TRANSPARENT PROCESS   OPEN ACCESS

Life Science Alliance

Check for updates

Splicing of enhancer-associated lincRNAs contributes to enhancer activity

Jennifer Y Tan[1] , Adriano Biasini[1] , Robert S Young[2] , Ana C Marques[1]

Transcription is common at active mammalian enhancers sometimes giving rise to stable enhancer-associated long intergenic noncoding RNAs (elincRNAs). Expression of elincRNA is associated with changes in neighboring gene product abundance and local chromosomal topology, suggesting that transcription at these loci contributes to gene expression regulation in *cis*. Despite the lack of evidence supporting sequence-dependent functions for most elincRNAs, splicing of these transcripts is unexpectedly common. Whether elincRNA splicing is a mere consequence of cognate enhancer activity or if it directly impacts enhancer function remains unresolved. Here, we investigate the association between elincRNA splicing and enhancer activity in mouse embryonic stem cells. We show that multiexonic elincRNAs are enriched at conserved enhancers, and the efficient processing of elincRNAs is strongly associated with their cognate enhancer activity. This association is supported by their enrichment in enhancer-specific chromatin signatures; elevated binding of co-transcriptional regulators; increased local intra-chromosomal DNA contacts; and strengthened *cis*-regulation on target gene expression. Our results support the role of efficient RNA processing of enhancer-associated transcripts to cognate enhancer activity.

Introduction

Enhancers are distal DNA elements that positively drive target gene expression (Banerji et al, 1981; Moreau et al, 1981; Li et al, 2016). These regulatory regions are DNase I hypersensitive, marked by histone 3 acetylation at lysine 27 (H3K27ac), and a high ratio of monomethylation versus trimethylation at histone 3 lysine 4 (H3K4me1 and H3K4me3, respectively). Together, these chromatin signatures are commonly used to annotate enhancers genome wide (Hoffman et al, 2012). Most active enhancers are also transcribed (De Santa et al, 2010; Kim et al, 2010; Kowalczyk et al, 2012). Relative to non-transcribed enhancers, those that give rise to enhancer-associated transcripts are more strongly associated with enhancer-specific chromatin signatures (Wang et al, 2011) and display higher levels of reporter activity both in vitro (Wu et al, 2014; Young et al, 2017) and in vivo (Andersson et al, 2014), supporting the link between enhancer transcription and *cis*-regulatory function. Whereas most enhancers transcribe short noncoding RNAs that are non-polyadenylated, unspliced, and short-lived from both the sense and antisense strands (eRNAs) (Kim et al, 2010), a subset of enhancers are predominantly transcribed in one direction (Natoli & Andrau, 2012) and produce enhancer-associated long intergenic noncoding transcripts that we refer to as elincRNAs (Marques et al, 2013). The asymmetry of transcriptional activity at these enhancers is at least in part due to differences in transcript stability. Specifically, and in contrast to eRNAs, elincRNAs are polyadenylated, relatively long, stable, and frequently spliced (Koch et al, 2011; Marques et al, 2013; Hon et al, 2017).

Enhancer transcription can increase local chromatin accessibility (Mousavi et al, 2013), modulate chromosomal interactions between cognate enhancer and target promoters (Lai et al, 2013), and regulate the load, pause, and release of RNA Polymerase II (RNAPlI) (Maruyama et al, 2014; Schaukowitch et al, 2014), ultimately contributing to enhanced expression of neighboring protein-coding genes (Orom et al, 2010; Marques et al, 2013). Recently, we showed that elincRNAs preferentially locate at topologically associating domain (TAD) boundaries, and their expression correlates with changes in local chromosomal architecture (Tan et al, 2017). Although the association between elincRNA transcription and enhancer activity is relatively well established, whether the molecular mechanisms underlying their functions depend on their transcript sequences has not yet been unequivocally demonstrated. Notably, consistent with the absence of nucleotide conservation at their exons (Marques et al, 2013), many elincRNA functions appear to rely on transcription alone (Yoo et al, 2012; Lai et al, 2013; Li et al, 2013; Hsieh et al, 2014; Alexanian et al, 2017).

Despite evidence that the functions of most elincRNAs is likely transcription dependent, a relatively large proportion of elincRNAs is not only stably transcribed but also undergoes splicing (Marques et al, 2013; Hon et al, 2017; Krchnakova et al, 2019). Recently, splicing of Blustr, a lincRNA expressed in mouse embryonic stem cells (mESCs) whose transcriptional start site initiates from an active enhancer (Mouse ENCODE Consortium et al, 2012), was shown to be sufficient to modulate the expression of its

[1]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland   [2]Medical Research Council Human Genetics Unit, Medical Research Council Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK

Correspondence: anaclaudia.marques@unil.ch; jennifer.tan@unil.ch

cognate protein-coding gene target in *cis* (Engreitz et al, 2016). Removal of the splicing signals in another elincRNA, Haunt, by replacing its endogenous locus with its cDNA, could not rescue its *cis*-regulatory function (Yin et al, 2015). Recently, the genome-wide analysis of enhancer transcription across multiple human cells lines (Gil & Ulitsky, 2018) corroborates candidate loci analyses, supporting the association between elincRNA splicing and cognate enhancer activity.

Here, we investigate the association between elincRNA splicing and developmentally regulated mESC enhancer's activity. We show that efficient splicing of multi-exonic elincRNAs associates with higher activity, cell-type–specific function and increased conservation of their cognate enhancers.

Results

To annotate enhancer-associated lincRNAs (elincRNAs), we took advantage of the extensive publicly available data for transcription and chromatin signatures in pluripotent mESC. We considered all intergenic mESC enhancers overlapping a DNase I–hypersensitive region (The ENCODE Project Consortium, 2012) and annotated their associated transcripts using a stringent approach that required the overlap between their transcriptional start site and the enhancer. This led to the identification of a relatively small, yet high confidence, set of enhancer-associated lincRNAs (n = 100, elincRNAs, Table S1) and eRNAs (n = 2,117). As expected (Xu et al, 2009; Andersson et al, 2014; Young et al, 2017), we found divergent transcription at all promoter and enhancer-associated transcriptional initiation regions (TIRs, Fig 1A–D). In contrast to eRNA-producing enhancers (Fig 1A), enhancers associated with elincRNAs (Fig 1B) have transcriptional profiles that resemble those of other promoter-associated mESC transcripts, including other mESC-expressed non–enhancer-associated lincRNAs (oth-lincRNAs) (Fig 1C) and protein-coding genes (Fig 1D).

Given the relatively small number of the stringently annotated elincRNAs, we also annotated elincRNAs using a less stringent approach. Analysis of this less stringent and more comprehensive set of mESC elincRNA (1,983 elincRNAs of which 211 are multi-exonic) is

described in the Supplemental Data 1 and fully supports the analysis of the stringently annotated set of mESC elincRNAs.

Multi-exonic elincRNAs are associated with stronger enhancer activity

Next, we investigated whether elincRNA splicing is linked to its cognate enhancer activity. Given that most enhancer activity is tissue specific, we first investigated the association between enhancer transcription and putative target expression during embryonic neurogenesis (Fraser et al, 2015). Similar to what was described previously (Marques et al, 2013), we found that elincRNA transcription positively correlated with changes in neighboring protein-coding gene abundance (Fig S1A). This association is 2.5-fold stronger for multi-exonic elincRNAs (median FD target transcription = 0.49) than their single-exonic counterparts (median FD target transcription = 0.19, $P < 0.05$, two-tailed Mann–Whitney U test, Fig 2A). As expected, no association was observed for other transcript classes, regardless of their splicing activity (Fig S1B).

