1 **The activity of human enhancers is modulated by the splicing of their**
2 **associated lncRNAs.**

3 Jennifer Y. Tan¹ and Ana C. Marques¹

4

5 ¹Department of Computational Biology, University of Lausanne, Switzerland

6 Corresponding authors: anaclaudia.marques@unil.ch; jennifer.tan@unil.ch.

7

1 **ABSTRACT**

2 Pervasive enhancer transcription is at the origin of more than half of all long
3 noncoding RNAs in humans. Transcription of enhancer-associated long
4 noncoding RNAs (eIncRNA) contribute to their cognate enhancer activity and
5 gene expression regulation in *cis*. Recently, splicing of eIncRNAs was shown
6 to be associated with elevated enhancer activity. However, whether splicing of
7 eIncRNA transcripts is a mere consequence of accessibility at highly active
8 enhancers or if eIncRNA splicing directly impacts enhancer function, remains
9 unanswered.

10 We analysed genetically driven changes in eIncRNA expression, in humans, to
11 address this outstanding question. We showed that splicing related motifs
12 within multi-exonic eIncRNAs evolved under selective constraints during human
13 evolution, suggesting the processing of these transcripts is unlikely to have
14 resulted from transcription across spurious splice sites. Using a genome-wide
15 and unbiased approach, we used nucleotide variants as independent genetic
16 factors to directly assess the causal relationship that underpin eIncRNA splicing
17 and their cognate enhancer activity. We found that the splicing of most
18 eIncRNAs is associated with changes in chromatin signatures at cognate
19 enhancers and target mRNA expression.

20 We conclude that efficient and conserved processing of enhancer-associated
21 eIncRNAs contributes to enhancer activity.

22

1 INTRODUCTION

2 Classically, enhancers are defined as regulatory DNA elements that positively
3 regulate temporal and spatial expression of their target genes, in *cis*.
4 Transcription activation by enhancers requires transcription factor-dependent
5 recruitment of coactivating complexes and three-dimensional genome
6 rearrangements that bring enhancers into close proximity of their target
7 promoters [reviewed in (Schoenfelder and Fraser 2019)].

8 Whereas enhancer activity, as classically defined, is solely determined by its
9 sequence and genomic location, most, if not all, active enhancers are
10 transcribed (De Santa et al. 2010; Kim et al. 2010; Kowalczyk et al. 2012). Initial
11 evidence of this phenomenon emerged from the characterisation of some of the
12 first identified enhancers (Ashe et al. 1997). Pervasive transcription of active
13 enhancers was later confirmed genome-wide, initially in neurons (Kim et al.
14 2010) and macrophages (De Santa et al. 2010) and more recently across a
15 number of human and mouse cell types (Andersson et al. 2014; Arner et al.
16 2015). These studies revealed that enhancer transcription often precedes
17 target promoter activation (De Santa et al. 2010) and is positively correlated
18 with target gene expression (De Santa et al. 2010; Kim et al. 2010; Arner et al.
19 2015). The majority [>95% (Natoli and Andrau 2012; Gil and Ulitsky 2018; Tan
20 et al. 2020)] of enhancers is transcribed bidirectionally and gives rise to
21 relatively short single exonic RNAs that are non-polyadenylated and unstable,
22 generally referred to as eRNAs (Kim et al. 2010). The remaining enhancers
23 transcribe enhancer-associated long noncoding RNAs [elncRNAs (Marques et
24 al. 2013; Gil and Ulitsky 2018; Tan et al. 2020)] predominantly in one direction
25 (Natoli and Andrau 2012). In contrast with eRNAs, elncRNAs are relatively
26 stable, polyadenylated and can be spliced (Natoli and Andrau 2012).

27 Transcription has been found to strengthen cognate enhancer activity through
28 various non-mutually exclusive mechanisms. For example, enhancer
29 transcription facilitates binding of enhancer factors, such as CREBBP (Bose et
30 al. 2017), or chromatin remodeling complex, including Cohesin and Mediator,

1 that in turn induce RNA-dependent changes in local chromatin conformation
2 (Lai et al. 2013; Mousavi et al. 2013; Hsieh et al. 2014). Enhancer transcription
3 has also been proposed to regulate the load, pause and release of RNA
4 Polymerase II (RNAPII) (Maruyama et al. 2014; Schaukowitch et al. 2014).

5 While the contribution of transcription to enhancer function is by now relatively
6 well established, the importance of the transcribed RNAs has only been
7 demonstrated for a handful of anecdotal transcripts (for example, (Lai et al.
8 2013; Lam et al. 2013; Li et al. 2013; Melo et al. 2013; Iott et al. 2014)). In
9 general, the absence of evolutionary constraint at eRNA/elncRNA exons
10 (Marques et al. 2013) argues that the function of most eRNAs/elncRNAs is
11 unlikely to be encoded within their sequences. Yet a large fraction of elncRNAs
12 is multi-exonic and recently, genome-wide analysis in multiple human cell lines
13 (Gil and Ulitsky 2018) and mouse Embryonic Stem Cells (Tan et al. 2020)
14 revealed that their splicing positively impacts enhancer function. Enhancers
15 that transcribe multi-exonic elncRNAs are enriched in enhancer-specific
16 chromatin signatures, elevated binding of co-transcriptional regulators,
17 increased local intra-chromosomal DNA contacts, and strengthened *cis*-
18 regulation on target gene expression (Gil and Ulitsky 2018; Tan et al. 2020).
19 What remains currently unclear is the causal relationship between elncRNA
20 splicing and enhancer activity. High enhancer activity may promote
21 transcription across splice/donor acceptor site-like sequence, which are
22 common in mammals (Wan and Larson 2018), leading to unregulated/noisy
23 splicing of elncRNAs. Arguing against this is evidence that mutations disrupting
24 elncRNA splicing directly decreased their cognate enhancer activity. For
25 example, removal of the intronic sequence of the locus encoding for the Haunt
26 elncRNA significantly reduced its *cis* regulatory function (Yin et al. 2015),
27 whereas mutations disrupting the splicing of Blustr, another enhancer-
28 associated lncRNA, were sufficient to negatively impact the expression of its
29 *cis* target (Engreitz et al. 2016). Whether the splicing of most elncRNAs directly
30 impacts enhancer activity, as exemplified by Blustr and Haunt, or whether it is
31 an inconsequential result of their cognate enhancer's high activity remains

1 unknown.

2 We used a combination of integrative and population genomics approaches to
 3 address this outstanding question. We showed that splicing related motifs
 4 within elncRNAs evolved under selective constraint, consistent with their
 5 processing being biologically relevant. Using a mediation-based approach, we
 6 performed causal inference testing using genetic variants as independent
 7 factors to test the underlying causal relationship between elncRNA splicing and
 8 their cognate enhancer function. We showed that splicing of most elncRNAs
 9 causally contribute to target gene regulation. Furthermore, nucleotide variants
 10 that disrupt elncRNA splicing significantly reduced cognate enhancer activity,
 11 supporting the functional importance of elncRNA splicing.

12

13

14

1 RESULTS

2 Splicing of eIncRNAs is associated with higher enhancer activity.

3 We integrated publicly available RNA sequencing data for human
4 lymphoblastoid cell lines [LCL (Lappalainen et al. 2013)] with enhancer
5 annotations from the ENCODE consortium (Hoffman et al. 2013) to distinguish
6 between predominantly bidirectionally (eRNAs, n=4433) and unidirectionally
7 transcribed (eIncRNAs, n=564) enhancers active in LCLs (Supplementary
8 Table ST1, Methods). The transcription profiles of eIncRNAs resemble that of
9 promoter-associated lncRNAs (plncRNAs, n=600) and protein-coding genes
10 (n=12,070, Supplementary Figure S1A). In LCLs, most (n=352, 62.4%)
11 eIncRNAs are multi-exonic, consistent with previous analysis in other human
12 cell lines (Gil and Ulitsky 2018).

