that impeding the co-transcriptional splicing of an enhancer-associated lncRNA (*A-ROD*) leads to increased chromatin retention

of the lncRNA transcript, and has a consequent repressive effect on transcription of the target gene *DKK1* (Ntini et al., 2018). Therefore, splicing of the enhancer-associated lncRNA may also have some downstream effects mediated by the lncRNA transcript itself. Upon its dissociation from chromatin, the nascent lncRNA may interact with protein factors required and recruited for transcription regulation of target gene expression within the spatial three-dimensional proximity of pre-established chromosomal loops (Ntini et al., 2018). The enhancer-transcribed lncRNA, upon its chromatin dissociation permitted by splicing, may also be involved in- and contribute to 'spatial amplification' of the *cis*-regulatory signal (Engreitz et al., 2016b). For instance, the chromatin-dissociated lncRNA could be bound by RNA recognition motif (RRM)-containing proteins, which bind also to DNA, and thereby get recruited to the promoters of genes within the spatial proximity of the enhancer (Figure 2B). Consequently, this would contribute to a quasi-*cis* mode of gene regulation (Ntini et al., 2018). In this model, the overall primary sequence conservation of the lncRNA is not a necessity, as RRMs can be short and degenerate or can even be emerging in secondary structures formed locally in the nascent chromatin-dissociated lncRNA. These hypotheses are intriguing but need to be further verified by large-scale or more specific analyses, such as bioinformatics prediction of local secondary structure elements in enhancer-associated lncRNAs (Gawronski et al., 2018), or by enrichment analysis of DNA/RNA-binding protein sites from the numerous available high-throughput data in public repositories.

Overall, the recent experimental and genome-wide derived results suggest that the feature of chromatin-dissociation may be an additional functional aspect of enhancer-produced lncRNAs, and therefore understanding the dynamics, underlying molecular mechanism(s), and mechanistic details of this process may help to further characterize the functional distinctions among the various non-coding RNAs transcribed from active enhancers. For instance, whether there is an overall association between the degree of chromatin-dissociation and the degree of splicing, and whether efficiently chromatin-dissociated enhancer-transcribed lncRNAs are more efficiently spliced compared to their chromatin-retained counterparts, are important questions to address.

An overview of the mechanistic scenarios involved in mediating the functional impact of lncRNA splicing on enhancer activity is presented in Figure 2.

In addition, apart from splicing, other cotranscriptional processes associated/coupled with lncRNA transcription, like transcription termination and the formation of a mature 3' end, might as well have a similar functional impact in shaping enhancer activity and contributing to *cis* gene regulation. In this direction, similarly to analyzing DNA motifs and the enrichment of transcription factor binding sites at the promoters of enhancer-associated lncRNAs (Gil and Ulitsky, 2018), characterizing the interactions of RNA-binding proteins with the nascent lncRNA transcripts and their functional impact in co-transcriptional RNA processes is urging. Therefore, in order to comprehensively design and pursue large-scale and long-term experimental analyses, the use of recently developed computational algorithms to predict RNA-binding protein interactions, and their specific application for predictions using RNA sequence-structure motifs, is of great importance (Cirillo et al., 2016; Heller et al., 2017; Krakau et al., 2017; Budach and Marsico, 2018). Such tools facilitate the analysis and characterization of functional RNA-protein interactions and create a meaningful starting dataset of potential candidates to further dissect their functions experimentally.

By combining and employing both experimental and computational strategies, locus specific genetic manipulation and complementary large-scale analyses, future research is expected to comprehensively

characterize the mechanistic details and functional impact of non-coding RNA transcription and its various products in the precise regulation of target gene expression. We have only started scratching the surface.

Acknowledgements

References

Amaral, P.P., Clark, M.B., Gascoigne, D.K., et al. (2011). lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res *39*, D146-151.

Amaral, P.P., Leonardi, T., Han, N., et al. (2018). Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. Genome Biol *19*, 32.

Amit, M., Donyo, M., Hollander, D., et al. (2012). Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. Cell Rep *1*, 543-556.

Andersson, R., Gebhard, C., Miguel-Escalada, I., et al. (2014a). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Andersson, R., Refsing Andersen, P., Valen, E., et al. (2014b). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. Nat Commun *5*, 5336.

Andersson, R., Sandelin, A., and Danko, C.G. (2015). A unified architecture of transcriptional regulatory elements. Trends Genet *31*, 426-433.

