Distinctive features of long non-coding RNA chromatin (dis-)association

Evgenia Ntini[1,2,*], Stefan Budach[1,2], Ulf A Vang Ørom[3], Annalisa Marsico[1,2,4,*]

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

[2]Freie Universität Berlin, 14195 Berlin

[3]Aarhus University, Department of Molecular Biology and Genetics, 8000 Aarhus, Denmark

[4]Institute of Computational Biology, Helmholtz Zentrum Muenchen, Munich, Germany

*Correspondence: ntini@molgen.mpg.de; annalisa.marsico@helmholtz-muenchen.de (Lead Contact)

Highlights

- Chromatin (dis-)association of lncRNAs can be modeled using nascent RNA sequencing from pulse-chase chromatin fractionation
- Distinct physical and functional characteristics contribute to lncRNA chromatin (dis-)association
- lncRNAs transcribed from enhancers display increased degree of chromatin dissociation
- lncRNAs of distinct degrees of chromatin association display differential binding probabilities for RNA-binding proteins (RBPs)

Summary

Long non-coding RNAs (lncRNAs) are involved in gene expression regulation in *cis* and *trans*. Although enriched in the chromatin cell fraction, to what degree this defines their broad range of functions remains unclear. In addition, the factors that contribute to lncRNA chromatin tethering, as well as the molecular basis of efficient lncRNA chromatin dissociation and its functional impact on enhancer activity and target gene expression, remain to be resolved. Here, we combine pulse-chase metabolic labeling of nascent RNA with chromatin fractionation and transient transcriptome sequencing to follow nascent RNA transcripts from their co-transcriptional state to their release into the nucleoplasm. By incorporating functional and physical characteristics in machine learning models, we find that parameters like co-transcriptional splicing contributes to efficient lncRNA chromatin dissociation. Intriguingly, lncRNAs transcribed from enhancer-like regions display reduced chromatin retention, suggesting that, in addition to splicing, lncRNA chromatin dissociation may contribute to enhancer activity and target gene expression.

Keywords

lncRNAs, enhancer, nascent RNA transcription, co-transcriptional splicing, enhancer-associated lncRNAs, processing, co-transcriptional splicing, chromatin tethering, chromatin dissociation, RNA-binding protein interactions, predictive models, machine learning

Introduction

Bidirectional nascent RNA transcription is a prominent characteristic of active enhancers, leading to the production of short-lived non-coding RNA transcripts termed eRNAs. eRNAs are short and non-spliced, thus unstable, potentially terminated by the Integrator complex[1] and subjected to rapid exosome degradation[2], thereby contributing to their observed chromatin enrichment and eliminated detection in steady-state whole-cell RNA data. eRNA production, measured by nascent RNA-sequencing, along with DNase I hypersensitivity and distinct histone marks (H3K27Ac, H3K4me1) (and CBP/p300 binding) demarcate active enhancers[2-6].

Intriguingly, a small subset of bidirectionally transcribed enhancers, about ~3 to 5 %, produce a more stable and spliced long non-coding RNA (lncRNA) elongating in one direction[3,7] (& Tan and Marques, biorXiv 2020), while about one third to one fourth of annotated lncRNAs overlap enhancer-like regions[8]. Those enhancer-associated lncRNAs (*elncRNAs*) are associated with stronger enhancer activity (aka. higher nascent RNA transcription, H3K27Ac histone mark, DNase accessibility) and their expression is associated with changes in putative target gene expression and local chromatin structure. This suggests that elncRNA production contributes to gene expression regulation *in cis*[3,7]. However, to what degree elncRNAs remain chromatin-associated (in a manner analogous to the observed eRNA chromatin enrichment), and the degree to which their function depends on their chromatin (dis-)association remains obscure. In addition, the mechanistic basis of their exerted regulation on target gene expression *in cis* is not well characterized, and it remains an open question whether all elncRNAs would follow the same mechanistic mode in gene expression regulation. For instance, we showed that the lncRNA *A-ROD* transcribed from an active enhancer at the anchor point of a chromosomal loop in *MCF-7* cells enhances the expression of its target gene *DKK1* upon its post-transcriptional chromatin dissociation and within a pre-established chromosomal proximity. Enforcing *A-ROD* chromatin retention, by splicing inhibiting morpholinos or targeting polyadenylation, suppresses target gene expression, suggesting that chromatin dissociation is an important feature of lncRNA mediated gene expression regulation *in cis*[9].