Consistent with their stronger association with neighboring protein-coding gene expression, chromatin signatures associated with high enhancer activity were found at enhancers that transcribe multi-exonic elincRNAs compared with those that give rise to either single-exonic elincRNAs or eRNAs. Specifically, multi-exonic elincRNA-producing enhancers were enriched for monomethylation of histone 3 lysine 4 (H3K4me1, Fig 2B), acetylation of histone 3 lysine 27 (H3K27ac, Fig 2C), and DNase I accessibility (DHSI, Fig 2D). Using a hypothesis-free approach, we found that relative to their unspliced counterparts, TIRs of multi-exonic elincRNAs were significantly enriched (false discovery rate < 0.05) for transcription factor–binding motifs required for the recruitment of the transcriptional co-activator cAMP-response element-binding protein (CREB)–binding protein (CREBBP) (Bedford et al, 2010), including Stat1, Egr1, Sp2, Smad3, and Klf5 (Table S2). For a subset of the enriched CREBBP-recruiting transcription factors with available chromatin immunoprecipitation (ChIP) sequencing data in mESCs and the CREBBP transcriptional co-activator, EP300 (Merika et al, 1998), we found experimental support for their more frequent binding at multi-exonic elincRNAs' TIRs (Figs 2E and S1C–E). Recently, direct binding of CREBBP to enhancer-associated RNAs was demonstrated to stimulate its histone acetylation activity and induce activation of target gene transcription (Bose et al, 2017). Our findings raise the possibility that
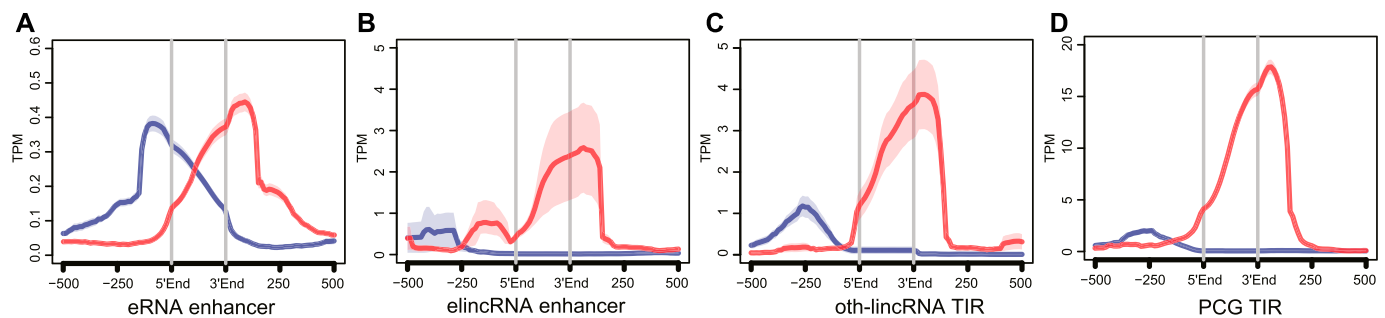

Figure 1. Stringent annotations of elincRNAs.
(A, B, C, D) Metagene plots of CAGE reads centered at transcription initiation regions (TIRs) of (A) eRNAs, (B) elincRNAs, (C) other mouse embryonic stem cell-expressed lincRNAs (oth-lincRNAs), and (D) protein-coding genes (PCGs). Sense (red) and antisense (blue) reads denote those that map to the same or opposite strand, respectively, as the direction of their cognate TIRs.
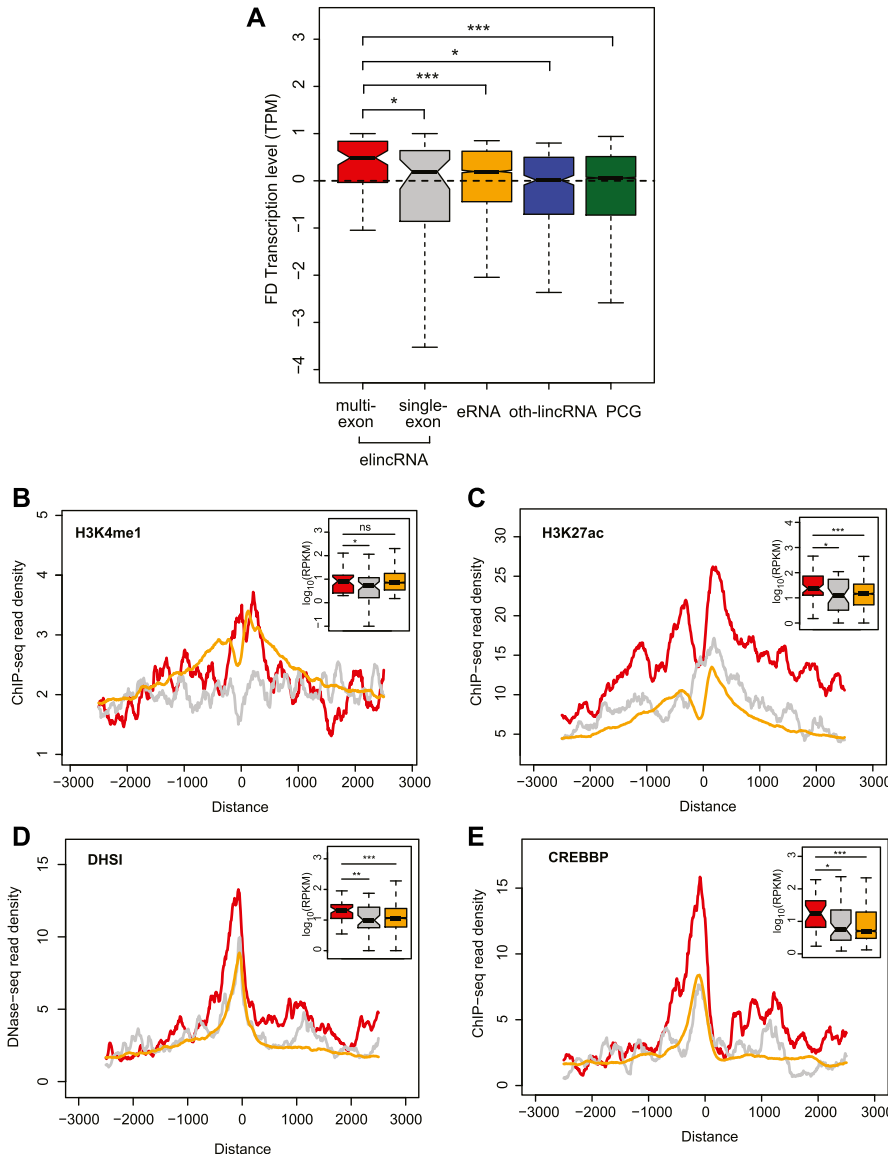
Figure 2. Multi-exonic elincRNAs are associated with higher enhancer activity.
(A) Distribution of the fold difference (FD) in transcription (measured as CAGE TPM) of the most proximal gene to multi-exonic (red) and single-exonic (grey) elincRNAs, eRNAs (yellow), other mouse embryonic stem cell-expressed lincRNAs (oth-lincRNAs, blue), and protein-coding genes (PCGs, green) both expressed in a same stage of embryonic neurogenesis. Fold difference of neighboring genes is calculated between the two cellular stages across neuronal differentiation, where the expression level of their reference locus (elincRNA, oth-lincRNA, or PCG) is maximal and minimal. **(B, C, D, E)** Metagene plots and distribution (figure insets) of (B) H3K4me1, (C) H3K27ac, (D) DNase I hypersensitive sites (DHSI), and (E) Crebbp ChIP-seq reads in mouse embryonic stem cells at transcription-initiation regions of multi-exonic (red) and single-exonic (grey) elincRNAs and eRNAs (yellow). Differences between groups were tested using a two-tailed Mann–Whitney U test. $*P < 0.05$; $***P < 0.001$.

multi-exonic elincRNAs are more likely to physically interact with CREBBP than are other enhancer-derived RNAs.

Multi-exonic elincRNAs are specifically associated with changes in local chromosomal architecture

Because *cis*-regulatory interactions are dependent on local chromosomal architecture, we examined whether the observed association between elincRNA splicing and enhanced neighboring gene expression was mediated through the modulation of their local chromosomal organization.

Analysis of their relative position within mESC TADs revealed that only multi-exonic elincRNA TIRs were significantly enriched at TAD boundaries and depleted at TAD centers ($P < 0.05$, Fig 3A). This suggests that elincRNAs' preferential location at TAD boundaries (Tan et al, 2017) is restricted to multi-exonic elincRNAs. Preferential

localization of multi-exonic elincRNA-transcribing enhancers at TAD boundaries, where chromosomal looping between enhancers and promoters frequently occurs (Symmons et al, 2014; Lupianez et al, 2015), is further supported by the enriched binding of protein factors implicated in the establishment and modulation of chromosomal topology (Bonev & Cavalli, 2016). Relative to their single-exonic counterparts, multi-exonic elincRNA-producing enhancers display evidence for higher binding of Ctcf (Fig S2A), subunits of the cohesin complex (Smc1a and Smc3), its cofactor Nipbl (Fig S2B–D), and the mediator complex (Med1 and Med3) (Fig S2E and F) in mESCs.