Arnold, C.D., Gerlach, D., Stelzer, C., et al. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074-1077.

Arnold, C.D., Zabidi, M.A., Pagani, M., et al. (2017). Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. Nat Biotechnol *35*, 136-144.

Berretta, J., and Morillon, A. (2009). Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO Rep *10*, 973-982.

Bogu, G.K., Vizan, P., Stanton, L.W., et al. (2015). Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. Mol Cell Biol *36*, 809-819.

Brinster, R.L., Allen, J.M., Behringer, R.R., et al. (1988). Introns increase transcriptional efficiency in transgenic mice. Proc Natl Acad Sci U S A *85*, 836-840.

Brockdorff, N. (2002). X-chromosome inactivation: closing in on proteins that bind Xist RNA. Trends Genet *18*, 352-358.

Budach, S., and Marsico, A. (2018). pysster: classification of biological sequences by learning sequence and structure motifs with convolutional neural networks. Bioinformatics *34*, 3035-3037.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., et al. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Cabianca, D.S., Casa, V., Bodega, B., et al. (2012). A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell *149*, 819-831.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

Cerase, A., Pintacuda, G., Tattermusch, A., et al. (2015). Xist localization and function: new insights from multiple levels. Genome Biol *16*, 166.

Cirillo, D., Blanco, M., Armaos, A., et al. (2016). Quantitative predictions of protein interactions with long noncoding RNAs. Nat Methods *14*, 5-6.

Cirillo, L.A., Lin, F.R., Cuesta, I., et al. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell *9*, 279-289.

Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., et al. (2012). Genome-wide analysis of long noncoding RNA stability. Genome Res *22*, 885-898.

Cong, L., Ran, F.A., Cox, D., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Core, L.J., Martins, A.L., Danko, C.G., et al. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311-1320.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., et al. (2014). Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res *24*, 1-13.

Dao, L.T.M., Galindo-Albarran, A.O., Castro-Mondragon, J.A., et al. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. Nat Genet *49*, 1073-1081.

De Santa, F., Barozzi, I., Mietton, F., et al. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

Derrien, T., Johnson, R., Bussotti, G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res *22*, 1775-1789.

Diao, Y., Fang, R., Li, B., et al. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. Nat Methods *14*, 629-635.

Dinger, M.E., Amaral, P.P., Mercer, T.R., et al. (2009). Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. Brief Funct Genomic Proteomic *8*, 407-423.

Dixon, J.R., Selvaraj, S., Yue, F., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Djebali, S., Davis, C.A., Merkel, A., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101-108.

Ebisuya, M., Yamamoto, T., Nakajima, M., et al. (2008). Ripples from neighbouring transcription. Nat Cell Biol *10*, 1106-1113.

Engreitz, J.M., Haines, J.E., Perez, E.M., et al. (2016a). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature *539*, 452-455.

Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016b). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat Rev Mol Cell Biol *17*, 756-770.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods *9*, 215-216.

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature *414*, 929-933.

Fraser, J., Ferrai, C., Chiariello, A.M., et al. (2015). Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol Syst Biol *11*, 852.

Fullwood, M.J., Liu, M.H., Pan, Y.F., et al. (2009a). An oestrogen-receptor-α-bound human chromatin interactome. Nature *462*, 58-64.

Fullwood, M.J., and Ruan, Y. (2009). ChIP-based methods for the identification of long-range chromatin interactions. J Cell Biochem *107*, 30-39.

Fullwood, M.J., Wei, C.L., Liu, E.T., et al. (2009b). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. Genome Res *19*, 521-532.

Gawronski, A.R., Uhl, M., Zhang, Y., et al. (2018). MechRNA: prediction of lncRNA mechanisms from RNA-RNA and RNA-protein interactions. Bioinformatics *34*, 3101-3110.

Gil, N., and Ulitsky, I. (2018). Production of Spliced Long Noncoding RNAs Specifies Regions with Increased Enhancer Activity. Cell Syst *7*, 537-547 e533.

Groff, A.F., Barutcu, A.R., Lewandowski, J.P., et al. (2018). Enhancers in the Peril lincRNA locus regulate distant but not local genes. Genome Biol *19*, 219.

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

Guttman, M., Amit, I., Garber, M., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

Hacisuleyman, E., Goff, L.A., Trapnell, C., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol *21*, 198-206.