A substantial portion of lncRNAs are enriched in the chromatin fraction, presumably tethered at their sites of transcription through elongating (transcriptionally engaged) Pol II, and are involved in regulation of proximal gene expression *in cis*[10-12]. However, intriguingly, lncRNAs transcribed from the anchor points of chromosomal loops and enhancer-like regions show significantly lower chromatin—to—nucleoplasmic enrichment at steady

65 state[9]. This may indicate that the process of chromatin dissociation, which relies on (co-transcriptional) RNA
66 maturation steps, could be important for the function of many enhancer-transcribed lncRNAs, acting *in cis* within
67 the spatial proximity of pre-established chromosomal loops[13].
68 A recent study additionally implicated U1 snRNP binding as a means of chromatin tethering for lncRNAs:
69 lncRNA exonic sequences are enriched in U1 recognition sites, while their gene bodies are depleted from 3'
70 splice sites (compared to mRNAs). This leads to persistent U1 snRNP binding —due to poor or inefficient
71 splicing efficiency—, which through additional protein interactions with transcriptionally engaged Pol II,
72 contributes to co-transcriptional lncRNA tethering (or post-transcriptional retargeting) to chromatin[14]. Intriguingly,
73 compared to other lncRNAs that are not enhancer-associated, elncRNAs display conserved splice sites and
74 significantly higher splicing efficiency, which is associated with local changes in chromatin states and positively
75 impacts their cognate enhancer activity[3,7,13]. Yet, a correlation between elncRNA splicing and chromatin-
76 association/dissociation has not been clarified. Although recent systems biology approaches strongly infer an
77 impact of elncRNA processing on enhancer activity, the role of elncRNA chromatin (dis-)association has not
78 been systematically examined.
79 In this work, we have combined pulse-chase metabolic labeling with chromatin fractionation and transient
80 transcriptome sequencing to follow nascent RNAs from the point of their transcription to their chromatin release
81 into the nucleoplasm. We have incorporated several parameters, physical and functional characteristics, in
82 machine learning models to predict distinct degrees of chromatin (dis-)association, and examined the
83 relationship between lncRNA chromatin dissociation and enhancer activity. Thus, two important questions are
84 addressed here: First, what are the parameters that contribute to distinct degrees of lncRNA chromatin
85 association or chromatin tethering. Second, whether increased chromatin dissociation of certain lncRNAs could
86 imply a functional potential, for instance by having an impact on— or shaping enhancer activity.
87

88 ## Results

89 ### Modeling chromatin (dis-)association of nascent RNA transcripts

90 To follow nascent RNA transcripts from their synthesis to their post-transcriptional chromatin dissociation we
91 performed nascent RNA sequencing from the chromatin-associated and nucleoplasmic fraction. We performed
92 4-thiouridine (4-SU) metabolic labeling of MCF-7 cells for an 8 min pulse, followed by 5, 10, 15 and 20 min
93 uridine chase (Methods). To additionally capture nascent RNA Pol II transcription in a high resolution and follow
94 transcription dynamics, we fragmented RNA prior to isolation of nascent RNA. Thus, our approach is similar to
95 'transient transcriptome sequencing' (*TT-seq*[15]) but coupled with chromatin fractionation and pulse-chase
96 labeling.