Enhancer-associated transcripts participate in enhancer-promoter looping by recruiting Cohesin or Mediator complexes to enhancer regions, which in turn stimulate cognate target gene transcription (Lai et al, 2013; Hsieh et al, 2014). Consistent with the role of multi-exonic elincRNAs and their underlying enhancers in cell-type–specific
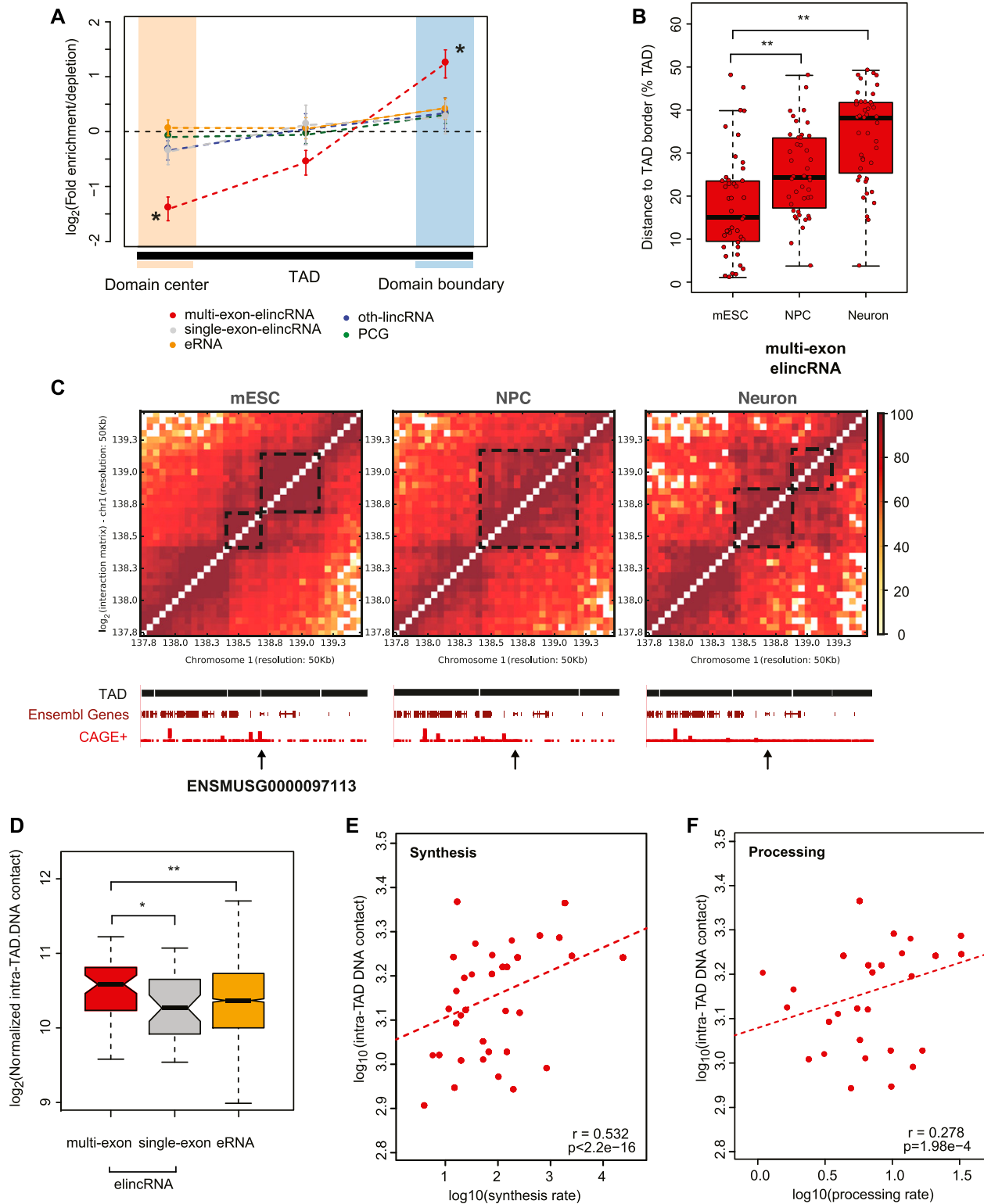
Figure 3. Multi-exonic elincRNAs are associated with modulation of local chromosomal architecture.
(A) Fold enrichment or depletion of multi-exonic (red) and single-exonic (grey) elincRNAs, eRNAs (yellow), other expressed lincRNAs (blue), and protein-coding genes (green) at boundaries (light blue shaded area) and center (light yellow shaded areas) of TADs. Significant fold differences are denoted with * ($P < 0.05$, permutation test) and standard deviation is shown with error bars. **(B)** Distribution of the distance between multi-exonic elincRNA transcription-initiation sites (red) to the nearest TAD border in mouse embryonic stem cells (mESCs), neuronal precursor cells (NPCs), and neurons. **(C)** Heat map displaying the amount of chromosomal interactions, measured using Hi-C data, at regions surrounding one multi-exonic elincRNA (ENSMUSG0000097113) in mESC, NPC, and Neuron. Dotted black squares denote TAD, which is

modulation of local chromosomal structure, we found that although, on average, the location of single-exonic enhancer-derived lincRNAs and eRNAs remained relatively unchanged with respect to their nearest TAD border (Fig S3A), the distance between TAD borders and multi-exonic elincRNA TIRs increases upon cell differentiation (Fig 3B and C). Multi-exonic elincRNA transcription is strongly correlated with the presence and maintenance of TAD boundaries across differentiation, supporting cell-type–specific functions of these enhancers (Fig S3B and C). Furthermore, supporting the tissue-specific activity and functions of multi-exonic elincRNA-transcribing enhancers, we found that genes in their vicinity are enriched in genes involved in mESC pluripotency maintenance (1.73-fold enrichment, $P < 0.05$, hypergeometric test) (Xu et al, 2013) and DNA binding and RNA transcription (Fig S3D).

To assess the impact of multi-exonic elincRNA on local chromosomal architecture, we next investigated the relationship between enhancer transcription and splicing and intra-TAD DNA contact density. We found that the frequency of DNA contacts within TADs that encompass multi-exonic elincRNA loci to be significantly higher than those containing other transcribed enhancers ($P < 0.05$, two-tailed Mann–Whitney U test, Fig 3D, see the Materials and Methods section). Furthermore, we found that the density of local chromosomal interactions correlated with the rate of transcription (Fig 3E) and processing (Fig 3F) of multi-exonic elincRNAs.

Activity of enhancers that transcribe multi-exonic elincRNAs is conserved

We reasoned that if splicing of enhancer-associated transcripts is biologically relevant, multi-exonic elincRNA-producing enhancers should be conserved during evolution. To test this hypothesis, we assessed the extent of enhancer conservation by overlapping the syntenic regions of transcribed mESC enhancers in humans with H1 ESC (hESC) enhancers (The ENCODE Project Consortium, 2012). We found that more than half (n = 57/100, 57%) of mESC enhancers that produce elincRNAs have conserved chromatin signatures at their syntenic regions in hESCs, a significantly higher proportion than those that produce eRNAs (n = 487/2,117, 23%, $P < 5 \times 10^{-13}$, two-tailed Fisher's exact test). Furthermore, relative to enhancers that transcribe single-exonic elincRNAs, those that express multi-exonic elincRNAs are twofold enriched among conserved enhancers ($P < 1 \times 10^{-4}$, two-tailed Fisher's exact test). Importantly, of the conserved enhancers with evidence of transcription in humans (n = 12/57, 21%), most give rise to multi-exonic elincRNAs in mESCs (n = 10/12, 83%), consistent with the conservation of the function and transcription of these enhancers during mammalian evolution.

Rapid elincRNA splicing is associated with efficient transcription

We next turned our attention to the mechanisms and sequences underlying the splicing of elincRNAs. Differences in GC content between intronic and exonic sequences are known to facilitate splice site recognition and increase splicing efficiency (Amit et al, 2012). The exons and introns of elincRNAs display distinct GC contents, similar to protein-coding genes and oth-lincRNAs (Fig 4A) (Schuler et al, 2014; Haerty & Ponting, 2015). Further supporting the biological relevance of elincRNA splicing, we found that their splice site (SS)–flanking regions are enriched in splicing-associated elements, including exonic splicing enhancers (Fig 4B) and U1 snRNP-binding motifs (Fig 4C). Relative to other multi-exonic lincRNAs, elincRNAs SSs also have a higher likelihood of being recognized by the splicing machinery (Fig S4A and B). Together, these results suggest elincRNA splicing is efficient.