Haerty, W., and Ponting, C.P. (2015). Unexpected selection to retain high GC content and splicing enhancers within exons of multiexonic lncRNA loci. RNA *21*, 333-346.

Hah, N., Danko, C.G., Core, L., et al. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622-634.

Hah, N., Murakami, S., Nagari, A., et al. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res *23*, 1210-1223.

He, Y., Gorkin, D.U., Dickel, D.E., et al. (2017). Improved regulatory element prediction based on tissue-specific local epigenomic signatures. Proc Natl Acad Sci U S A *114*, E1633-E1640.

Heintzman, N.D., Stuart, R.K., Hon, G., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Heller, D., Krestel, R., Ohler, U., et al. (2017). ssHMM: extracting intuitive sequence-structure motifs from high-throughput RNA-binding protein data. Nucleic Acids Res *45*, 11004-11018.

Henriques, T., Scruggs, B.S., Inouye, M.O., et al. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev *32*, 26-41.

Hicks, M.J., Yang, C.R., Kotlajich, M.V., et al. (2006). Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. PLoS Biol *4*, e147.

Hoffman, M.M., Ernst, J., Wilder, S.P., et al. (2013). Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res *41*, 827-841.

Hoki, Y., Kimura, N., Kanbayashi, M., et al. (2009). A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. Development *136*, 139-146.

Hon, C.C., Ramilowski, J.A., Harshbarger, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. Nature *543*, 199-204.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. Mol Cell *52*, 473-484.

Kapranov, P., Cheng, J., Dike, S., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484-1488.

Khalil, A.M., Guttman, M., Huarte, M., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A *106*, 11667-11672.

Kim, S., Kim, H., Fong, N., et al. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. Proc Natl Acad Sci U S A *108*, 13564-13569.

Kim, T.K., Hemberg, M., Gray, J.M., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, T.K., and Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. Cell *162*, 948-959.

Kirk, J.M., Kim, S.O., Inoue, K., et al. (2018). Functional classification of long non-coding RNAs by k-mer content. Nat Genet *50*, 1474-1482.

Koch, F., and Andrau, J.C. (2011). Initiating RNA polymerase II and TIPs as hallmarks of enhancer activity and tissue-specificity. Transcription *2*, 263-268.

Koch, F., Fenouil, R., Gut, M., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat Struct Mol Biol *18*, 956-963.

Kowalczyk, M.S., Hughes, J.R., Garrick, D., et al. (2012). Intragenic enhancers act as alternative promoters. Mol Cell *45*, 447-458.

Krakau, S., Richard, H., and Marsico, A. (2017). PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. Genome Biol *18*, 240.

Kwek, K.Y., Murphy, S., Furger, A., et al. (2002). U1 snRNA associates with TFIIH and regulates transcriptional initiation. Nat Struct Biol *9*, 800-805.

Lalmansingh, A.S., Karmakar, S., Jin, Y., et al. (2012). Multiple modes of chromatin remodeling by Forkhead box proteins. Biochim Biophys Acta *1819*, 707-715.

Latos, P.A., Pauler, F.M., Koerner, M.V., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science *338*, 1469-1472.

Le Hir, H., Nott, A., and Moore, M.J. (2003). How introns influence and enhance eukaryotic gene expression. Trends Biochem Sci *28*, 215-220.

Li, G., Ruan, X., Auerbach, R.K., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84-98.

Liao, J., He, Q., Li, M., et al. (2016). LncRNA MIAT: Myocardial infarction associated and more. Gene *578*, 158-161.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lin, M.F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics *27*, i275-282.

Liu, F., Marquardt, S., Lister, C., et al. (2010). Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. Science *327*, 94-97.

Lubas, M., Andersen, P.R., Schein, A., et al. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. Cell Rep *10*, 178-192.

Lupianez, D.G., Kraft, K., Heinrich, V., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.

Mahat, D.B., Kwak, H., Booth, G.T., et al. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat Protoc *11*, 1455-1476.

Mali, P., Yang, L., Esvelt, K.M., et al. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Marquardt, S., Raitskin, O., Wu, Z., et al. (2014). Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. Mol Cell *54*, 156-165.

Marques, A.C., Hughes, J., Graham, B., et al. (2013). Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. Genome Biol *14*, R131.

McCracken, S., Fong, N., Yankulov, K., et al. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.

Mele, M., Mattioli, K., Mallard, W., et al. (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. Genome Res *27*, 27-37.