97 To model chromatin dissociation we extracted read coverage from the last exon, as we did not block new
98 transcription initiation events during the pulse-chase experiment (in the case of overlapping transcript isoforms
99 the longest transcript was selected; Methods). This was done to minimize transcriptional input from new
100 transcription initiation events during the pulse-chase time period and be closer to the transcript 3' end, thus
101 better reflecting capturing full-length transcripts. We see chromatin-associated read coverage decrease over
102 time and nucleoplasmic read coverage increase (Suppl. Figure S1A). We determined chromatin association as
103 the ratio CHR/(CHR+NP) at each time point and kept only transcripts with a defined ratio 0 to 1 at all time points
104 (NAs discarded, n = 15,157 transcripts). As expected, we see an overall decrease in the transcript chromatin
105 association over pulse-chase time (Figure 1A). We therefore fitted these ratios on an exponential decay curve to
106 extract a 'chromatin-association halftime' : [halftime = -(Intercept+ln2) / k]. For further analysis, we kept only
107 entries that fit the exponential decay curve with a p-value < 0.05 (n = 12,391 transcripts, of which 2,077 are
108 lncRNAs; Methods). We then split the dataset in 3 equal-size quantiles based on the calculated chromatin
109 association halftime, i.e. 'fast', 'medium' and 'slow' released transcripts (Figure 1B, 1C, Suppl. Figure S1B) (the
110 latter correspond to chromatin-retained transcripts). Alternatively, transcripts were clustered into 3 groups of
111 fast, medium and slow released using the CHR/(CHR+NP) ratios from the five time points as an input to k-
112 means clustering (Suppl. Figure S1C). In general, there is a good agreement between the two methods of
113 grouping, with the 3 groups of k-means clustering showing corresponding chromatin association halftimes
114 (Suppl. Fig. S1D). Although significantly shorter and with a smaller number of exons as previously reported[16],
115 lncRNAs show on average greater chromatin association halftimes compared to mRNAs (Suppl. Figures S1E-
116 G). Chromatin association halftimes extracted this way reflect the chromatin association ratios at steady state
117 (Suppl. Figure S1H). We find 872 fast, 499 medium and 706 slow-released lncRNAs (Suppl. Table 1). Two
118 representative lncRNAs are *A-ROD* as a fast-released, and *PVT1* as a slow-released, chromatin-retained
119 transcript (Figure 1D).
120 Nascent RNA sequencing from the chromatin associated fraction allows to follow Pol II transcriptional dynamics
121 in high resolution: application of a short metabolic pulse and RNA fragmentation prior nascent RNA purification,
122 as in the original TT-seq protocol[15], combined with chromatin fractionation further enriches for nascent RNA
123 reads[17]. By metagene analysis to profile nascent RNA transcription, we obtain Pol II transcriptional profiles
124 similar to the original TT-seq[15] (Figure 1E). Nascent RNA sequencing from the chromatin-associated fraction
125 also captures promoter-associated divergent transcription producing short unstable antisense transcripts
126 (PROMPTs)[18,19]. We note here that lncRNA loci produce higher upstream antisense transcription compared to
127 mRNAs which extends beyond the typical PROMPT length (~200 nt) (Figure 1E, lower right panel). This is most
128 probably because many lncRNAs arise upstream and antisense to protein coding genes (and the observed

upstream antisense signal is due to the associated mRNA transcription). Interestingly, we observe that fast-released lncRNAs display stronger upstream antisense signal, suggesting that fast-released lncRNAs originate more often upstream antisense of protein coding genes. Indeed, by plotting the interdistance to closest antisense protein coding gene TSS, we find that fast-released lncRNAs display on average significantly smaller values (Supplementary Figure S1 I). Notably, about half of fast-released lncRNAs originate within less than 1 kb antisense to mRNA TSS (either upstream or internal antisense) (Supplementary Figure S1 J). An example is the fast-released lncRNA *GATA3-AS1* transcribed upstream and antisense of *GATA3* (Supplementary Figure S1 M). As expected, ENCODE annotated lncRNAs with the biotype 'antisense' are enriched in fast-released transcripts (odds ratio 1.4657, p-value = 4.217e-06), whereas *de novo* assembled lncRNA transcripts from the chromatin-associated data not overlapping ENCODE annotations (Methods) are enriched in the slow released/chromatin-retained transcripts (odds ratio 2.070958, p-value 1.917e-11).

Nascent RNA sequencing coupled with chromatin fractionation reveals major co-transcriptional nascent RNA processing and some degree of post-transcriptional splicing

Nascent RNA sequencing from the chromatin-associated and nucleoplasmic fraction at different pulse-chase time points allows to track the progress of co- and post-transcriptional splicing. To measure splicing we used high confidence introns (Methods) and extracted splicing efficiency by calculating the ratio of split to non-split reads at the 3' splice site as in ref[20]. By plotting the cumulative fraction of intron splicing efficiencies from all time points and samples, we observe that most of the introns undergo extensive splicing co-transcriptionally while at chromatin, within the first 10-15 min of transcription (Figure 1F, Suppl. Figure S1K), and co-transcriptional splicing efficiency dynamics[21] (SED; Methods) is significantly higher compared to post-transcriptional nucleoplasmic SED (Figure 1G). These results are in agreement with recent reports that the majority of splicing occurs co-transcriptionally (Reimer et al., bioRxiv 2020). We then calculated the extent of post-transcriptional splicing (after chromatin dissociation) relative to co-transcriptional splicing (while at chromatin) (as the difference between chromatin and nucleoplasmic splicing efficiency, normalized to chromatin; Methods). This was done at intron and transcript level (by extracting a mean processing efficiency from a transcript's high-confidence introns; Methods). We observe that introns of fast-released lncRNAs, and respectively fast-released lncRNA transcripts undergo the least additional post-transcriptional splicing upon chromatin dissociation (Supplementary Figure S1 L i-ii), suggesting that most of their processing has been concluded co-transcriptionally while at chromatin. Overall, mRNAs may undergo some further post-transcriptional processing to a higher degree compared to lncRNAs (Suppl. Fig. S1 L iii). This is in agreement with recent findings using single molecule RNA FISH suggesting that some post-transcriptional splicing can occur upon chromatin dissociation, after