To assess whether increased density of splicing-associated motifs at multi-exonic elincRNA reflect efficient transcript splicing at these loci, we determined their transcriptome-wide rates of splicing in mESCs. We performed 4-thiouridine (4sU) metabolic labeling of RNA for 15, 30, and 60 min. Ribo-depleted total RNA from the total and newly transcribed fractions was sequenced and used to estimate transcriptome-wide rates of synthesis, splicing, and degradation in mESCs using INSPEcT (de Pretis et al, 2015) (Fig S4C). Consistent with previous reports, lincRNAs as a class were significantly less efficiently spliced than protein-coding genes (Mele et al, 2017; Mukherjee et al, 2017). However, compared with other lincRNAs, those transcribed from enhancers were 1.5-fold more rapidly processed (Fig 4D) and a higher proportion of their introns (14%) have undergone complete splicing (Fig 4E, $P < 0.05$, two-tailed Mann–Whitney U test, Table S3). The splicing efficiency of elincRNAs was comparable with that of protein-coding genes (Fig 4D and E). No significant differences were found in the synthesis and degradation rates between elincRNAs and other lincRNAs ($P > 0.05$ two-tailed Mann–Whitney U test, Fig S4D).

We found the exons of multi-exonic elincRNA evolved neutrally (Fig S5), suggesting efficient splicing of these transcripts was not maintained to preserve the assembly of evolutionarily conserved and likely functional sequence motifs within their primary transcripts. Given the well-established coupling between splicing and transcription (Brinster et al, 1988; Le Hir et al, 2003) and higher splicing efficiency of elincRNA 5′ exons (Fig 5A, $P < 0.05$, two-tailed Mann–Whitney U test), which was not detected for mRNAs or oth-lincRNAs (Fig 5A), we questioned if splicing was instead associated with higher transcription of multi-exonic elincRNA loci. Consistent with this hypothesis, we found multi-exonic elincRNA transcripts were more rapidly synthesized than their single-exonic counterparts (Fig 5B). This higher transcriptional activity was further supported by elevated levels of engaged RNA Polymerase II (RNAPII, Fig 5C) at their TIRs and lower RNAPII promoter-proximal stalling relative to other noncoding transcripts, as shown by their relatively low ratio between RNAPII reads mapping to their TIR relative to their gene body (Travelling Ratio, Fig 5D, $P < 0.05$, two-tailed Mann–Whitney U test, see the Materials and Methods section). Furthermore, relative to other non-spliced ncRNAs, multi-exonic elincRNA TIRs and gene bodies were enriched in phosphorylated serine 5 (S5P) and serine 2 (S2P) (Fig 5E and F) at

also represented by the black bars below the heat map. Gene browser view of the corresponding region displaying Ensembl gene models (dark red lines) and CAGE read density (red lines) at each cell stage. **(D)** Distribution of the average amount of chromosomal contacts within mESC TADs that contain multi-exonic (red) and single-exonic (grey) elincRNAs and eRNAs (yellow). **(E, F)** DNA–DNA contacts within multi-exonic elincRNA-containing mESC TADs (log10, y-axis) as a function of their respective (E) synthesis rate or (F) processing rate (log10, red points, Spearman's correlation). Differences between groups were tested using a two-tailed Mann–Whitney U test. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; NS $P > 0.05$.
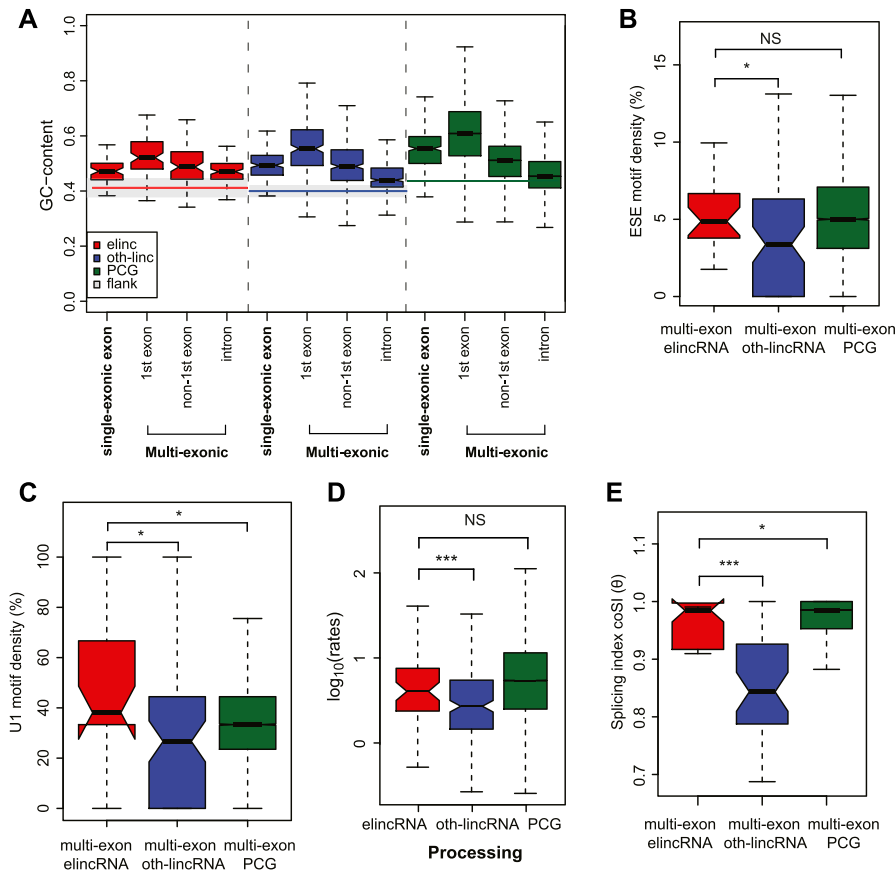
Figure 4. elincRNA splicing is efficient.
(A) Distribution of the GC content of exons and introns of single- and multi-exonic elincRNAs (red), other expressed lincRNAs (blue), protein-coding genes (green), and their respective flanking regions (grey). **(B, C)** Distribution of the density of predicted (B) exonic splicing enhancers (ESEs) and (C) U1 spliceosome RNAs (snRNPs) within multi-exonic elincRNAs (red), other expressed lincRNAs (blue), and protein-coding genes (green). **(D)** Distribution of the average processing rates for elincRNAs (red), other expressed lincRNAs (blue), and protein-coding genes (green). **(E)** Distribution of the splicing index, coSI (θ) for multi-exonic elincRNAs (red), other expressed lincRNAs (blue), and protein-coding genes (green). Differences between groups were tested using a two-tailed Mann–Whitney U test. $*P < 0.05$; $***P < 0.001$; NS $P > 0.05$.

RNAPII C-terminal domain, respectively, further supporting their high transcription initiation (Ho & Shuman, 1999), efficient transcription elongation, and co-transcriptional splicing (Komarnitsky et al, 2000; Gu et al, 2013).

Discussion

Although most active enhancers show no preference in the direction of transcription initiation or elongation and produce short and unstable eRNAs bidirectionally (Andersson et al, 2014), a fraction is expressed predominantly in one direction and give rise to elincRNAs that can be spliced (Marques et al, 2013; Hon et al, 2017). Whether differences in the directionality and transcript structure of enhancer-associated transcription underlie differences in enhancer activity remains unknown. Here, we address this question and provide evidence that enhancer-associated transcript splicing directly impact cognate enhancer function. Specifically, we found that elincRNAs, particularly those that undergo splicing, are transcribed from enhancers whose activity was conserved during mammalian evolution and are highly active. The association between elincRNA splicing and cognate enhancer activity is supported by their enrichment in enhancer epigenetic signatures; greater fold increase in putative *cis*-target expression; and the modulation of local chromosomal architecture. Our results in mouse are also consistent with recent work in human cells, which also supports that multi-exonic lincRNAs are often

transcribed from highly active enhancers (Gil & Ulitsky, 2018). Given the paucity of evidence supporting a sequence-dependent mechanism for most elincRNAs and their poor exonic nucleotide conservation, unexpectedly, we found splicing of elincRNAs is efficient.