Melgar, M.F., Collins, F.S., and Sethupathy, P. (2011). Discovery of active enhancers through bidirectional expression of short transcripts. Genome Biol *12*, R113.

Mikhaylichenko, O., Bondarenko, V., Harnett, D., et al. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev *32*, 42-57.

Mouse, E.C., Stamatoyannopoulos, J.A., Snyder, M., et al. (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol *13*, 418.

Muerdter, F., Boryn, L.M., and Arnold, C.D. (2015). STARR-seq - principles and applications. Genomics *106*, 145-150.

Mukherjee, N., Calviello, L., Hirsekorn, A., et al. (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. Nat Struct Mol Biol *24*, 86-96.

Nagano, T., Mitchell, J.A., Sanz, L.A., et al. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717-1720.

Nagari, A., Murakami, S., Malladi, V.S., et al. (2017). Computational Approaches for Mining GRO-Seq Data to Identify and Characterize Active Enhancers. Methods Mol Biol *1468*, 121-138.

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., et al. (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res *11*, 833-849.

Nguyen, T.A., Jones, R.D., Snavely, A.R., et al. (2016). High-throughput functional comparison of promoter and enhancer activities. Genome Res *26*, 1023-1033.

Nitsche, A., Rose, D., Fasold, M., et al. (2015). Comparison of splice sites reveals that long noncoding RNAs are evolutionarily well conserved. RNA *21*, 801-812.

Nojima, T., Tellier, M., Foxwell, J., et al. (2018). Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. Mol Cell *72*, 970-984 e977.

Nora, E.P., Lajoie, B.R., Schulz, E.G., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.

Ntini, E., Louloupi, A., Liz, J., et al. (2018). Long ncRNA A-ROD activates its target gene DKK1 at its release from chromatin. Nat Commun *9*, 1636.

Orom, U.A., Derrien, T., Beringer, M., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.

Pandya-Jones, A., Bhatt, D.M., Lin, C.H., et al. (2013). Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. RNA *19*, 811-827.

Paralkar, V.R., Taborda, C.C., Huang, P., et al. (2016). Unlinking an lncRNA from Its Associated cis Element. Mol Cell *62*, 104-110.

Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol Cell *36*, 178-191.

Pertea, M., Pertea, G.M., Antonescu, C.M., et al. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol *33*, 290-295.

Ponjavic, J., Ponting, C.P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res *17*, 556-565.

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

Purmann, A., Toedling, J., Schueler, M., et al. (2007). Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality. Genomics *89*, 580-587.

Rajagopal, N., Srinivasan, S., Kooshesh, K., et al. (2016). High-throughput mapping of regulatory DNA. Nat Biotechnol *34*, 167-174.

Rigo, F., and Martinson, H.G. (2009). Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. RNA *15*, 823-836.

Sahlen, P., Abdullayev, I., Ramskold, D., et al. (2015). Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. Genome Biol *16*, 156.

Sanyal, A., Lajoie, B.R., Jain, G., et al. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109-113.

Schuler, A., Ghanbarian, A.T., and Hurst, L.D. (2014). Purifying selection on splice-related motifs, not expression level nor RNA folding, explains nearly all constraint on human lincRNAs. Mol Biol Evol *31*, 3164-3183.

Senner, C.E., Nesterova, T.B., Norton, S., et al. (2011). Disruption of a conserved region of Xist exon 1 impairs Xist RNA localisation and X-linked gene silencing during random and imprinted X chromosome inactivation. Development *138*, 1541-1550.

Shiraki, T., Kondo, S., Katayama, S., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A *100*, 15776-15781.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-286.

Sigova, A.A., Abraham, B.J., Ji, X., et al. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978-981.

Soufi, A., Garcia, M.F., Jaroszewicz, A., et al. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555-568.

Stojic, L., Niemczyk, M., Orjalo, A., et al. (2016). Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions. Nat Commun *7*, 10406.

Symmons, O., Uslu, V.V., Tsujimura, T., et al. (2014). Functional and topological characteristics of mammalian regulatory domains. Genome Res *24*, 390-400.

Tan, J.Y., Sirey, T., Honti, F., et al. (2015). Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells. Genome Res *25*, 655-666.

Tan, J.Y., Smith, A.A.T., Ferreira da Silva, M., et al. (2017). cis-Acting Complex-Trait-Associated lincRNA Expression Correlates with Modulation of Chromosomal Architecture. Cell Rep *18*, 2280-2288.