161 transcription is completed, and potentially while nascent RNA transcripts localize to speckles (Coté et al.,
162 bioRxiv 2020). That slow-released transcripts show overall more extensive post-transcriptional splicing (Suppl.
163 Fig. S1 L) is also in agreement with a model where completely synthesized nascent RNA transcripts move
164 slowly through a transcription site proximal zone (without being tethered to chromatin or the transcription site
165 anymore), while they can undergo additional post-transcriptional splicing (Coté et al., bioRxiv 2020).
166

167 **Different degrees of chromatin association correlate with distinct physical (and functional)**
168 **characteristics of nascent RNA transcripts**

169 We observe that lncRNAs show on average significantly lower co-transcriptional splicing efficiency compared to
170 mRNAs (Figure 2A left), which is in agreement with what was previously reported measuring splicing using
171 either steady-state or nascent RNA data[20,22]. In addition, fast-released mRNAs, but not lncRNAs, show on
172 average higher mean transcript splicing efficiencies compared to slow-released/chromatin-retained transcripts
173 (Figure 2A right). However, the minimum splicing efficiency per transcript (i.e. splicing efficiency of the worst
174 spliced intron) is significantly higher for fast-released lncRNAs compared to chromatin-retained transcripts,
175 suggesting that splicing of a slowly or inefficiently processed intron may act as a kinetic 'bottleneck' for nascent
176 RNA transcript chromatin dissociation (Figure 2B). As expected, lncRNAs show on average significantly higher
177 alternative splicing compared to mRNAs (intron *psi* value extracted as in ref[21]), and chromatin-retained lncRNAs
178 undergo significantly higher alternative splicing compared to fast-released transcripts (Suppl. Fig. S2A).
179 By extracting the promoter-associated transcriptional pausing index (Methods) using MCF-7 available Pol II P-
180 Ser2 ChIP-seq or GRO-seq data[23], we find that mRNAs show significantly higher pausing index as previously
181 reported[20,24] (Suppl. Fig. S2B-D, S2F). Interestingly, fast-released lncRNAs, but not mRNAs, display higher
182 pausing index compared to chromatin-retained transcripts, suggesting that transcriptional activity *per se* may
183 relate to lncRNA chromatin dissociation or tethering (Suppl. Fig. S2C, S2D). In agreement, we find significantly
184 different levels of transcriptionally engaged Pol II over the first Kb downstream of TSS for fast versus slow-
185 released lncRNAs, but not mRNAs which are overall more transcriptionally active (Supplementary Fig. S2E,
186 S2F). Taken together, these observations are in agreement with a recent report that promoters of lncRNAs
187 show distinct transcriptional burst kinetics compared to mRNAs (lower burst frequencies; Johnsson et al.,
188 bioRxiv 2020), and suggest that within lncRNAs, promoters of fast-released transcripts tend to be more
189 transcriptionally active and display higher degree of Pol II pausing compared to chromatin-retained lncRNA
190 transcripts. Interestingly, transcriptional pausing index was previously associated with lncRNA nuclear export[24].
191 We note here that by extracting transcription bi-directionality score (or divergent-transcription score) using GRO-
192 seq (as antisense/sense signal from 1 Kb around TSS) we reach the same conclusion by using chromatin