The coupling between splicing and transcription at multi-exonic elincRNAs, particularly those at promoter-proximal exons, is also consistent with the well-established synergy between splicing and transcription (Furger et al, 2002; Damgaard et al, 2008). Our results expand on these earlier findings and reveal a novel link between elincRNA splicing and enhancer activity that in turn impact target expression. We propose that higher enhancer transcription facilitates the binding of molecular factors, such as CREBBP, the Cohesin and Mediator complexes, at their cognate enhancers, which were recently shown to induce local chromatin remodeling and conformation in an RNA-dependent manner (Lai et al, 2013; Hsieh et al, 2014; Bose et al, 2017), ultimately leading to the stronger enhancer activity observed at these loci (Fig 6).

We further propose that some enhancers associated with eRNA transcription (Andersson et al, 2014), which generally turn over rapidly during mammalian evolution (Villar et al, 2015), have evolved molecular features, including splicing that strengthened their transcription and led to increased cognate enhancer activity by facilitating the recruitment of enhancer factors in a RNA-dependent manner (Fig 6). This is in concordance with evidence that novel exon-containing transcript isoforms show increased expression (Merkin et al, 2015) and that the acquisition
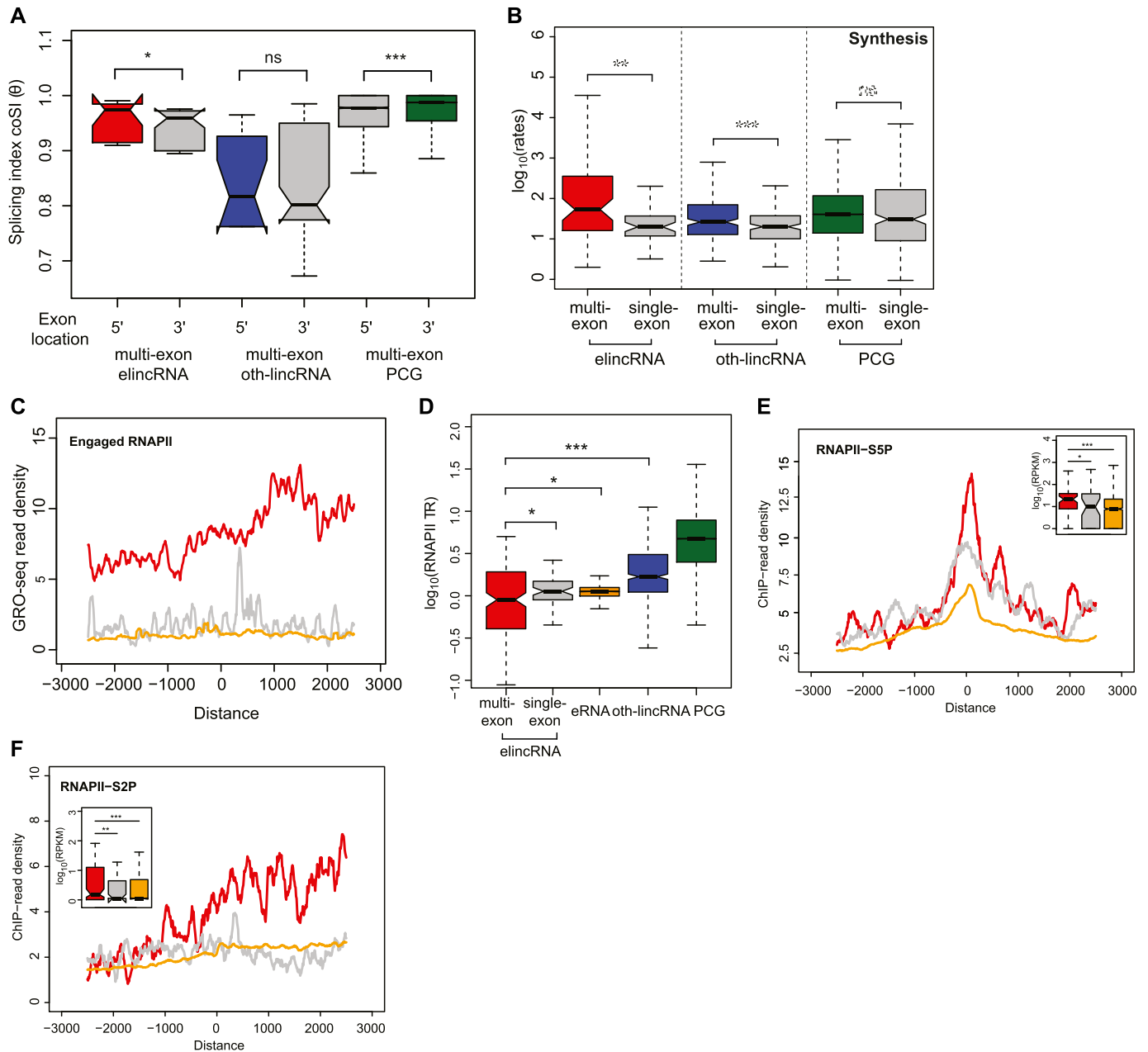
Figure 5. elincRNA 5' end exon splicing associates with increased transcription.
(A) Distribution of the splicing index, coSI (θ) of introns located at the 5' or 3' ends of multi-exonic elincRNAs (red), other expressed lincRNAs (blue), and protein-coding genes (green). **(B)** Distribution of the RNA synthesis rates of multi-exonic elincRNAs (red), other expressed lincRNAs (blue), and protein-coding genes (green), as well as their single-exonic counterparts (grey). **(C)** Metagene plot of mouse embryonic stem cells GRO-seq reads centered at transcription initiation region of multi-exonic (red) and single-exonic (grey) elincRNAs and eRNAs (yellow). **(D)** Distribution of RNAPII travelling ratio (TR) for multi-exonic (red) and single-exonic (grey) elincRNAs, eRNAs (yellow), other expressed lincRNAs (blue), and protein-coding genes (green). **(E, F)** Metagene plots and distribution (figure insets) of ChIP-seq reads for RNAPII with (E) phosphorylated serine 5 (S5P) and (F) phosphorylated serine 2 (S2P) at their C-terminal domain centered at transcription initiation regions of multi-exonic (red) and single-exonic (grey) elincRNAs and eRNAs (yellow). Differences between groups were tested using a two-tailed Mann–Whitney U test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS $P > 0.05$.

of splicing and polyadenylation signals at newly evolved transcriptional initiation sites, which are intrinsically bidirectional (Jin et al, 2017), can favor the preservation of the preferred transcription direction (Almada et al, 2013; Carelli et al, 2018).

Further work is now required to establish the mechanisms underlying the evolution of efficient splicing of elincRNAs and how the processing of these transcripts facilitates recruitment of enhancer factors. Furthermore, inhibition or enhancement of splicing can be achieved through targeted approaches, such as using small molecules or antisense oligos (Spitali & Aartsma-Rus, 2012). Our results open new avenues for modulating enhancer activity through targeting elincRNA processing.
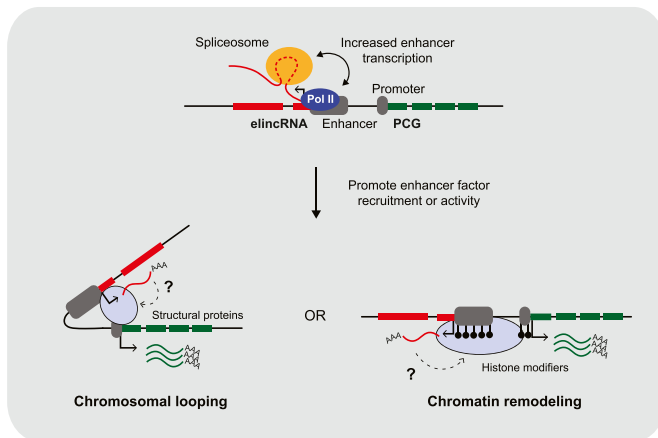
Figure 6. Proposed model of how elincRNA splicing strengthens enhancer activity through chromatin remodeling.
(Top panel) Enhancers (large grey box) can be transcribed by RNA Polymerase II (Pol II, blue circle) and give rise to multi-exonic elincRNA (red boxes) transcripts (red line) whose introns (dashed red line) are co-transcriptionally spliced (spliceosome, yellow circle). The synergistic interaction between elincRNA splicing and Pol II activity increases enhancer transcription, which in turn strengthens *cis*-regulation of nearby protein-coding gene targets (PCG, green, promoter as small grey box). (Bottom panel) Increased elincRNA transcription promotes RNA-dependent recruitment or activity of enhancer factors, for example: (left) structural proteins (blue shaded circle) or and (right) histone modifiers (blue shaded oval). The mechanism by which multi-exonic elincRNAs interact with enhancer factors remains unknown (question mark).