Tippens, N.D., Vihervaara, A., and Lis, J.T. (2018). Enhancer transcription: what, where, when, and why? Genes Dev *32*, 1-3.

Trapnell, C., Williams, B.A., Pertea, G., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

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

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26-46.

van Arensbergen, J., FitzPatrick, V.D., de Haas, M., et al. (2017). Genome-wide mapping of autonomous promoter activity in human cells. Nat Biotechnol *35*, 145-153.

Vucicevic, D., Corradin, O., Ntini, E., et al. (2015). Long ncRNA expression associates with tissue-specific enhancers. Cell Cycle *14*, 253-260.

Wang, K.C., Yang, Y.W., Liu, B., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120-124.

Wang, L., Park, H.J., Dasari, S., et al. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res *41*, e74.

Wang, X., Arai, S., Song, X., et al. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature *454*, 126-130.

Weintraub, A.S., Li, C.H., Zamudio, A.V., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573-1588 e1528.

Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? Cell *155*, 990-996.

Yin, Y., Yan, P., Lu, J., et al. (2015). Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. Cell Stem Cell *16*, 504-516.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., et al. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556-559.

Figure 1 Strategy used for enhancer annotation and identification of enhancer-associated lncRNAs. (**A**) The annotation of enhancers starts from the mapping of DNAse I hypersensitivity sites (DHSs) that reflect increased chromatin accessibility. CAGE (Andersson et al., 2014a) and/or nascent RNA sequencing data like GRO-seq (Hah et al., 2013) or PRO-seq (Mahat et al., 2016) are used to detect substantial Pol II transcription emanating bidirectionally from the DHS. The unstable eRNAs, which are short, non-spliced, and non-polyadenylated, are the products of bidirectional Pol II transcription activity at the enhancer and can be detected using *de novo* transcript assembly in GRO-seq data. Apart from Pol II and general transcription factors (GTFs) like CBP/p300 binding at the DHS, the presence of active histone marks like H3K27Ac and H3K4me1 validate the enhancer. eRNAs are terminated early by polyadenylation (pAS)-like signals and are rapidly