193 associated nascent RNA sequencing from time point 0 ('CHR0', Figure 1E), showing that fast-released lncRNAs
194 display significantly higher antisense (divergent) transcription (Supplementary Figure S2 G), which is most
195 probably due to their enrichment in originating near and antisense of protein-coding gene TSS (Suppl. Figures
196 S1 I-J).
197 Chromatin dissociation of nascent RNA transcript is coupled to transcription termination and 3' end formation.
198 We thus generated transcription metagene profiles around the transcript 3' end site (TES) using ChIP-seq signal
199 from transcriptionally engaged Pol II phosphorylated at Ser2 (P-ser2 Pol II occupancy) or strand-specific GRO-
200 seq read coverage. To account for annotation discrepancies, we extracted *de novo* putative (pA)
201 polyadenylation sites from ENCODE available MCF-7 nuclear polyA+ RNA-seq data using ContextMap[25].
202 Although in general there is good agreement between the annotated transcript 3' ends and the *de novo*
203 extracted pA sites (Suppl. Fig. S2 H), for increased positional accuracy we used the latter for further analyses
204 (i.e. assigned a transcript 3' end to closest and stronger ContextMap predicted pA site, Methods). P-Ser2 Pol II
205 metagene profiles around TES resemble the ones obtained by mNET-seq[20], revealing polyadenylation
206 associated Pol II pausing in a 2 Kb window downstream of TES of mRNAs, but not lncRNAs (Supplementary
207 Figure S2 I, left). In conjunction, mRNAs display significantly higher transcription termination index compared to
208 lncRNAs, as previously reported[20] (Supplementary Figure S2 I, right). GRO-seq metagene analysis profiles of
209 transcriptionally engaged Pol II verify these results (Figure 2C, Suppl. Figure S2 K). In particular, we find no
210 significant difference in the transcription termination index (extracted using GRO-seq) between fast and slow
211 released mRNA transcripts, indicating no significant differences in polyadenylation-associated TES-downstream
212 Pol II pausing (Suppl. Figure S2 K). This could suggest no significant differences in transcription termination
213 efficiencies *per se*. Yet, by extracting a Pol II 'travel index' (as the ratio of strand-specific GRO-seq signal from
214 the region 2.5 to 5 Kb downstream of TES to the first 2.5 Kb downstream of TES where the polyadenylation-
215 associated pausing resides[20]; Methods), we note that Pol II of slow-released transcripts tends to travel further
216 beyond the polyadenylation-associated pausing site, which would be in support of chromatin tethering via
217 ongoing transcription[10] (Suppl. Figure S2 L; S2 J; Figure 2C right panel) (or that ongoing transcription may
218 contribute to chromatin tethering and slow release of nascent RNA transcript). In the case of lncRNAs, we do
219 not observe a polyadenylation-associated TES-downstream Pol II accumulation or pausing, which is in
220 agreement with mNET-seq data suggesting polyadenylation-independent transcription termination modes[20].
221 Notably, and more evidently observed in normalized metagene transcriptional profiles, Pol II tends to transcribe
222 further beyond the TES of slow-released lncRNA transcripts (Suppl. Figure S2 J, lower panels). In agreement,
223 by extracting travel (readthrough) ratios using chromatin-associated nascent RNA-seq from time point 0
224 ('CHR0') we find that slow-released chromatin-retained nascent RNA transcripts, either mRNAs or lncRNAs,