Materials and Methods

Identification of enhancer-associated transcripts

We considered mESC ENCODE intergenic enhancers (61,877 mESCs enhancers) (Bogu et al, 2015) to be transcribed if they overlapped DNase I–hypersensitive sites (Mouse ENCODE Consortium et al, 2012) and a cap anaysis gene expression (CAGE) cluster (Fraser et al, 2015) in the corresponding cell type (n = 2,217). We considered all mESC-expressed lincRNAs (Tan et al, 2015) and Ensembl-annotated protein-coding genes (version 70) with at least one CAGE read overlapping (by > 1 nucleotide) their first exon and an mESC CAGE cluster on the same strand. One hundred transcribed enhancers overlapped lincRNA CAGE clusters (Table S1). The remaining CAGE clusters were TIRs associated with 13,143 protein-coding genes and 317 other non–enhancer-associated mESC-expressed lincRNAs (oth-lincRNAs).

Metagene profiles of CAGE reads centered at mESC enhancers and gene TIRs were plotted using NGSplot (Shen et al, 2014). Sense and antisense reads denote those that map to the same or opposite strand, respectively, as the direction of their cognate CAGE clusters. For eRNAs, direction is defined as the direction with the highest number of CAGE clusters. In cases of equal CAGE clusters on either direction, enhancer direction is randomly assigned.

eRNAs were included only in analyses that do not require transcript models because eRNAs, by definition, are non-polyadenylated, unspliced, and shorted-lived (Kim et al, 2010).

We annotated a larger set of elincRNAs using a more permissive criterion by considering all mESC lincRNAs whose 5′ end is within 500 bp of an enhancer to be enhancer-derived. Using this approach, we identified 211 multi-exonic and 1,772 single-exonic elincRNAs. Corresponding

figures for the analysis of this more comprehensive yet less stringent set of elincRNAs can be found in the Supplemental Data 1.

Metagene analysis of binding enrichment at elincRNAs

Enrichment of histone modifications, transcription factor binding, and gene expression levels were assessed using publicly available mESC ChIP-seq and RNA-seq data sets. Downloaded data sets are listed in Table S4.

For all downloaded data sets, adaptor sequences were first removed from sequencing reads with Trimmomatic (version 0.33) (Bolger et al, 2014) and then aligned to the mouse reference genome (mm9) using HISAT2 (version 2.0.2) (Kim et al, 2015).

Metagene profiles of sequencing reads centered at gene TIRs were visualized using HOMER v4.7 (Heinz et al, 2010).

Analysis of preferential location and chromosomal contact within TADs

mESC TADs (Fraser et al, 2015) were divided into five equal size segments where the two most external bins on either side of the TAD were considered as TAD boundaries and the middle bin as the center of TAD. Enrichment or depletion of enhancer-associated transcripts was estimated for each TAD region, relative to the expectation, using the Genome Association Tester (Heger et al, 2013). Specifically, TAD positional enrichment was compared with a null distribution obtained by randomly sampling 10,000 times (with replacement) the segments of the same length and matching the GC content as the tested loci within mappable intergenic regions of TADs (as predicted by ENCODE [Hoffman et al, 2013]). To control for potential confounding variables that correlate with the GC content, such as gene density, the genome was divided into segments of 10 kb and assigned to eight isochore bins in the enrichment analysis. The frequency of chromosomal interactions within TADs was calculated using mESCs Hi-C contact matrices (Fraser et al, 2015), as previously described (Tan et al, 2017).

Enhancer activity across embryonic neurogenesis

Level of gene transcription initiation (CAGE-based TPM (transcripts per kilobase million) at TIRs) at each of the three stages of neuronal differentiation (mESC to NPC to neuron) was downloaded from Fraser et al (2015). Each locus was paired with its genomically closest protein-coding gene, considered here as its putative *cis*-target. Only pairs where both loci were expressed in at least one embryonic neurogenesis stage were considered. For each gene, the two stages where the locus of interest was most highly or lowly expressed were determined and used to calculate the fold difference between the expression difference of its putative *cis*-target, as described previously (Marques et al, 2013).

Prediction of enriched transcription factor motifs at mESC enhancers

We predicted DNA motifs for transcription factors enriched at multi-exonic elincRNA TIRs (±500 bp from the center of TIRs) relative to those that transcribe single-exonic elincRNAs and eRNAs. Enrichment of motifs

of at least 8mer was predicted using FIMO (Grant et al, 2011). Enriched motifs matching with known transcription factor–binding sites (JASPAR 2016 [Mathelier et al, 2016]) were predicted using TOMTOM (Gupta et al, 2007) with default parameters.

Expression conservation analysis

Syntenic regions of mESC (mm9) genetic elements in human (hg19) were determined using liftOver with the following parameters: -minMatch = 0.2, -minBlocks = 0.01 (Meyer et al, 2013). Regions within the ENCODE Data Analysis Consortium Blacklisted Regions (Hoffman et al, 2013) were excluded from this analysis.

We considered all transcribed mESC ENCODE intergenic enhancers (Bogu et al, 2015) to be conserved in enhancer activity if their syntenic region overlaps human ESC H1 (hESC) ENCODE enhancers (Bogu et al, 2015) by one or more base pairs. Conservation of elincRNA transcription and splicing at syntenic mESC enhancers in humans was assessed using hESC CAGE (Hon et al, 2017) and PolyA-selected RNA-seq (The ENCODE Project Consortium, 2012) data. Conserved hESC enhancers that overlapped an hESC CAGE cluster and RNA-seq reads were considered to be conserved in transcription. Those that overlapped RNA-seq reads that span across exon–intron junctions were considered to be conserved in splicing.

4sU metabolic labeling of mESCs and RNA extraction

Mouse DTCM23/49 XY mESCs were cultured at 37°C with 5% CO_2 in Knockout DMEM (#10829-018; Thermo Fisher Scientific) supplemented with 15% FBS (#16000-044; Thermo Fisher Scientific), 1% antibiotic penicillin/streptomycin (15070063; Thermo Fisher Scientific), 0.01% recombinant mouse leukemia inhibitory factor protein (#ESG1107; Merck), and 0.06 mM 2-mercaptoethanol (#31350-010; Thermo Fisher Scientific), on 0.1% gelatin-coated cell culture dishes. When confluent, the culture was divided into two and passaged eight times. Five million mESCs of two biological replicates were seeded and allowed to grow to 70–80% confluency (~1 d). RNA was labeled with 4sU (T4509; Sigma-Aldrich) and nascent RNA was isolated after the general procedure as previously described (Dolken et al, 2008). Specifically, 4sU was added to the growth medium (final concentration of 200 μM), and the cells were incubated at 37°C for 15, 30, or 60 min. The plates were washed once with 1× PBS and RNA was extracted using TRIzol (#15596-026; Thermo Fisher Scientific). 100 μg of extracted RNA was incubated for 2 h at room temperature with rotation in 1/10 volume of 10× biotinylation buffer (Tris–HCl pH 7.4, 10 mM EDTA) and 2/10 volume of biotin-HPDP (1 mg/ml in dimethylformamide [#21341; Thermo Fisher Scientific]). RNA was extracted using phenol:chloroform:isoamyl alcohol (P3803-400ML; Sigma-Aldrich). Equal volume of biotinylated RNA and prewashed Dynabeads MyOne Streptavidin T1 beads (#65601; Thermo Fischer Scientific) was added to 2× B&W buffer (10 mM Tris–HCl, pH 7.5, 1 mM EDTA, and 2M NaCl [#65601; Thermo Fisher Scientific]) and incubated at room temperature for 15 min under rotation. The beads were then separated from the mixture using DynaMag-2 Magnet (#12321D; Thermo Fisher Scientific). After removing the supernatant, the beads were washed with 1× B&W three times. Biotinylated RNA was recovered from the supernatant after 1 min of incubation with RLT buffer (RNeasy kit, #74104; QIAGEN) and

purified using the RNeasy kit according to the manufacturer's instructions.

RNA sequencing, mapping, and quantification of metabolic rates

Total RNA libraries were prepared from 10 ng of DNase-treated total and newly transcribed RNA using Ovation RNA-seq and sequenced on Illumina HiSeq 2500 (average of 50 million reads per library).