turned-over by the exosome, hence they are not readily detectable in steady-state RNA-seq data. (**B**) In some cases (estimated ~5% of active enhancers), Pol II transcription initiating intrinsically bidirectionally at active enhancers is preferentially elongated in one direction, most probably due to the accumulation of activating random mutations leading to enrichment of U1 sites and splicing signals, which suppress early pAS (Wu and Sharp, 2013; Gil and Ulitsky, 2018). The production of long, spliced, and polyadenylated lncRNAs, which are stable and thus readily detectable in steady-stade RNA-seq data, is associated with higher enhancer activity; the latter is reflected by increased chromatin accessibility (DHS) and enrichment of positive histone modifications: H3K27Ac (enhancer mark), H3K4me3 (promoter mark), and H3K36me3 (a hallmark of transcription elongation). (**C**) The genomic topology of the predicted active enhancer is also examined, and its position relative to anchor points of chromosomal loops is characterized from Pol II ChIA–PET interactions and/or TAD boundaries. One or more putative target genes are identified based on long-range 3D interactions, and induced effects on target gene expression are examined in conditions of enhancer perturbation, including genetic manipulation of the enhancer-associated lncRNA locus. (**D**) Flow diagram of the steps employed for enhancer annotation and enhancer-associated lncRNA identification.
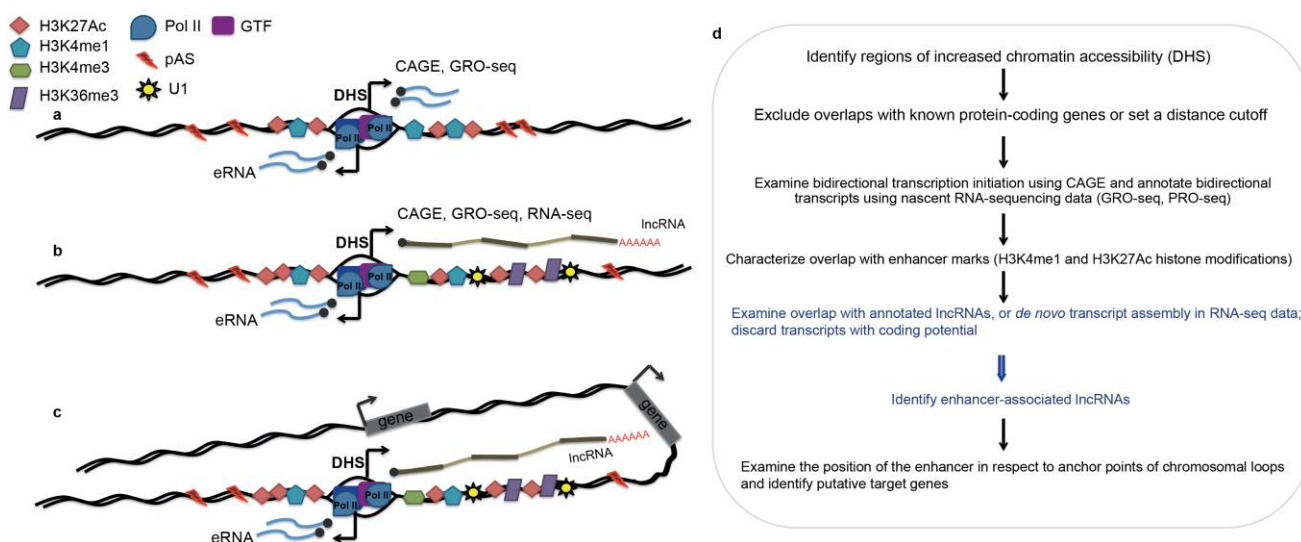
Figure 2 Possible mechanisms explaining the functional impact of lncRNA splicing on enhancer activity. (**A**) Splicing factors cotranscriptionally engaged at the conserved lncRNA splice sites may in turn recruit activating factors like histone modification enzymes and chromatin remodelers, resulting in an overall chromatin opening of the locus and a positive effect on *cis*-gene regulation. This happens while the lncRNA is tethered on chromatin, at its site of transcription, during cotranscriptional processing. (**B**) Splicing permits dissociation of the nascent lncRNA transcript from chromatin, which in turn interacts with additional proteins contributing to spatial amplification of the *cis*-regulatory signal. The two distinct possible mechanisms are not mutually exclusive.
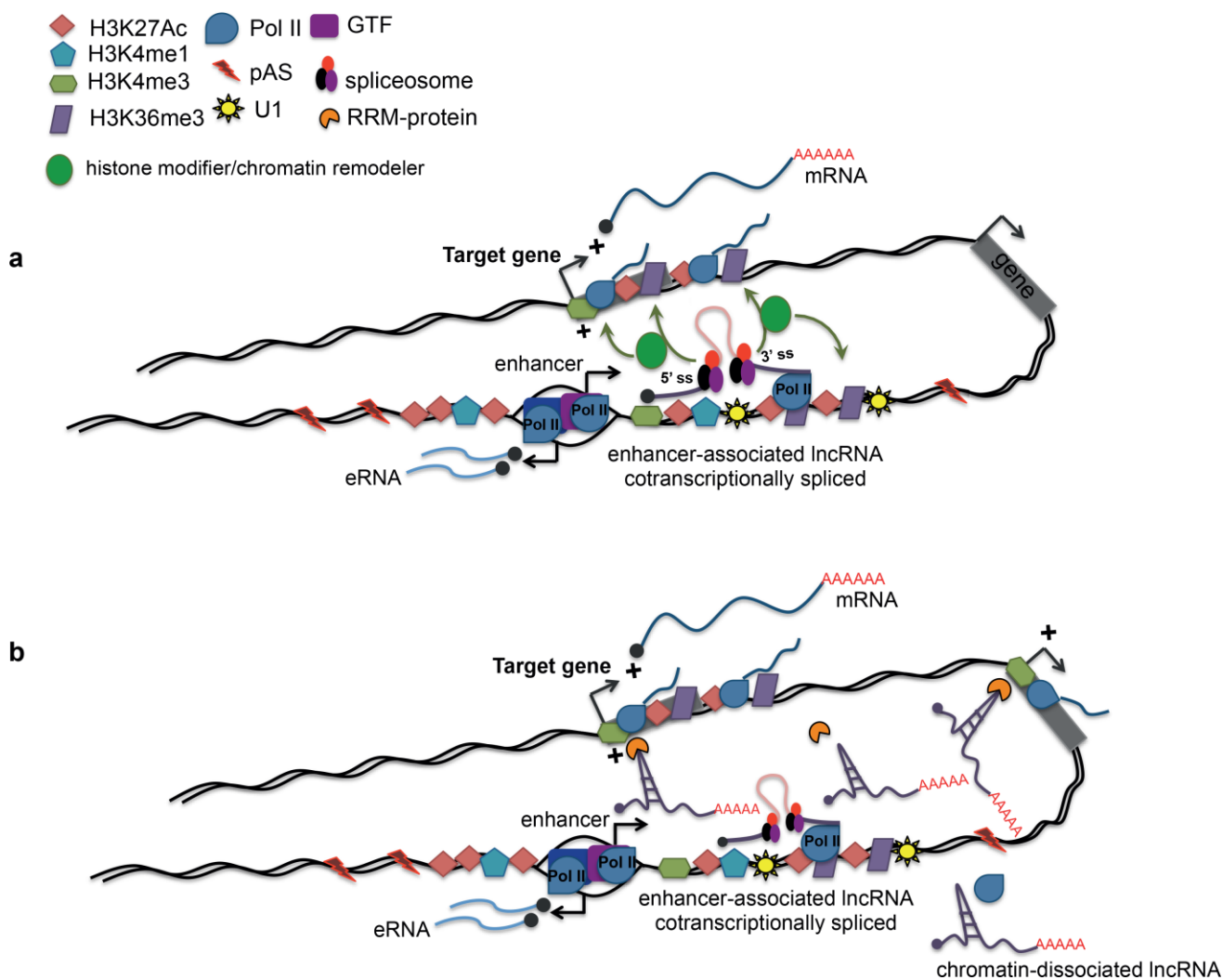
Table 1 Glossary.