225 exhibit higher readthrough transcription (Figure 2D). Taken together with the observed inefficient splicing of
226 slow-released transcripts (Figure 2A-B), these results are in agreement with a crosstalk between splicing,
227 transcription and transcription termination[26,27], and with recent findings that inefficient splicing associates with
228 readthrough transcription (Reimer et al., biorXiv 2020).
229
230 **Different degrees of chromatin association demarcated by distinct chromatin states**
231 We then examined whether distinct degrees of nascent RNA transcript chromatin association would relate to
232 distinct chromatin states. Notably, for all histone marks associated with transcriptional activity (H3K4me3,
233 H3K4me1, H3K27Ac) we see significant differences in the promoter regions around the TSS of fast, medium
234 and slow-released lncRNAs, but not for mRNAs (Figure 2E). By extracting the ratio H3K4me1 to H3K4me3
235 around the TSS, we observe that the fast-released lncRNAs resemble mRNAs in terms of promoter activity
236 (Suppl. Fig. S3 A), while slow-released/chromatin-retained lncRNAs display on average higher signals of
237 repressive histone marks like H3K9me3 and H3K27me3 (Suppl. Fig. S3 B). Profiles of total Pol II occupancy
238 (POL2RA ChIP-seq) confirm the differences in the transcriptional activity among distinct degrees of chromatin
239 association for lncRNAs (Suppl. Figure S3 C). Notably, fast-released lncRNAs are transcribed from regions with
240 significantly greater chromatin accessibility (measured by DNase-seq, Figure 2F), and display significantly
241 higher CTCF and YY1 binding (for the latter, Avocado calculated binding probability[28]; Methods) (Suppl. Figure
242 S3 D-E). This is important, as both factors are associated with chromatin looping, and YY1 in particular
243 promotes enhancer-promoter chromatin loops by forming protein dimers and facilitating DNA interactions[29]. The
244 negative correlation between the extracted chromatin association halftime and looping scores (i.e. promoter-
245 overlapping ChIA-PET nodes; Methods) is greater for lncRNAs compared to mRNAs (Pearson's correlation -
246 0.267 vs. -0.107, respectively). Notably, promoters of fast-released lncRNAs display significantly higher ChIA-
247 PET scores (Figure 2G, Suppl. Fig. S3 F), indicating that they are/tend to be transcribed from the anchor points
248 of chromosomal loops.
249
250 **Enhancer-associated lncRNAs (elncRNAs) do not remain chromatin-associated**
251 We therefore examined the association of distinct degrees of lncRNA chromatin dissociation with enhancer
252 activity. For this purpose we used the FANTOM5[2,6] human 'permissive' enhancers expanded by transcribed
253 enhancers defined by NET-CAGE[4]. We filtered that these enhancers should be transcriptionally active in MCF-7
254 cells by GRO-seq measurement, ending up with 10,008 high-confidence bidirectionally transcribed enhancers
255 (Fig. 3A). About 2.5 % of bidirectionally transcribed enhancers have an lncRNA TSS (derived from the analyzed
256 dataset 2,077 lncRNAs) within an interdistance < 2 kb which is reminiscent to what was previously reported[3,7].

257 Thus, those lncRNAs can be regarded as enhancer-associated elncRNAs[7] and their cognate enhancers as la-
258 EPCs[3]. Notably, fast-released lncRNAs are significantly enriched in elncRNAs (odds ratio 1.68, p-value
259 0.0001398). On the other hand, ~7.6 % of the bidirectionally transcribed enhancers have an mRNA TSS (from
260 the 10,314 analyzed) within less than 2 Kb interdistance, however fast-released mRNAs are not enriched in this
261 subset (odds ratio 0.97). This suggests that transcribed enhancers are more likely to be associated with a fast-
262 released lncRNA. In other words, when bidirectionally transcribed enhancers are associated with an lncRNA (at
263 ~3-5 %), then this is more likely to be a fast-released lncRNA transcript. Analogously, we find that elncRNAs
264 (defined at an interdistance < 2 kb to closest enhancer midpoint; Fig. 3C) are enriched in fast-released lncRNAs
265 (odds ratio ~1.8, p-value 0.002586), whereas mRNAs with an interdistance < 2 kb to closest enhancer midpoint
266 are not enriched in fast-released mRNAs (Fig 3B). Notably, elncRNAs show significantly higher association to
267 anchor points of chromatin loops (measured by score of overlapping ChIA-PET nodes; Fig. 3D), and display on
268 average significantly lower chromatin-association halftimes (p-value = 5.116e-06; Fig. 3E), (while, as a control,
269 fast-released mRNAs are not enriched in interdistances less than 2 kb to enhancer midpoint: odds ratio
270 0.8409119, p-value 0.03862). This is similar to what was previously published, that lncRNAs transcribed from
271 enhancer-like regions display on average higher ChIA-PET scores on their overlapping promoter regions[9].
272 Notably, although lncRNAs as a class display higher chromatin association halftimes compared to mRNAs,
273 elncRNAs escape this rule by showing significantly lower chromatin association halftimes (Fig. 3E), which is in
274 agreement with elncRNAs being enriched in fast-released transcripts. In conclusion, we show here that
275 enhancer-associated or rather, enhancer-transcribed lncRNAs (elncRNAs, equivalent to la-EPCs[3]), in addition
276 to increased splicing efficiencies[3,7], also show increased degrees of chromatin dissociation.
277