13 As described for other human and mouse cells, enhancers that transcribe multi-
14 exonic eIncRNAs are more active (Gil and Ulitsky 2018; Tan et al. 2020). In
15 particular, compared to enhancers that give rise to either single-exonic
16 eIncRNAs or eRNAs, those transcribing multi-exonic eIncRNAs have higher
17 levels of chromatin modifications associated with enhancer function (Figure 1A-
18 C), including mono-methylation of Histone 3 Lysine 4 (H3K4me1, Figure 1A),
19 acetylation of Histone 3 lysine 27 (H3K27ac, Figure 1B) and DNase I
20 accessibility (DHSI, Figure 1C). Multi-exonic eIncRNA-transcribing enhancers
21 are also enriched for binding of histone acetyl transferase (HAT) P300 (Figure
22 1D), whose recruitment also distinctly marks active enhancers (Merika et al.
23 1998).

24 Consistent with previous analysis (Gil and Ulitsky 2018; Tan et al. 2020),
25 relative to all other transcribed enhancers, multi-exonic eIncRNA enhancers are
26 enriched in CTCF binding (Figure 1E), significantly enriched at LCL promoter-
27 enhancer loop domains (fold enrichment=1.3, FDR=7x10⁻³, Methods) and their
28 topologically associating domains (TAD) display a significantly higher
29 frequency of intra-TAD DNA-DNA contacts (p<0.02, two-tailed Mann-

1 Whitney U test, Figure 1F).

Figure 1. Multi-exonic elncRNAs are transcribed from highly active enhancers. Metagen plots of (A) H3K4me1, (B) H3K27ac, (C) DNase I hypersensitive sites (DHSI), (D) P300 and (E) CTCF ChIP-seq reads in LCLs at promoters of multi-exonic (red) and single-exonic (grey) elncRNAs, and eRNAs (yellow). (F) Distribution of the average amount of DNA:DNA contacts within LCL TADs containing multi-exonic (red) and single-exonic (grey) elncRNAs and eRNAs (yellow). Differences between groups were tested using a two-tailed Mann-Whitney U test. * $p < 0.05$; *** $p < 0.001$.

In summary, our analysis in LCLs supports that enhancers transcribing multi-exonic elncRNAs are more active, as previously described in other cell types (Gil and Ulitsky 2018; Tan et al. 2020).

6 elncRNA splicing was preserved by purifying selection

The splicing of elncRNAs may be a regulated and conserved process. Alternatively, it can be a mere by-product of inconsequential transcription across splice-site donor/acceptors favoured by increased accessibility at their highly active cognate enhancers. To distinguish between these two possibilities, we first analysed splicing related motifs of elncRNAs. As previously described (Schuler et al. 2014; Haerty and Ponting 2015; Tan et al. 2020) the GC contents of exons and introns (Amit et al. 2012), a determinant of efficient splice-site

1 recognition, of multi-exonic elncRNAs is similar to that of protein-coding genes
2 and promoter-associated plncRNAs, and as observed in mESCs (Tan et al.
3 2020) (Supplementary Figure S2A, two-tailed Mann-Whitney test). Compared
4 to plncRNAs, regions flanking elncRNA splice-sites are enriched in splicing-
5 associated elements, including U1 snRNP binding motifs (Supplementary
6 Figure S2B, $p=0.05$) and a comprehensive (Supplementary Figure S2C, $p=0.04$)
7 (Fairbrother et al. 2002) and a stringent set (Supplementary Figure S2D, $p=0.05$,
8 two-tailed Mann-Whitney test)(Caceres and Hurst 2013) of exonic splicing
9 enhancers (ESEs). This increase in density of splicing-related motifs in
10 elncRNAs support their increased splicing efficiency relative to plncRNAs
11 (Supplementary Figure S2E, $p=5 \times 10^{-11}$, two-tailed Mann-Whitney U test),
12 consistent with what was previously reported in mESCs (Tan et al. 2020).

13 If splicing of elncRNAs is important for enhancer function, one would expect
14 selection to have prevented the accumulation of deleterious mutations in their
15 splicing-associated motifs during evolution. We estimated the nucleotide
16 substitution rate, between human and mouse, for elncRNA splicing-related
17 sequence motifs and compared it to that of neutrally evolving neighbouring
18 ancestral repeats (ARs, Figure 2A). We found that splice donor/acceptor sites,
19 as well as U1 and ESE sites, evolved under significant selective constraint by
20 accumulating fewer substitutions during mammalian evolution than neutrally
21 evolving ARs ($p<0.002$, permutation test, Figure 2A). Consistent with the
22 observed differences in splicing efficiency, we found that splicing-associated
23 motifs within elncRNAs (including SS, U1 and ESEs, Figure 2B-D), evolved
24 significantly slower ($p<2.2 \times 10^{-16}$, two-tailed Mann-Whitney U test) than those
25 within plncRNAs.

26 We next turned our attention to the evolution of elncRNA splicing during recent
27 human history. Relative to neutrally evolving ARs and plncRNAs, single
28 nucleotide polymorphisms (SNPs) within elncRNA splice sites and splicing
29 motifs were enriched in alleles with low derived allele frequency (DAF < 0.1)
30 ($p<0.05$, two-tailed Fisher's exact test, Figure 2E), consistent with their

Figure 2. Splicing of elncRNAs has evolved under purifying selection. (A) Distribution of pairwise nucleotide substitution rate between human and mouse for 1,000 randomly subsampled sets of local ancestral repeats (ARs) with matching GC-content and size as splicing related motifs (grey). The observed rate observed at splice sites (d_{SS} , solid line arrow), ESE (d_{ESE} , dash line arrow), and U1 (d_{U1} , dotted line arrow) sites within multi-exonic elncRNAs. Distribution of nucleotide substitution rate relative to that of randomly subsampled local ARs (d_{AR}) for (B) splice sites, (C) ESEs, and (D) U1s within multi-exonic elncRNA (red), plncRNA (blue) and protein-coding gene (dark green). Differences between groups were tested using a two-tailed Mann-Whitney U test. *** $p < 0.001$. (E) Distribution of derived allele frequency (DAF) for single nucleotide variants at splicing motifs (splice sites, ESEs and U1s) within multi-exonic elncRNAs (red), plncRNAs (blue), protein-coding genes (dark green), and ARs (black). The insert illustrates variants with low derived allele frequency (DAF<0.1). Differences between groups were tested using a two-tailed Fisher's exact test. * $p < 0.05$; *** $p < 0.001$; NS $p > 0.05$.

1 evolution under selective constraint. Whereas, the frequency of SNPs with
2 DAF<0.1 that overlap splicing-associated sequences within elncRNAs is
3 statistically indistinguishable from that of protein-coding genes ($p=0.93$, two-
4 tailed Fisher's exact test, Figure 2E), we cannot exclude this is in part a
5 consequence of the lack of power to identify differences. Evidence that
6 selection has purged deleterious mutations at elncRNA splicing-associated
7 motifs during mammalian and recent human history, supports the functional
8 relevance of elncRNA processing.

1 **Disruption of elncRNA splicing decreases target expression**

2 We used genotyping data (The 1000 Genomes Project Consortium 2012) and
3 gene expression data (Lappalainen et al., 2013) from LCLs derived from 373
4 healthy individuals to identify nucleotide variants that disrupt elncRNA splicing
5 (Figure 3A). This unbiased approach allowed us to directly test whether
6 changes in elncRNA splicing impact enhancer activity.