Hundred-nucleotide-long single-end reads were first mapped to mouse ribosomal RNA (rRNA) sequences with STAR v2.5.0 (Dobin et al, 2013). On average, 20% of reads were mapped to rRNA reads. Reads that do not map to rRNA (36 million on average) were then aligned to intronic and exonic sequences using STAR and quantified using RSEM (Li & Dewey, 2011). Principal component analysis of read counts was performed to demonstrate separation between newly transcribed (labeled) and total RNA (Fig S1D). Rates of synthesis, processing, and degradation were independently inferred using biological duplicates at each labeling points using the INSPEcT Bioconductor package v1.8.0 (de Pretis et al, 2015). Biotype differences in the average rate across the three labeling times were used in the analyses (Table S2).

GC composition

Only mESC genes with multi-exonic transcripts (two or more exons) were considered for this analysis. We computed GC content separately for the first and all remaining exons, as well as the introns, for each gene and their flanking intergenic sequences of the same length, after excluding the 500 nucleotides immediately adjacent to annotations, as previously described (Haerty & Ponting, 2015).

Identification of splicing-associated motifs

We predicted the density of mouse exonic splicing enhancer motifs (identified in Fairbrother et al (2002)) within mESC transcripts, as described previously (Haerty & Ponting, 2015). Exonic nucleotides (50 nt) flanking the SSs of internal transcript exons (>100 nt) were considered in the analysis, after masking the 5 nt immediately adjacent to SS to avoid SS-associated nucleotide composition bias (Fairbrother et al, 2002; Yeo & Burge, 2004). Canonical U1 sites (GGUAAG, GGUGAG, and GUGAGU) adjacent to 5′ SSs (three exonic nt and six intronic nt flanking the 5′ SS) were predicted as previously described (Almada et al, 2013). FIMO (Grant et al, 2011) was used to search for perfect hexamer matches within these sequences. For each exon, we estimated the SS strength using MaxENT (Yeo & Burge, 2004). SS scores were calculated using the −3 exonic nt to +6 intronic nt and −20 intronic nt to +3 exonic nt flanking the 5′ SS and 3′ SS, respectively.

Splicing efficiency

The efficiency of splicing was assessed by estimating the fraction of transcripts for each gene where its introns were fully excised using bam2ssj (Pervouchine et al, 2013). The splicing index, coSI (θ), represents the ratio of total RNA-seq reads spanning exon–exon splice junctions (excised intron) over those that overlap exon–intron junctions (incomplete excision) (Tilgner et al, 2012).

RNAPII stalling

Distribution of RNAPII across the gene TIR and body, commonly used as an indicator of promoter-proximal RNAPII stalling and efficient transcription elongation, was estimated by calculating the travelling ratio and by using mESC RNAPII ChIP-seq data (Brookes et al, 2012). The travelling ratio represents relative read density at gene TIRs divided by that across the gene body (Reppas et al, 2006).

Statistical tests

All statistical analyses were performed using the R software environment for statistical computing and graphics (R Development Core Team, 2008).

Data access

The raw and processed 4sU sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus under accession number GSE111951. Most analyses were performed using standard publicly available command-line tools, as detailed in the Materials and Methods section.

Supplementary Information

Supplementary Information is available at https://doi.org/10.26508/lsa.202000663.

Acknowledgements

Author Contributions

JY Tan: conceptualization, data curation, formal analysis, supervision, funding acquisition, investigation, methodology, project administration, and writing—original draft, review, and editing.
A Biasini: conceptualization, data curation, formal analysis, investigation, methodology, and writing—original draft, review, and editing.
RS Young: formal analysis and writing—review and editing.
AC Marques: conceptualization, formal analysis, supervision, funding acquisition, project administration, and writing—original draft, review, and editing.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

References

Alexanian M, Maric D, Jenkinson SP, Mina M, Friedman CE, Ting CC, Micheletti R, Plaisance I, Nemir M, Maison D, et al (2017) A transcribed enhancer dictates mesendoderm specification in pluripotency. *Nat Commun* 8: 1806. doi:10.1038/s41467-017-01804-w

Almada AE, Wu XB, Kriz AJ, Burge CB, Sharp PA (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499: 360–363. doi:10.1038/nature12349

Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, Lev-Maor G, Burstein D, Schwartz S, Postolsky B, et al (2012) Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep* 1: 543–556. doi:10.1016/j.celrep.2012.03.013

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507: 455–461. doi:10.1038/nature12787

Banerji J, Rusconi S, Schaffner W (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27: 299–308. doi:10.1016/0092-8674(81)90413-x

Bedford DC, Kasper LH, Fukuyama T, Brindle PK (2010) Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. *Epigenetics* 5: 9–15. doi:10.4161/epi.5.1.10449

Bogu GK, Vizan P, Stanton LW, Beato M, Di Croce L, Marti-Renom MA (2015) Chromatin and RNA maps reveal regulatory long noncoding RNAs in mouse. *Mol Cell Biol* 36: 809–819. doi:10.1128/mcb.00955-15

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. doi:10.1093/bioinformatics/btu170

Bonev B, Cavalli G (2016) Organization and function of the 3D genome. *Nat Rev Genet* 17: 661–678. doi:10.1038/nrg.2016.112

Bose DA, Donahue G, Reinberg D, Shiekhattar R, Bonasio R, Berger SL (2017) RNA binding to CBP stimulates histone acetylation and transcription. *Cell* 168: 135–149.e22. doi:10.1016/j.cell.2016.12.020

Brinster RL, Allen JM, Behringer RR, Gelinas RE, Palmiter RD (1988) Introns increase transcriptional efficiency in transgenic mice. *Proc Natl Acad Sci U S A* 85: 836–840. doi:10.1073/pnas.85.3.836

Brookes E, de Santiago I, Hebenstreit D, Morris KJ, Carroll T, Xie SQ, Stock JK, Heidemann M, Eick D, Nozaki N, et al (2012) Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell* 10: 157–170. doi:10.1016/j.stem.2011.12.017

Carelli FN, Liechti A, Halbert J, Warnefors M, Kaessmann H (2018) Repurposing of promoters and enhancers during mammalian evolution. *Nat Commun* 9: 4066. doi:10.1038/s41467-018-06544-z

Damgaard CK, Kahns S, Lykke-Andersen S, Nielsen AL, Jensen TH, Kjems J (2008) A 5′ splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell* 29: 271–278. doi:10.1016/j.molcel.2007.11.035

de Pretis S, Kress T, Morelli MJ, Melloni GE, Riva L, Amati B, Pelizzola M (2015) INSPEcT: A computational tool to infer mRNA synthesis, processing and degradation dynamics from RNA- and 4sU-seq time course experiments. *Bioinformatics* 31: 2829–2835. doi:10.1093/bioinformatics/btv288

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G (2010) A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* 8: e1000384. doi:10.1371/journal.pbio.1000384

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21. doi:10.1093/bioinformatics/bts635

Dolken L, Ruzsics Z, Radle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P, et al (2008) High-resolution gene

expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* 14: 1959–1972. doi:10.1261/rna.1136108

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539: 452–455. doi:10.1038/nature20149

Fairbrother WG, Yeh RF, Sharp PA, Burge CB (2002) Predictive identification of exonic splicing enhancers in human genes. *Science* 297: 1007–1013. doi:10.1126/science.1073774

Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, Barbieri M, Moore BL, Kraemer DC, Aitken S, et al (2015) Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol Syst Biol* 11: 852. doi:10.15252/msb.20156492

Furger A, O'Sullivan JM, Binnie A, Lee BA, Proudfoot NJ (2002) Promoter proximal splice sites enhance transcription. *Genes Dev* 16: 2792–2799. doi:10.1101/gad.983602

Gil N, Ulitsky I (2018) Production of spliced long noncoding RNAs specifies regions with increased enhancer activity. *Cell Syst* 7: 537–547.e3. doi:10.1016/j.cels.2018.10.009

Grant CE, Bailey TL, Noble WS (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27: 1017–1018. doi:10.1093/bioinformatics/btr064

Gu B, Eick D, Bensaude O (2013) CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res* 41: 1591–1603. doi:10.1093/nar/gks1327

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS (2007) Quantifying similarity between motifs. *Genome Biol* 8: R24. doi:10.1186/gb-2007-8-2-r24

Haerty W, Ponting CP (2015) Unexpected selection to retain high GC content and splicing enhancers within exons of multiexonic lncRNA loci. *RNA* 21: 333–346. doi:10.1261/rna.047324.114