278 **Prediction of lncRNA chromatin dissociation in machine learning models**

279 We then incorporated several of the functional and physical characteristics in machine learning to predict
280 chromatin dissociation of lncRNAs. We applied logistic regression with a ten-times cross-validation to predict
281 fast versus slow-released transcripts (Figure 4A lncRNA, 4B mRNA). In agreement with the distribution of the
282 individual parameters (Figure 2, Suppl. Fig. S2, S3) we find that the transcript exon density (previously used as
283 a proxy for splicing activity[3]), splicing efficiency of the transcript's worst processed intron and chromatin states
284 associated with promoter transcriptional activity (H3K4me3 and H3K4me1) have significant coefficients in
285 predicting fast-released lncRNAs, whereas SNRP70 enrichment across the locus (mean of fold-enrichment from
286 ChIP-seq peaks; Methods) define slow-released, chromatin-retained lncRNAs. The latter is in agreement with
287 Yin et al. (2020) suggesting U1-mediated chromatin retention of inefficiently processed transcripts[14]. That P-
288 Ser2 Pol II coverage over gene body is significant in predicting slow-released lncRNAs could confirm that slow-

289 released lncRNAs are tethered to chromatin through transcriptionally engaged Pol II[10] and that transcriptional
290 activity could contribute to U1 snRNP-mediated tethering of inefficiently processed transcripts[14]. In contrast to
291 the exon density (which reflects overall splicing activity) and the splicing efficiency of the worst spliced intron,
292 potentially acting as a kinetic bottleneck in nascent RNA transcript chromatin release, and while those two
293 parameters confidently predict lncRNA chromatin dissociation, we (paradoxically) find the transcript's mean
294 splicing efficiency as an important predictive parameter of chromatin association. This could be explained if
295 some (or one, or few) easy to process introns achieve high splicing efficiency during their prolonged stay on
296 chromatin, thereby contributing to increasing the mean splicing efficiency of the host transcript. On the other
297 hand, and in agreement with the analyzed distributions (Figure 2; Suppl. Fig. S2, S3), chromatin states of
298 mRNA loci do not contribute to defining chromatin association of nascent mRNAs (Figure 4B). Similarly to
299 lncRNAs, exon density and splicing efficiency of the worst spliced intron predict mRNA chromatin dissociation,
300 while high SNRP70 enrichment over the transcription unit predicts slow-released mRNAs as well, suggesting
301 that chromatin association of slow-released mRNAs could be at least partially achieved through persistent U1
302 snRNP binding to inefficiently processed transcripts.
303 Apart from logistic regression, we also applied linear regression to predict the chromatin association halftime
304 (continuous value) as a multivariate function of several parameters (Supplementary Figure S4 A), as well as 2-
305 class random forest (Supplementary Figure S4 B), reaching similar results regarding the weight of parameters in
306 predicting fast versus slow-released transcripts.
307
308 **Distinct RNA binding proteins are predicted to bind transcripts of different degrees of chromatin**
309 **association**
310 We then asked whether lncRNAs of different degrees of chromatin association would interact with distinct RNA
311 binding protein (RBP) activities. For this, we used the ENCODE-available eCLIP data[30] from HepG2 cells as a
312 proxy dataset. As most lncRNAs are expressed in a cell-type specific manner, we trained the *pysster* algorithm[31]
313 on mRNA or lncRNA sequences with overlapping RBP binding sites to acquire full-length transcript binding
314 probabilities (by extracting the median score from positions that score above a pre-defined cutoff; Methods). We
315 then incorporated these in random forest machine learning models to predict fast versus slow-released lncRNA
316 or mRNA transcripts in 10 times cross-validation, with a mean accuracy of ~0.81 and ~0.8 respectively (Suppl.
317 Figure S4C). Interestingly, we find RBPs with high binding probabilities which are commonly important in
318 specifying chromatin association of both lncRNAs and mRNAs. These include factors with additional DNA
319 binding activity (localizing to chromatin) like the KH-domain containing factors KHSRP and KHDRBS1, FUBP3
320 and SUGP2 which display increased binding probabilities for chromatin-retained/slow-released transcripts,