7 We identified 4 variants that disrupt the splice-donor or acceptor sites of 4
8 elncRNAs that have at least one predicted *cis*-target (Supplementary Table ST2,
9 Methods). Since enhancer-promoter interactions rely on local chromosomal
10 looping, enhancers and their associated transcripts often do not target the
11 nearest gene in linear distance (Javierre et al. 2016). We unbiasedly predicted
12 putative elncRNA targets as those that are jointly associated to the same
13 expression quantitative trait loci (eQTL) variants (Methods), supporting their
14 potential co-regulation (Tong et al. 2017).

15 We used an approach that explores spliced reads (intronic spanning reads) to
16 quantify differential abundance of spliced isoforms in individuals with reference
17 and alternative alleles at splice sites (Li et al. 2018). Individuals that carry
18 nucleotide variants disrupting canonical splice donor/acceptor site have
19 significantly decreased transcript splicing efficiency relative to individuals with
20 the reference canonical splice sites allele (GT-AG) (Figure 3B, Supplementary
21 Figure S3). As expected, the impact of SS variants on gene splicing efficiency
22 depends on the total number of alternative transcripts and exons and ranges
23 from 11% to 24% for elncRNA with 6 to 2 number of exons, respectively
24 (Supplementary Figure S3). Consistent with a direct role of splicing in the
25 modulation of enhancer function, this natural mutational study revealed that
26 disruption of elncRNA splicing, which does not impact their protein-coding
27 target splicing ($p > 0.3$, two-tailed Mann-Whitney U test, Figure 3C,
28 Supplementary Figure S3), is associated with significant decrease in target
29 expression ($p < 0.05$, two-tailed Mann-Whitney U test, Figure 3D,
30 Supplementary Figure S3). The association between of splice site mutations

1 and gene expression was restricted to predicted elncRNA targets genes, as the
2 expression levels of other nearby was unaffected ($p > 0.1$, two-tailed Mann-
3 Whitney U test, Figure 3E, Supplementary Figure S3). In addition to changes
4 in target protein-coding gene levels, the relative abundance of elncRNAs was
5 also significantly decreased in individuals that carried variants that altered
6 elncRNA splice donor or splice acceptor sites (Figure 3F, Supplementary
7 Figure S3). This is consistent with the well-established synergy between RNA
8 processing and transcription and the impact of co-transcriptional splicing on
9 gene expression (Furger et al. 2002; Damgaard et al. 2008). We replicated this
10 mutational study using 89 samples of Yoruba (YRI) population from the
11 Geuvadis dataset (Supplementary Figure S4) and the analysis of these results
12 consistently supports that elncRNA splicing variants impact cognate enhancer
13 function.

Figure 3. Disrupted elncRNA splicing impacts *cis*-gene regulation. (A) Example of a splice site variant that potentially disrupt elncRNA splicing. In contrast to individuals that carry the reference genome allele at a canonical 3' splice site, those with alternative alleles have carry any nucleotide but A (denoted as B) at the first position of the acceptor site. Distribution of the median fold difference, between individuals that carry alternative or reference alleles at elncRNA splice sites, in splicing index of (B) multi-exonic elncRNAs and (C) target protein coding genes; (D) target and (E) non-target gene expression levels (RPKM); and (F) elncRNA expression. Differences between groups were tested using a two-tailed Mann-Whitney U test. * $p < 0.05$; *** $p < 0.001$; NS $p > 0.05$.

1 elncRNA splicing causally mediate target expression.

2 The results of these mutational studies support the role of splicing in modulating
 3 enhancer activity and suggest that this is at least in part a consequence of
 4 increased transcription of multi-exonic elncRNAs. However, the
 5 interdependence between elncRNA splicing and expression together with the
 6 fact that splice site mutations may also affect enhancer factor binding make it
 7 difficult to directly use splice site mutations to disentangle the distinct roles of
 8 splicing and transcription on enhancer function. Furthermore, although
 9 unbiased, splice site mutations were only found for a subset of elncRNAs,
 10 restricting the extent of our analysis. To overcome these limitations, we first
 11 used multi-variate regression to investigate the relationship between splicing of
 12 elncRNAs and the expression of their putative targets genome-wide. First, we
 13 found that elncRNA splicing is correlated with target expression levels
 14 (Supplementary Figure S5A) and this correlation is significantly higher than
 15 compared to that with their proximal non-targets ($p < 2.2 \times 10^{-16}$, two-tailed Mann-
 16 Whitney *U* test, Supplementary Figure S5B). Next, we found by adding
 17 elncRNA splicing to its expression significantly, albeit moderately, improved the
 18 predicative power of its target gene expression ($p = 0.02$, two-tailed Mann-
 19 Whitney *U* test, Supplementary Figure S5C). This is in contrast to elncRNA
 20 proximal non-targets, whose abundance was not explained by elncRNA
 21 splicing ($p = 0.8$, two-tailed Mann-Whitney *U* test, Supplementary Figure S5C),
 22 supporting that the impact of elncRNA splicing on gene expression is target-
 23 specific.

24 Next, we expanded the number of elncRNA loci considered in our mutational
 25 study by identifying SNPs associated with the amount of splicing at multi-exonic
 26 elncRNAs (sQTLs) (Figure 4A). To directly assess the impact of elncRNA
 27 splicing on enhancer activity independent of its transcription, we excluded sQTL
 28 variants that are also associated with elncRNA expression (eQTLs) from this
 29 analysis (Methods). The average distance between these variants and
 30 elncRNA transcriptional start sites and their associated-enhancer boundaries is

1 63.7 kb and 72.2 kb, respectively (Supplementary Figure S6A-B), eliminating
2 the potential confounding impact of sQTLs on enhancer factor binding. We
3 identified sQTLs for 25% (88/352) of multi-exonic elncRNAs. Of these, more
4 than half (49, 56%) had at least one sQTL that was also associated with their
5 target expression (80 elncRNA-target pairs, 2,354 joint seQTLs).

Figure 4. Impact of elncRNA splicing on *cis*-gene regulation in the human population. (A) RP11-112J3 (ENSG00000227388) is a multi-exonic elncRNA whose splicing is associated with genotype of the SNP variant (rs12337802), which is also associated to the expression level of an elncRNA target, GBA2 (ENSG00000070610). (Top panel) Distribution of the splicing index of elncRNA (RP11-112J3, red) and target expression (ENSG00000070610, green) in samples across the population that carry different alleles of the SNP variant (rs12337802). Spearman's rho and p-values are shown. (Bottom panel) Genome browser illustrating the genomic positions of the elncRNA (RP11-112J3, red), its target (GBA2, green), and their associated SNP (rs12337802, black). (B) Four models of causal inference testing that predict the relationship between joint seQTL variant (black box) associations with the splicing (sQTL) of multi-exonic elncRNAs (red boxes) and the expression level (eQTL) of their target protein-coding genes (green boxes). Schematic representation of the models of joint seQTL associations: (1) the variants are independently associated with elncRNA splicing and target expression (independent model); (2) direct association between the variant and elncRNA splicing mediates the indirect association between that and target expression (causal model); (3) direct association between the variant with target expression mediates the indirect association between that and elncRNA splicing (reactive model); and (4) the causative interaction between elncRNA splicing and target expression is more complex (undecided model). Direct associations are depicted as solid lines and indirect associations as dash lines. (C) Scatterplot depicting causal inference testing local FDR associated with each the four models (as illustrated in B). Number and proportion of joint seQTLs are provided in brackets for each model. Dotted red lines denote significance threshold at local FDR < 0.1. (D) Proportion of joint elncRNA seQTLs causally or non-causally predicted to mediate target gene expression at the 5' end (red) or 3' end (grey) ends of the transcripts. Differences between groups were tested using a two-tailed Fisher's exact test. *** $p < 0.001$.