Term	Description	Reference
ChromHMM	Software developed to learn and characterize chromatin states, by integrating chromatin datasets, such as ChIP-seq of histone modifications. It employs a multivariate Hidden Markov Model, which models the presence or absence of each histone mark.	(Ernst and Kellis, 2012)
CAGE	<u>C</u>apped <u>A</u>nalysis of <u>G</u>ene <u>E</u>xpression is a technique that maps the 5' ends of capped RNA transcripts, thereby capturing transcription initiation sites genome-wide.	(Shiraki et al., 2003)
GRO-seq	<u>G</u>lobal <u>R</u>un <u>O</u>n sequencing is a technique that maps transcriptionally engaged RNA polymerase II (Pol II) genome-wide, capturing nascent RNA transcription. In this method, Pol II is allowed to run on in the presence of a labeled nucleotide analog (5'-bromouridine), which is incorporated into newly transcribed RNA.	(Core et al., 2008)
ATAC-seq	<u>A</u>ssay for <u>T</u>ransposable <u>A</u>ccessible <u>C</u>hromatin followed by deep <u>s</u>equencing is a technique that captures open chromatin sites. It is used to identify regions of open chromatin, such as nucleosome-free positions in regulatory regions. It is based on the transposition of sequencing adapters into native chromatin *in vitro*. Essentially, it probes chromatin accessibility using transposons *in vitro*.	(Buenrostro et al., 2013)
STARR-seq	<u>S</u>elf-<u>T</u>ranscribing <u>A</u>ctive <u>R</u>egulatory <u>R</u>egion <u>s</u>equencing measures enhancer activity (enhancer strength) genome-wide, by assaying candidate sequences from any source of DNA using a reporter system *in vitro*.	(Arnold et al., 2013)
ChIA-PET	<u>Ch</u>romatin <u>I</u>nteraction <u>A</u>nalysis by <u>P</u>aired-<u>E</u>nd <u>T</u>ag sequencing provides large-scale analysis of long-range chromatin interaction networks. In this technique, cross-linked chromatin interaction sites bound by specific proteins are enriched by chromatin immunoprecipitation, and the remote DNA elements interacting in close spatial proximity are connected through proximity ligation.	(Fullwood et al., 2009a,b; Fullwood and Ruan, 2009)
TADs	<u>T</u>opologically <u>A</u>ssociating <u>D</u>omains are considered the basic organizational units of chromosome architecture, covering hundreds of kilobases to several million bases in length. DNA sequences within a TAD physically interact with each other more frequently than with sequences outside the TAD.	(Dixon et al., 2012; Nora et al., 2012)
Hi-C	Hi-C is based on chromosome conformation capture (3C) technique. It captures the 3D architecture of entire genomes, by coupling proximity ligation of DNA fragments tethered together in spatial proximity (enriched by crosslinking) with high-throughput sequencing, to produce genome-wide interaction maps.	(Lieberman-Aiden et al., 2009)