Heger A, Webber C, Goodson M, Ponting CP, Lunter G (2013) GAT: A simulation framework for testing the association of genomic intervals. *Bioinformatics* 29: 2046–2048. doi:10.1093/bioinformatics/btt343

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576–589. doi:10.1016/j.molcel.2010.05.004

Ho CK, Shuman S (1999) Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* 3: 405–411. doi:10.1016/s1097-2765(00)80468-2

Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS (2012) Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* 9: 473–476. doi:10.1038/nmeth.1937

Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen PM, Bilmes JA, Birney E, et al (2013) Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res* 41: 827–841. doi:10.1093/nar/gks1284

Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J, et al (2017) An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* 543: 199–204. doi:10.1038/nature21374

Hsieh CL, Fei T, Chen Y, Li T, Gao Y, Wang X, Sun T, Sweeney CJ, Lee GS, Chen S, et al (2014) Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci U S A* 111: 7319–7324. doi:10.1073/pnas.1324151111

Jin Y, Eser U, Struhl K, Churchman LS (2017) The ground state and evolution of promoter region directionality. *Cell* 170: 889–898.e10. doi:10.1016/j.cell.2017.07.006

Kim D, Langmead B, Salzberg SL (2015) HISAT: A fast spliced aligner with low memory requirements. *Nat Methods* 12: 357–360. doi:10.1038/nmeth.3317

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465: 182–187. doi:10.1038/nature09033

Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la Chapelle AL, Heidemann M, Hintermair C, et al (2011) Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* 18: 956–963. doi:10.1038/nsmb.2085

Komarnitsky P, Cho EJ, Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* 14: 2452–2460. doi:10.1101/gad.824700

Kowalczyk MS, Hughes JR, Garrick D, Lynch MD, Sharpe JA, Sloane-Stanley JA, McGowan SJ, De Gobbi M, Hosseini M, Vernimmen D, et al (2012) Intragenic enhancers act as alternative promoters. *Mol Cell* 45: 447–458. doi:10.1016/j.molcel.2011.12.021

Krchnakova Z, Thakur PK, Krausova M, Bieberstein N, Haberman N, Muller-McNicoll M, Stanek D (2019) Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5' splice-site sequences due to weak interactions with SR proteins. *Nucleic Acids Res* 47: 911–928. doi:10.1093/nar/gky1147

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R (2013) Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494: 497–501. doi:10.1038/nature11884

Le Hir H, Nott A, Moore MJ (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* 28: 215–220. doi:10.1016/s0968-0004(03)00052-5

Li B, Dewey CN (2011) RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC bioinformatics* 12: 323. doi:10.1186/1471-2105-12-323

Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, et al (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498: 516–520. doi:10.1038/nature12210

Li W, Notani D, Rosenfeld MG (2016) Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. *Nat Rev Genet* 17: 207–223. doi:10.1038/nrg.2016.4

Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161: 1012–1025. doi:10.1016/j.cell.2015.04.004

Marques AC, Hughes J, Graham B, Kowalczyk MS, Higgs DR, Ponting CP (2013) Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biol* 14: R131. doi:10.1186/gb-2013-14-11-r131

Maruyama A, Mimura J, Itoh K (2014) Non-coding RNA derived from the region adjacent to the human HO-1 E2 enhancer selectively regulates HO-1 gene induction by modulating Pol II binding. *Nucleic Acids Res* 42: 13599–13614. doi:10.1093/nar/gku1169

Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, et al (2016) JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 44: D110–D115. doi:10.1093/nar/gkv1176

Mele M, Mattioli K, Mallard W, Shechner DM, Gerhardinger C, Rinn JL (2017) Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res* 27: 27–37. doi:10.1101/gr.214205.116

Merika M, Williams AJ, Chen G, Collins T, Thanos D (1998) Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Mol Cell* 1: 277–287. doi:10.1016/s1097-2765(00)80028-3

Merkin JJ, Chen P, Alexis MS, Hautaniemi SK, Burge CB (2015) Origins and impacts of new mammalian exons. *Cell Rep* 10: 1992–2005. doi:10.1016/j.celrep.2015.02.058

Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, et al (2013) The UCSC genome

browser database: Extensions and updates 2013. *Nucleic Acids Res* 41: D64–D69. doi:10.1093/nar/gks1048

Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P (1981) The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* 9: 6047–6068. doi:10.1093/nar/9.22.6047

Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager GL, Sartorelli V (2013) eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* 51: 606–617. doi:10.1016/j.molcel.2013.07.022

Mouse ENCODE Consortium, Stamatoyannopoulos JA, Snyder M, Hardison R, Ren B, Gingeras T, Gilbert DM, Groudine M, Bender M, Kaul R, et al (2012) An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* 13: 418. doi:10.1186/gb-2012-13-8-418

Mukherjee N, Calviello L, Hirsekorn A, de Pretis S, Pelizzola M, Ohler U (2017) Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat Struct Mol Biol* 24: 86–96. doi:10.1038/nsmb.3325

Natoli G, Andrau JC (2012) Noncoding transcription at enhancers: General principles and functional models. *Annu Rev Genet* 46: 1–19. doi:10.1146/annurev-genet-110711-155459

Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, 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

Pervouchine DD, Knowles DG, Guigo R (2013) Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics* 29: 273–274. doi:10.1093/bioinformatics/bts678

Reppas NB, Wade JT, Church GM, Struhl K (2006) The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. *Mol Cell* 24: 747–757. doi:10.1016/j.molcel.2006.10.030

Schaukowitch K, Joo JY, Liu X, Watts JK, Martinez C, Kim TK (2014) Enhancer RNA facilitates NELF release from immediate early genes. *Mol Cell* 56: 29–42. doi:10.1016/j.molcel.2014.08.023

Schuler A, Ghanbarian AT, Hurst LD (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. doi:10.1093/molbev/msu249

Shen L, Shao N, Liu X, Nestler E (2014) ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 15: 284. doi:10.1186/1471-2164-15-284

Spitali P, Aartsma-Rus A (2012) Splice modulating therapies for human disease. *Cell* 148: 1085–1088. doi:10.1016/j.cell.2012.02.014

Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F (2014) Functional and topological characteristics of mammalian regulatory domains. *Genome Res* 24: 390–400. doi:10.1101/gr.163519.113

Tan JY, Sirey T, Honti F, Graham B, Piovesan A, Merkenschlager M, Webber C, Ponting CP, Marques AC (2015) Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells. *Genome Res* 25: 655–666. doi:10.1101/gr.181974.114

Tan JY, Smith AAT, Ferreira da Silva M, Matthey-Doret C, Rueedi R, Sonmez R, Ding D, Kutalik Z, Bergmann S, Marques AC (2017) cis-acting complex-trait-associated lincRNA expression correlates with modulation of chromosomal architecture. *Cell Rep* 18: 2280–2288. doi:10.1016/j.celrep.2017.02.009

R Development Core Team (2008) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org

The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74. doi:10.1038/nature11247

Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigo R (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* 22: 1616–1625. doi:10.1101/gr.134445.111

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, et al (2015) Enhancer evolution across 20 mammalian species. *Cell* 160: 554–566. doi:10.1016/j.cell.2015.01.006

Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, et al (2011) Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474: 390–394. doi:10.1038/nature10006

Wu H, Nord AS, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA, Visel A (2014) Tissue-specific RNA expression marks distant-acting developmental enhancers. *PLoS Genet* 10: e1004610. doi:10.1371/journal.pgen.1004610

Xu H, Baroukh C, Dannenfelser R, Chen EY, Tan CM, Kou Y, Kim YE, Lemischka IR, Ma'ayan A (2013) ESCAPE: Database for integrating high-content published data collected from human and mouse embryonic stem cells. *Database* 2013: bat045. doi:10.1093/database/bat045

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457: 1033–1037. doi:10.1038/nature07728

Yeo G, Burge CB (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11: 377–394. doi:10.1089/1066527041410418

Yin Y, Yan P, Lu J, Song G, Zhu Y, Li Z, Zhao Y, Shen B, Huang X, Zhu H, 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. doi:10.1016/j.stem.2015.03.007

Yoo EJ, Cooke NE, Liebhaber SA (2012) An RNA-independent linkage of noncoding transcription to long-range enhancer function. *Mol Cell Biol* 32: 2020–2029. doi:10.1128/mcb.06650-11

Young RS, Kumar Y, Bickmore WA, Taylor MS (2017) Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome Biol* 18: 242. doi:10.1186/s13059-017-1379-8