1

2 Given that quantitative trait variants were identified using a relatively small
3 cohort (373 LCL samples) and that cohort size is an important factor in QTL
4 calling, we reasoned that the relatively small fraction of elncRNAs (~14%,
5 49/352) with joint seQTL associations is a consequence of our study being
6 under-powered. Using a dataset with nearly 85-times more samples (31,684
7 blood samples, eQTLGen consortium) (Võsa et al. 2018), we found nearly 80%
8 of elncRNAs (70/88) splicing QTLs are jointly associated to target expression
9 (171 elncRNA-target pairs). Compared to the smaller LCL data, target-seQTL
10 associations were considerably higher (3.45×10^{61} times, $p < 2.2 \times 10^{-16}$, two-tailed
11 Mann-Whitney U test, Supplementary Figure S6C). In addition, using the larger
12 dataset, target eQTLs that were not identified to be associated with elncRNA
13 splicing by the smaller LCL set (non-seQTLs) were associated to a similar level
14 as those identified to be jointly associated to elncRNA splicing (seQTLs) ($p=0.3$,
15 two-tailed Mann-Whitney U test, Supplementary Figure S6D). This
16 corroborates that our analysis is limited by the sample size of the LCL data and

1 we are likely to be under-estimating the prevalence of elncRNAs whose splicing
2 contribute to enhancer activity. However, 80% of the expression data used by
3 the eQTLGen consortium was generated using microarray technology which
4 cannot be used to predict splicing frequency. Given this limitation and the
5 restricted access to its sensitive data, we could not carry out detailed analysis
6 using the eQTLGen data and continued to use the LCL dataset.

7 Different models can account for the joint association between a variant and
8 elncRNA splicing or its target expression (joint seQTL, Figure 4B) (Li et al.
9 2010). This can occur if the variant is independently associated with elncRNA
10 splicing and target expression (independent model) or as the result of the
11 complex interaction between splicing and expression (undecided model)
12 Alternatively, variant can be associated with the target expression that in turn
13 mediates splicing of elncRNA (reactive model). Finally, cases where the seQTL
14 is associated with elncRNA splicing that in turn mediate target expression
15 (causal model) would support a functional role of elncRNA splicing. For each
16 seQTL, we established the most likely model of association using causal
17 inference analysis (Wang and Michoel 2017). With a false discovery rate of 10%,
18 we found that most associations between seQTLs (Figure 4C) and target
19 expression is causally mediated by elncRNA splicing (n=1328, 56.4%). These
20 causal seQTLs support the role of splicing for 61% of elncRNAs in the
21 regulation of target expression (1 to 2 targets on average). Of the remaining
22 variants, most are independently associated with elncRNA splicing and target
23 expression (n=969, 41.2%), which is likely in part due to the relatively high false
24 negative rates of this type of analysis (Wang and Michoel 2017). To account for
25 the effect of linkage disequilibrium and noise in genotyping data, we repeated
26 the analysis using only seQTLs that are the best (ie. most significant) variant
27 associated with elncRNA splicing. Using this conservative set of seQTLs, we
28 identified a similar proportion of elncRNAs (60%) whose splicing is predicted to
29 causally regulate their target expression (Supplementary Figure S7),
30 corroborating the causal impact of elncRNA splicing on their cognate enhancer
31 function.

1 Importantly, while the fraction of seQTL associated with 5' and 3' end splice
2 junctions is similar, 90% of those that support elncRNA splicing as a mediator
3 of target expression are located at the 5' end of the transcript, which is
4 consistent with the known synergy between transcription and 5' end splicing
5 (Figure 4D).

6 **elncRNA splicing impacts local chromatin state**

7 Evidence that variants associated with changes in elncRNA splicing also
8 mediate target gene expression support the direct role of splicing on enhancer
9 function. To assess if elncRNA splicing impacts cognate enhancer function by
10 directly impinging on its activity, we investigated changes in chromatin
11 signatures at elncRNA cognate enhancers. This was possible thanks to the
12 recent release of genome-wide histone modification data, including H3K4me1
13 and H3K27ac, for a subset of 150 LCL samples (Delaneau et al. 2019).
14 Consistent with a synergistic contribution of splicing to enhancer activity,
15 individuals that carried nucleotide variants that altered elncRNA splice site also
16 had significantly lowered cognate enhancer activity, which we estimated using
17 levels of H3K4me1 and H3K27ac (Heintzman et al. 2009) at cognate enhancers
18 (Figure 5A), supporting that splicing at elncRNA loci impacts local enhancer-
19 associated epigenetic state.

20 To extend this analysis genome-wide, we next identified nucleotide variants
21 associated with elncRNA splicing and chromatin modification (scQTL, 1,932
22 and 534 scQTLs involving 30 and 23 elncRNAs, for H3K4me1 and H3K27ac,
23 respectively). We determined the most likely relationship between scQTLs,
24 elncRNA splicing and chromatin signatures using causal inference analysis
25 (Wang and Michoel 2017). We found that a large proportion of scQTL variants
26 (855 (44.3%) and 397 (74.3%) scQTLs for H3K4me1 and H3K27ac,
27 respectively) associated independently to elncRNA splicing and chromatin
28 mark levels. This is expected given the relatively small number of samples
29 available, which has likely reduced the power of this analysis as discussed
30 above. Despite this limitation, we predicted that splicing of a sizable fraction of

1 elncRNAs modulates chromatin state (40% and 48% for H3K4me1 and
2 H3K27ac, respectively, Figure 5D,E). These fractions were significantly higher
3 compared to the fraction of elncRNAs whose splicing were predicted to be
4 impacted by chromatin modification changes (3% and 4% for H3K4me1 and
5 H3K27ac, respectively; $p < 0.05$, two-tailed Fisher's exact test, Figure 5D,E). We
6 conclude that splicing a sizable fraction of enhancer-associate transcripts
7 directly impacts the activity of their cognate enhancer activity.

8

9

Figure 5. Impact of elncRNA splicing on cognate enhancer chromatin signatures in the human population.

(A) Example of a splice site variant that potentially disrupt elncRNA splicing. In contrast to individuals that carry the reference genome allele at a canonical 3' splice site, those with alternative alleles have carry any nucleotide but A (denoted as B) at the first position of the acceptor site. Distribution of the median fold difference in chromatin signature, between individuals that carry alternative or reference alleles at elncRNA splice sites, in (B) H3K4me1 and (C) H3K27ac. (D) Scatterplot depicting joint elncRNA scQTLs in which elncRNA splicing is causally or non-causally predicted to mediate enhancer chromatin signatures, (D) H3K4me1 and (E) H3K27ac, respectively, using causal inference testing as illustrated using local FDR associated with the four models (as shown in Figure 4B). Dotted black lines denote significance threshold at local FDR < 0.1 . Differences between groups were tested using a two-tailed Mann-Whitney U test. ** $p < 0.01$.

10

11

1 DISCUSSION

2 We and others have previously shown that splicing of multi-exonic eIncRNAs is
3 strongly associated with high enhancer activity in several human and mouse
4 cell lines (Gil and Ulitsky 2018; Tan et al. 2020). The direct contribution of
5 splicing to enhancer activity has been shown for a few candidate eIncRNA
6 transcripts (Yin et al. 2015; Engreitz et al. 2016). However, it remains
7 uninvestigated whether elevated enhancer activity resulting from the splicing of
8 these eIncRNAs is a general phenomenon that would explain their previously
9 described genome-wide associations (Gil and Ulitsky 2018; Tan et al. 2020).

10 To gain initial insights into this, we first investigated the evolution of splicing-
11 associated motifs, including splice sites, within eIncRNAs. We found sequences
12 that support eIncRNA splicing evolved under purifying selection, an
13 evolutionary signature of functionality. Interestingly, splicing related motifs were
14 previously found to be amongst the most highly conserved sequences within
15 lncRNAs (Haerty and Ponting 2015). With at least half of all lncRNA transcripts
16 initiating at enhancers (Marques et al. 2013; Hon et al. 2017) and a sizable
17 proportion of these undergoing splicing (Gil and Ulitsky 2018; Tan et al. 2020),
18 it is likely the relative high sequence constraint observed at splicing-associated
19 motifs within lncRNAs (Haerty and Ponting 2015) can be attributed to enhancer
20 transcribed lncRNAs. Consistent with this, we found splicing-related motifs
21 within the less efficiently processed promoter-associated lncRNAs (plncRNAs)
22 (Tan et al. 2020) to be less constrained.

23 We investigated the direct contribution of eIncRNA splicing to cognate enhancer
24 activity using a statistical population genomics approach in human
25 lymphoblastoid cell lines (LCLs). We found splicing of at least 60% of eIncRNAs
26 causally mediates changes in their target gene expression levels. Interestingly,
27 a sizable proportion (40%) of eIncRNA splicing was also predicted to mediate
28 changes in chromatin signatures, including H3K4me1 and H3K27ac, at
29 enhancers. The present analysis is limited by the size of the cohort analysed.
30 Relatively small sample sizes reduce the ability to confidently identify

1 quantitative trait loci (Beavis 1998). This is illustrated by 1.4-fold increase in the
2 number of seQTL detected when using data from an much larger cohort
3 (eQTLGen (Võsa et al. 2018)) than the one from Geuvadis (Lappalainen et al.
4 2013). We anticipate that the use of larger cohorts should result in an increase
5 in the percentage of elncRNAs for which splicing can be confidently shown to
6 mediate target expression in *cis*.

7 Our analysis of elncRNA splice-motif evolution and the impact of elncRNA
8 processing on enhancer function and activity support that elncRNA splicing is
9 unlikely a spurious by-product of enhancer activity. This is consistent with
10 evidence supporting conserved elncRNA splicing and their lack of exonic
11 constraint (Haerty and Ponting 2013; Tan et al. 2020), suggesting the
12 contribution of elncRNA splicing to the modulation of enhancer activity go
13 beyond the processing of their mature transcripts.

14 Whereas the precise molecular mechanisms underlying this association remain
15 currently unknown, a number of mechanisms can support the role of elncRNA
16 splicing on enhancer activity. Splicing has long been known to directly impact
17 the rate of transcription (Le Hir et al. 2003). For example, DNA elements
18 embedded within introns have been shown to contribute to transcriptional
19 regulation (Sleckman et al. 1996) and components of the spliceosome can
20 directly enhance RNAPII initiation (Kwek et al. 2002) and transcript elongation
21 (Fong and Zhou 2001). Furthermore, it was recently shown that novel exon
22 splicing events can result in transcription initiation at novel exon proximal
23 promoters, possibly through the recruitment of transcription machinery by
24 splicing factors (Fiszbein et al. 2019).

25 The synergy between splicing and transcription may also be influenced by
26 changes in local chromatin environment. For example, splicing pattern changes
27 have been shown to correlate with histone modification and factors that alter
28 chromatin structure, including chromatin signatures at active enhancer regions
29 (Luco et al. 2010). On the other hand, tethering of enhancer-associated
30 transcripts at their site of transcription has been shown to alter local chromatin

1 environment (Lai et al. 2013; Li et al. 2013; Hsieh et al. 2014; Bose et al. 2017).
 2 Chromatin-bound lncRNAs have been recently shown to be enriched in U1
 3 small nuclear ribonucleoprotein (snRNP) binding, a protein essential for the
 4 recognition of nascent RNA 5' splice site and assembly of the spliceosome (Yin
 5 et al. 2020). The depletion of U1 snRNP reduced the tethering of lncRNAs to
 6 chromatin, suggesting a mechanism by which elncRNA splicing can strengthen
 7 enhancer activity (Yin et al. 2020).

8 The link between splicing and transcription, together with evidence of RNA-
 9 dependent recruitment and activity of some enhancer factors, provide initial
 10 clues to what might be the molecular mechanism(s) underlying the role of
 11 elncRNA splicing in the modulation of enhancer activity, as supported by the
 12 present work.

13

METHODS

Identification of enhancer-associated transcripts

We defined enhancer-associated transcripts by considering all ENCODE intergenic enhancers in human GM12878 lymphoblastoid cell line (LCL. 98,529 enhancers)(Bogu et al. 2015) overlapping DNase I hypersensitive sites (Encode Project Consortium 2012) and a CAGE cluster (Andersson et al. 2014) in LCLs (n=4997). We considered all LCL-expressed lncRNAs (Tan et al. 2017) and Ensembl annotated protein coding genes (version 75). LncRNAs whose 5' end is within 500bp of a LCL enhancer were considered as being transcribed from the enhancer (n=564 elncRNAs, Supplementary Table ST1), and all remaining enhancers were presumed to transcribe eRNAs (n=4433). Transcripts within 500bp of ENCODE LCL promoters (Encode Project Consortium 2012) were classified as protein-coding genes (n=12,070) or promoter-associated lncRNAs (plncRNAs, n=600) depending on their biotype annotation. We found 62.4% of elncRNAs (n=352), 60.5% of plncRNAs (n=363), and 94.4% of protein-coding genes (n=11,392) to be multi-exonic.

Read mapping and quantification

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 human reference genome (hg19) using HISAT2 (version 2.0.2) (Kim et al. 2015). Since available Geuvadis RNA sequencing and genotype data, LCL CAGE data and enhancer predictions, were all originally mapped to GRCh37 (hg19), our analyses were performed using the same human genome assembly.

Metagene analysis of elncRNAs

Direction of transcription was assessed with metagene profiles of CAGE reads centered at LCL enhancers and gene transcriptional start sites (TSS) plotted using NGSplot (version 2.4) (Shen et al. 2014) as in (Tan et al. 2020).

1 Enrichment of histone modifications, transcription factor binding, and gene
2 expression levels were assessed using publicly available GM12878 DNase-seq,
3 ChIP-seq and RNA-seq data sets (Supplementary Table ST3). Metagene
4 profiles of sequencing reads centered at gene TSSs were visualized using
5 HOMER (version 4.7) (Heinz et al. 2010).

6 ***Analysis of chromosomal architecture***

7 Enrichment of enhancer-associated transcripts at LCL loop anchors (Rao et al.
8 2014), relative to expectation, was assessed using the Genome Association
9 Tester (GAT, version 1.3.6) (Heger et al. 2013). Specifically, loop positional
10 enrichment was compared to a null distribution obtained by randomly sampling
11 10,000 times (with replacement) segments of the same length and matching
12 GC content as the tested loci within mappable intergenic regions (as predicted
13 by ENCODE (Hoffman et al. 2013)). To control for potential confounding
14 variables that correlate with GC content, such as gene density, the genome
15 was divided into segments of 10 kb and assigned to eight isochore bins in the
16 enrichment analysis.

17 The frequency of chromosomal interactions within topologically associating
18 domains (TADs) was calculated using LCL Hi-C contact matrices (Rao et al.
19 2014), as previously described (Tan et al. 2017).

20 ***Identification of splicing-associated motifs***

21 Density of human exonic splicing enhancers (ESEs), using a set of 238
22 computationally predicted motifs (Fairbrother et al. 2002) and a more stringent
23 set of 54 motifs identified across multiple studies (Caceres and Hurst 2013),
24 were predicted within LCL transcripts as previously described (Haerty and
25 Ponting 2015). Specifically, exonic nucleotides (50 nt) flanking splice sites (SS)
26 of internal transcript exons (> 100 nt) were considered in the analysis, after
27 masking the 5 nt immediately adjacent to SS to avoid splice site-associated
28 nucleotide composition bias (Fairbrother et al. 2002; Yeo and Burge 2004).
29 Canonical U1 sites (GGUAAG, GGUGAG, GUGAGU) adjacent to 5' splice sites

(3 exonic nt and 6 intronic nt flanking the 5' SS) were predicted as previously described (Almada et al. 2013). FIMO (MEME version 4.12) (Grant et al. 2011) was used to search for perfect hexamer matches within these sequences.

GC content

For all LCL expressed genes with at least 2 exons, we computed the GC content for the first exon, remaining exons, and introns for each gene separately, as well as their flanking intergenic sequences of the same length (after excluding the 500 nucleotides immediately adjacent to annotations).

Splicing efficiency

Transcript splicing efficiency was estimated by computing the proportion of fully excised introns using bam2ssj (IPSA version 3.3) (Pervouchine et al. 2013) of transcripts for each gene. Using long non polyA-selected RNA-seq data in GM12878 (Encode Project Consortium 2012), we calculated the splicing index, coSI (θ), which represents the ratio of reads that span exon-exon splice junctions (excised intron) over those that overlap exon-intron junctions (incomplete excision) (Tilgner et al. 2012).

elncRNA conservation across evolution and in humans

To assess selective constraints of transcripts, we estimated their pairwise nucleotide substitution rates. First, pairwise alignment between human and mouse of the different splicing-associated features (including canonical splice sites, exonic splicing enhancers and canonical U1 sites) were separately concatenated for multi-exonic elncRNAs, plncRNAs and protein-coding genes. We used BASEML from the PAML package [version 4.9, REV substitution model (Yang 1997)] to estimate pairwise nucleotide substitution rates of each splicing motif. To determine whether the splicing motifs have evolved under significant purifying selection, we compared the observed nucleotide substitution rate to that of randomly selected pseudo-splicing motif sequences of the same length and GC content from neighbouring (within 1Mb) neutrally

1 evolving sequences ancestral repeats (ARs, (Lunter et al. 2006)). We repeated
2 this process 1000 times to obtain a distribution of the expected substitution
3 rates.

4 To assess the conservation of splicing motifs during modern human evolution,
5 we determined their derived allele frequency (DAF), as previously described
6 (Haerty and Ponting 2013). We identified common single nucleotide
7 polymorphisms (SNPs) in the human population (dbSNP build 150) mapped
8 within splicing motifs of elncRNAs, plncRNAs and protein-coding genes. DAF
9 was calculated using information on ancestral allele and the frequency of these
10 SNPs in the European population obtained from the 1000 Genomes Project
11 (The 1000 Genomes Project Consortium 2012). To determine the expected
12 background distribution, we compared DAF spectrum of the observed splicing
13 motifs to that of randomly selected sequences of the same length and GC
14 content from local ARs.

15 ***Mapping of molecular quantitative trait loci (QTLs)***

16 Expression values (RPKM) of multi-exonic elncRNAs and protein-coding genes
17 in EBV-transformed LCLs derived from 373 individuals of European descent
18 (CEU, GBR, FIN and TSI) were quantified (as described in (Tan et al. 2017)).
19 The corresponding processed genotypes were downloaded from EBI
20 ArrayExpress (accession E-GEUV-1) (Lappalainen et al. 2013). Quantification
21 of alternative splicing events was estimated using LeafCutter (version 1.0) (Li
22 et al. 2018). Single nucleotide polymorphisms (SNPs) located within the same
23 TAD as the genes of interest were tested for association with splicing (sQTLs)
24 and with expression levels (eQTLs) of elncRNAs and protein-coding genes.
25 Only SNPs with minor allele frequency (MAF) greater than 5% were considered
26 in the QTL analyses. sQTLs and eQTLs were estimated using FastQTL (version
27 2.184) (Ongen et al. 2016). To assess the significance of the correlation globally,
28 we permuted the splicing or expression levels of each gene 1000 times and
29 noted the maximum permuted absolute regression coefficient ($r_{\max}$). We
30 considered only sQTLs or eQTLs with an observed absolute regression

coefficient (r_{obs}) greater than 95% of all permuted r_{max} values to be significant (Lappalainen et al. 2013). We further performed Benjamini-Hochberg multiple testing correction to estimate FDR (<5%) for all SNPs within the same TAD. Putative protein-coding gene targets of multi-exonic elncRNAs were predicted as those that reside within the same TADs and whose expression levels were associated to the same SNP variant as the expression of the elncRNAs (ie. associated to the same eQTL). We found 88 elncRNAs to have putative protein-coding gene targets and are associated with at least one sQTL ($n=26,741$).

Levels of histone modification marks (H3K4me1 and H3K27ac, CPM) that overlap enhancers associated with elncRNAs were downloaded for the 150 European individuals with available chromatin data (Delaneau et al. 2019). To assess the relationship between elncRNA splicing and local histone modifications that mark enhancer elements, we calculated chromatin QTLs (cQTLs) associated with cognate enhancers of elncRNAs in the same way as sQTLs and eQTLs.

Impact of genetic variation at elncRNA splice sites on cis-gene expression

We considered all SNPs located at elncRNA splice sites and estimated the fold difference in elncRNA splicing and steady state abundance, as well as in their putative target expression and chromatin marks at their cognate enhancer elements, between individuals that carry the reference or alternative alleles of these variants (Supplementary Table ST2).

Causality inference between elncRNA splicing and nearby protein-coding gene expression

To infer the causal relationship between elncRNA splicing and putative target gene expression, we focused on QTLs that are associated with both splicing of elncRNAs (sQTL) and their putative target gene abundance (eQTL), and we refer to these variants as joint seQTLs. elncRNA sQTL variants that were also associated with elncRNA expression level or splicing of their putative *cis*-target

1 genes were excluded from the analysis (n=21,650 out of 26,741). In total, we
2 found 49 eIncRNAs with seQTLs shared with 80 target genes.

3 For all triplets of seQTL – eIncRNA splicing – target gene expression (n=2,349),
4 we performed causal inference testing using a Bayesian approach as
5 implemented by Findr (Wang and Michael 2017) by testing the models: (1) the
6 independent model where seQTL variants are independently associated with
7 eIncRNA splicing and target gene expression; (2) the causal model with
8 eIncRNA splicing as the molecular mediator of gene expression; (3) the reactive
9 model where gene expression mediates eIncRNA splicing; and (4) the
10 undecided model where causative interaction between eIncRNA splicing and
11 target gene expression is more complex (Li et al. 2010).

12 The same causal inference testing was performing between eIncRNA splicing
13 (sQTL) and chromatin marks (cQTL) at their cognate enhancers by identifying
14 scQTLs. We found 30 and 23 eIncRNAs whose splicing-QTL is also associated
15 with H3K4me1 (n=1391 triplets) and H3K27ac (n=532 triplets) marks,
16 respectively, at their cognate enhancers.

17 ***Linear regression***

18 We used multi-variable regression to assess whether eIncRNA splicing
19 contributes to their cognate enhancer activity by comparing the difference in the
20 proportion of variance in target expression is explained by adding eIncRNA
21 splicing to the model (Target-expression ~ eIncRNA-expression compared to
22 Target-expression ~ eIncRNA-expression + eIncRNA-splicing).

23 ***Statistical tests***

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

26 ***Data Access***

- 1 Analyses were performed using publicly available command-line tools.
- 2 Versions and deviations from parameters used are as detailed in the Methods.
- 3 All scripts used to parse the results are available upon request.

4 **Competing interests**

- 5 The authors declare that they have no competing interests.

6 **Acknowledgements**

- 7 We thank members of the Marques group for valuable comments and
- 8 discussion. We thank Zoltán Kutalik and Diogo Ribeiro for discussion on
- 9 population genomics analysis; Olivier Delanau for early access to chromatin
- 10 mark quantification. This work is funded by the Swiss National Science
- 11 Foundation grant (PP00P3_179065 to ACM).

12 **Authors' contributions**

- 13 JYT and ACM designed the study. JYT performed the experiments and
- 14 analysed the results. JYT and ACM discussed the results. ACM supervised the
- 15 study. JYT and ACM wrote the manuscript. All authors approved the manuscript.

16

17

1 REFERENCES

- 2 Almada AE, Wu XB, Kriz AJ, Burge CB, Sharp PA. 2013. Promoter directionality is controlled
3 by U1 snRNP and polyadenylation signals. *Nature* **499**: 360-U141.
- 4 Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, Lev-Maor G, Burstein D, Schwartz
5 S, Postolsky B et al. 2012. Differential GC content between exons and introns
6 establishes distinct strategies of splice-site recognition. *Cell Rep* **1**: 543-556.
- 7 Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X,
8 Schmidl C, Suzuki T et al. 2014. An atlas of active enhancers across human cell types
9 and tissues. *Nature* **507**: 455-461.
- 10 Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, Lennartsson A, Ronnerblad
11 M, Hrydziuszko O, Vitezic M et al. 2015. Transcribed enhancers lead waves of
12 coordinated transcription in transitioning mammalian cells. *Science* **347**: 1010-1014.
- 13 Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ. 1997. Intergenic transcription and
14 transinduction of the human beta-globin locus. *Genes & development* **11**: 2494-2509.
- 15 Beavis WD. 1998. QTL analysis: Power, precision, and accuracy. In: *Paterson AH, editor*
16 *Molecular dissection of complex traits* Boca Raton, FL: CRC Press.
- 17 Bogu GK, Vizan P, Stanton LW, Beato M, Di Croce L, Marti-Renom MA. 2015. Chromatin and
18 RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Molecular and cellular*
19 *biology* **36**: 809-819.
- 20 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
21 data. *Bioinformatics* **30**: 2114-2120.
- 22 Bose DA, Donahue G, Reinberg D, Shiekhhattar R, Bonasio R, Berger SL. 2017. RNA Binding
23 to CBP Stimulates Histone Acetylation and Transcription. *Cell* **168**: 135-149 e122.
- 24 Caceres EF, Hurst LD. 2013. The evolution, impact and properties of exonic splice enhancers.
25 *Genome biology* **14**: R143.
- 26 Damgaard CK, Kahns S, Lykke-Andersen S, Nielsen AL, Jensen TH, Kjems J. 2008. A 5' splice
27 site enhances the recruitment of basal transcription initiation factors in vivo. *Molecular*
28 *cell* **29**: 271-278.
- 29 De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei
30 CL, Natoli G. 2010. A large fraction of extragenic RNA pol II transcription sites overlap
31 enhancers. *PLoS biology* **8**: e1000384.
- 32 Delaneau O, Zazhytska M, Borel C, Giannuzzi G, Rey G, Howald C, Kumar S, Ongen H,
33 Popadin K, Marbach D et al. 2019. Chromatin three-dimensional interactions mediate
34 genetic effects on gene expression. *Science* **364**.
- 35 Encode Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human
36 genome. *Nature* **489**: 57-74.
- 37 Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M,
38 Lander ES. 2016. Local regulation of gene expression by lncRNA promoters,
39 transcription and splicing. *Nature* **539**: 452-455.

- 1 Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing
2 enhancers in human genes. *Science* **297**: 1007-1013.
- 3 Fiszbein A, Krick KS, Begg BE, Burge CB. 2019. Exon-Mediated Activation of Transcription
4 Starts. *Cell* **179**: 1551-1565 e1517.
- 5 Fong YW, Zhou Q. 2001. Stimulatory effect of splicing factors on transcriptional elongation.
6 *Nature* **414**: 929-933.
- 7 Furger A, O'Sullivan JM, Binnie A, Lee BA, Proudfoot NJ. 2002. Promoter proximal splice sites
8 enhance transcription. *Genes & development* **16**: 2792-2799.
- 9 Gil N, Ulitsky I. 2018. Production of Spliced Long Noncoding RNAs Specifies Regions with
10 Increased Enhancer Activity. *Cell systems* **7**: 537-547 e533.
- 11 Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif.
12 *Bioinformatics* **27**: 1017-1018.
- 13 Haerty W, Ponting CP. 2013. Mutations within lncRNAs are effectively selected against in fruitfly
14 but not in human. *Genome biology* **14**: R49.
- 15 Haerty W, Ponting CP. 2015. Unexpected selection to retain high GC content and splicing
16 enhancers within exons of multiexonic lncRNA loci. *Rna* **21**: 333-346.
- 17 Heger A, Webber C, Goodson M, Ponting CP, Lunter G. 2013. GAT: a simulation framework
18 for testing the association of genomic intervals. *Bioinformatics* **29**: 2046-2048.
- 19 Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart
20 RK, Ching CW et al. 2009. Histone modifications at human enhancers reflect global
21 cell-type-specific gene expression. *Nature* **459**: 108-112.
- 22 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass
23 CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-
24 regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**:
25 576-589.
- 26 Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen
27 PM, Bilmes JA, Birney E et al. 2013. Integrative annotation of chromatin elements from
28 ENCODE data. *Nucleic acids research* **41**: 827-841.
- 29 Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, Denisenko E,
30 Schmeier S, Poulsen TM, Severin J et al. 2017. An atlas of human long non-coding
31 RNAs with accurate 5' ends. *Nature* **543**: 199-204.
- 32 Hsieh CL, Fei T, Chen Y, Li T, Gao Y, Wang X, Sun T, Sweeney CJ, Lee GS, Chen S et al.
33 2014. Enhancer RNAs participate in androgen receptor-driven looping that selectively
34 enhances gene activation. *Proceedings of the National Academy of Sciences of the*
35 *United States of America* **111**: 7319-7324.
- 36 Iott NE, Heward JA, Roux B, Tsitsiou E, Fenwick PS, Lenzi L, Goodhead I, Hertz-Fowler C,
37 Heger A, Hall N et al. 2014. Long non-coding RNAs and enhancer RNAs regulate the
38 lipopolysaccharide-induced inflammatory response in human monocytes. *Nature*
39 *communications* **5**: 3979.
- 40 Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW,
41 Varnai C, Thiecke MJ et al. 2016. Lineage-Specific Genome Architecture Links

1 Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**:
2 1369-1384 e1319.

3 Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory
4 requirements. *Nature methods* **12**: 357-360.

5 Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-
6 Haley K, Kuersten S et al. 2010. Widespread transcription at neuronal activity-regulated
7 enhancers. *Nature* **465**: 182-187.

8 Kowalczyk MS, Hughes JR, Garrick D, Lynch MD, Sharpe JA, Sloane-Stanley JA, McGowan
9 SJ, De Gobbi M, Hosseini M, Vernimmen D et al. 2012. Intragenic enhancers act as
10 alternative promoters. *Molecular cell* **45**: 447-458.

11 Kwek KY, Murphy S, Furger A, Thomas B, O'Gorman W, Kimura H, Proudfoot NJ, Akoulitchev
12 A. 2002. U1 snRNA associates with TFIIF and regulates transcriptional initiation.
13 *Nature structural biology* **9**: 800-805.

14 Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhata R. 2013.
15 Activating RNAs associate with Mediator to enhance chromatin architecture and
16 transcription. *Nature* **494**: 497-501.

17 Lam MT, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU,
18 Kim AS, Kosaka M et al. 2013. Rev-Erbs repress macrophage gene expression by
19 inhibiting enhancer-directed transcription. *Nature* **498**: 511-515.

20 Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-
21 Porta M, Kurbatova N, Griebel T, Ferreira PG et al. 2013. Transcriptome and genome
22 sequencing uncovers functional variation in humans. *Nature* **501**: 506-511.

23 Le Hir H, Nott A, Moore MJ. 2003. How introns influence and enhance eukaryotic gene
24 expression. *Trends in biochemical sciences* **28**: 215-220.

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

28 Li Y, Tesson BM, Churchill GA, Jansen RC. 2010. Critical reasoning on causal inference in
29 genome-wide linkage and association studies. *Trends in genetics : TIG* **26**: 493-498.

30 Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, Pritchard JK. 2018.
31 Annotation-free quantification of RNA splicing using LeafCutter. *Nature genetics* **50**:
32 151-158.

33 Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of
34 alternative splicing by histone modifications. *Science* **327**: 996-1000.

35 Lunter G, Ponting CP, Hein J. 2006. Genome-wide identification of human functional DNA using
36 a neutral indel model. *PLoS computational biology* **2**: e5.

37 Marques AC, Hughes J, Graham B, Kowalczyk MS, Higgs DR, Ponting CP. 2013. Chromatin
38 signatures at transcriptional start sites separate two equally populated yet distinct
39 classes of intergenic long noncoding RNAs. *Genome biology* **14**: R131.

- 1 Maruyama A, Mimura J, Itoh K. 2014. Non-coding RNA derived from the region adjacent to the
2 human HO-1 E2 enhancer selectively regulates HO-1 gene induction by modulating
3 Pol II binding. *Nucleic acids research* **42**: 13599-13614.
- 4 Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Oude Vrielink JA, Elkon R, Melo
5 SA, Leveille N, Kalluri R et al. 2013. eRNAs are required for p53-dependent enhancer
6 activity and gene transcription. *Molecular cell* **49**: 524-535.
- 7 Merika M, Williams AJ, Chen G, Collins T, Thanos D. 1998. Recruitment of CBP/p300 by the
8 IFN beta enhanceosome is required for synergistic activation of transcription. *Mol Cell*
9 **1**: 277-287.
- 10 Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager GL, Sartorelli
11 V. 2013. eRNAs Promote Transcription by Establishing Chromatin Accessibility at
12 Defined Genomic Loci. *Molecular cell* **51**: 606-617.
- 13 Natoli G, Andrau JC. 2012. Noncoding transcription at enhancers: general principles and
14 functional models. *Annual review of genetics* **46**: 1-19.
- 15 Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. 2016. Fast and efficient QTL mapper
16 for thousands of molecular phenotypes. *Bioinformatics* **32**: 1479-1485.
- 17 Pervouchine DD, Knowles DG, Guigo R. 2013. Intron-centric estimation of alternative splicing
18 from RNA-seq data. *Bioinformatics* **29**: 273-274.
- 19 R Development Core Team. 2008. R: A language and environment for statistical computing. R
20 Foundation for Statistical Computing.
- 21 Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL,
22 Machol I, Omer AD, Lander ES et al. 2014. A 3D map of the human genome at kilobase
23 resolution reveals principles of chromatin looping. *Cell* **159**: 1665-1680.
- 24 Schaukowitch K, Joo JY, Liu X, Watts JK, Martinez C, Kim TK. 2014. Enhancer RNA facilitates
25 NELF release from immediate early genes. *Molecular cell* **56**: 29-42.
- 26 Schoenfelder S, Fraser P. 2019. Long-range enhancer-promoter contacts in gene expression
27 control. *Nature reviews Genetics* **20**: 437-455.
- 28 Schuler A, Ghanbarian AT, Hurst LD. 2014. Purifying selection on splice-related motifs, not
29 expression level nor RNA folding, explains nearly all constraint on human lincRNAs.
30 *Mol Biol Evol* **31**: 3164-3183.
- 31 Shen L, Shao N, Liu X, Nestler E. 2014. ngs.plot: Quick mining and visualization of next-
32 generation sequencing data by integrating genomic databases. *BMC Genomics* **15**:
33 284.
- 34 Sleckman BP, Gorman JR, Alt FW. 1996. Accessibility control of antigen-receptor variable-
35 region gene assembly: role of cis-acting elements. *Annual review of immunology* **14**:
36 459-481.
- 37 Tan JY, Biasini A, Young RS, Marques AC. 2020. Splicing of enhancer-associated lincRNAs
38 contributes to enhancer activity. *Life science alliance* **3**.
- 39 Tan JY, Smith AAT, Ferreira da Silva M, Matthey-Doret C, Rueedi R, Sonmez R, Ding D, Kutalik
40 Z, Bergmann S, Marques AC. 2017. cis-Acting Complex-Trait-Associated lincRNA

1 Expression Correlates with Modulation of Chromosomal Architecture. *Cell Rep* **18**:
2 2280-2288.

3 The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from
4 1,092 human genomes. *Nature* **491**: 56-65.

5 Tilgner H, Knowles DG, Johnson R, Davis CA, Chakraborty S, Djebali S, Curado J, Snyder M,
6 Gingeras TR, Guigo R. 2012. Deep sequencing of subcellular RNA fractions shows
7 splicing to be predominantly co-transcriptional in the human genome but inefficient for
8 lncRNAs. *Genome research* **22**: 1616-1625.

9 Tong P, Monahan J, Prendergast JG. 2017. Shared regulatory sites are abundant in the human
10 genome and shed light on genome evolution and disease pleiotropy. *PLoS genetics*
11 **13**: e1006673.

12 Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A,
13 Kreuzhuber R, Kasela S et al. 2018. Unraveling the polygenic architecture of complex
14 traits using blood eQTL metaanalysis. *bioRxiv* doi:10.1101/447367: 447367.

15 Wan Y, Larson DR. 2018. Splicing heterogeneity: separating signal from noise. *Genome*
16 *biology* **19**: 86.

17 Wang L, Michoel T. 2017. Efficient and accurate causal inference with hidden confounders from
18 genome-transcriptome variation data. *PLoS computational biology* **13**: e1005703.

19 Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood.
20 *Comput Appl Biosci* **13**: 555-556.

21 Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications
22 to RNA splicing signals. *J Comput Biol* **11**: 377-394.

23 Yin Y, Lu JY, Zhang X, Shao W, Xu Y, Li P, Hong Y, Cui L, Shan G, Tian B et al. 2020. U1
24 snRNP regulates chromatin retention of noncoding RNAs. *Nature* **580**: 147-150.

25 Yin Y, Yan P, Lu J, Song G, Zhu Y, Li Z, Zhao Y, Shen B, Huang X, Zhu H et al. 2015. Opposing
26 Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene
27 Activation during Embryonic Stem Cell Differentiation. *Cell stem cell* **16**: 504-516.
28

29

A.

B.

C.

D.

E.

F.

A.

B.

C.

D.

Supplementary Figure S1

A.

B.

C.

— Multi-exonic elncRNA — Single-exonic elncRNA — eRNA

Supplementary Figure S3

A.

XLOC_013468 (6 exons, 1 target, snp_4_185303531)

B.

ENSG00000250903 (4 exons, 1 target, snp_6_2399099)

C.

ENSG00000227388 (2 exons, 7 targets, snp_9_35790151)

D.

ENSG00000205786 (5 exons, 5 targets, snp_19_35897981)

Supplementary Figure S5

A.

B.

C.

Model 1: target-expression ~ elncRNA-expression

Model 2: target-expression ~ elncRNA-expression + elncRNA-splicing

Supplementary Figure S6

Supplementary Figure S7