321 either lncRNAs or mRNAs (Suppl. Figure S4 C lower panels). Interestingly, CSTF2 involved in 3' end
322 formation[32] is also enriched in slow-released transcripts, perhaps reflecting persistent binding and unresolved
323 RNA-protein complexes in the case of inefficient transcription termination and 3' end formation. The exosome
324 component EXOSC5 is also enriched in slow-released transcripts, implying chromatin-associated clearance of
325 inefficiently processed nascent RNA transcripts. Among type-specific RBPs, DROSHA is an interesting
326 candidate significantly enriched in fast-released lncRNAs, but not mRNAs, perhaps suggesting some
327 involvement in promoting lncRNA chromatin dissociation in a causal manner (Suppl. Figure S4 C upper panels).
328 Intriguingly, DROSHA was found important for pA-signal-independent transcription termination and 3' end
329 formation of lncRNAs serving as miRNA hosts[33]. Yet, the observed DROSHA enrichment (increased RNA
330 binding probability) specifically in fast-released lncRNAs could also suggest post-transcriptional processing of
331 nucleoplasmic-enriched lncRNAs. Although we do not find any significant enrichment of lncRNA miRNA hosts in
332 the fast-released lncRNA category (since the numbers are quite small to infer statistical significance; only 34 of
333 the 2,077 analyzed, expressed in MCF-7 lncRNAs host miRNAs), a more careful and closer examination would
334 be required to conclude about microprocessor involvement in lncRNA transcription termination (and 3' end
335 formation) as an applying mechanism. Additional lncRNA-specific factors with increased RNA binding
336 probabilities predictive for fast-released lncRNAs are NONO (involved in splicing), and XRN2 and CSTF2T,
337 involved in transcription termination and 3' end formation[34]. Since all three of them have DNA binding activity
338 and localize to chromatin, this suggests that their predicted binding could be co-transcriptional and their activity
339 may contribute to promoting chromatin dissociation of nascent lncRNA transcripts. Experimental examination by
340 assessing the chromatin (dis-)association of nascent RNA transcripts in differential conditions upon RBP factor
341 knock-down would validate these predictions and substantiate a specific candidate involvement in promoting
342 efficient chromatin release or tethering.
343

344 **Discussion**

345 lncRNAs constitute a large heterogeneous class with a broad range of functions in regulation of gene
346 expression (regulation of transcription *in cis* and *in trans*), RNA processing and chromatin states[12,35], while a
347 common feature that distinguishes lncRNAs from mRNAs is reduced splicing efficiency[22]. The exerted functions
348 of lncRNAs largely depend on their subcellular localization where they can differentially interact with distinct
349 RNA-binding proteins and posit local target specificity. Previous computational efforts aimed to generate
350 predictive models of lncRNA subcellular localization (nuclear versus cytoplasmic enrichment) using steady-state
351 RNA-sequencing, and showed that inefficient splicing and intron retention is a major predictor of nuclear
352 localization[24]. It is however an outstanding question what underlies the observed lncRNA chromatin enrichment

353 (usually referred to as chromatin retention or chromatin tethering). lncRNAs may remain tethered to chromatin
354 via ongoing Pol II transcription[10] (since inhibiting Pol II transcription elongation abolished lncRNA chromatin
355 tethering[10]), while the function of chromatin bound *cis* acting lncRNAs in regulation of proximal gene expression
356 and local chromatin structure is mostly coupled to their ongoing transcription[11,36,37]. DNA elements in the *cis*-
357 acting, chromatin-tethered lncRNAs may be key: for instance, loop interactions between the promoter of the
358 chromatin-tethered lncRNA *PVT1* and its intragenic enhancers antagonize interactions with the neighboring
359 *MYC* gene promoter[38].
360 Yin et al. (2020)[14] implicate persistent U1 snRNP binding as a means of lncRNA chromatin tethering, which
361 relies on U1 site enrichment in lncRNA exons, depletion of 3' splice sites and/or inefficient splicing, and U1
362 snRNP70 protein interactions with transcriptionally engaged Pol II. Interestingly, a previous study indicated that
363 the overall lower splicing efficiency of lncRNAs (compared to mRNAs) is not due to defects in the U1-PAS axis
364 which is very similar to mRNAs[22]. In agreement, we also find here SNRNP70 recruitment as a major predictive
365 factor of lncRNA chromatin retention. In Yin et al. (2020), U1 inhibition dampened the chromatin association of
366 both well and poorly spliced lncRNAs, suggesting that a kinetic effect due to delayed release of unspliced (or
367 inefficiently/poorly spliced) nascent RNA cannot be the major determinant for lncRNA chromatin retention. In
368 agreement, the transcript's mean co-transcriptional splicing efficiency is a major predictor of chromatin
369 dissociation for mRNAs but not lncRNAs. However, the splicing efficiency of the worst spliced intron per
370 transcript has a significantly high coefficient in predicting chromatin dissociation, suggesting that it might
371 function as a "bottleneck" for lncRNA chromatin release. Thus, nascent RNA splicing kinetics may at least
372 partially contribute to lncRNA chromatin dissociation. Future experimental examination by point-mutating
373 specific splice sites to enhance (or abolish) splicing will help to definitely validate the impact of co-transcriptional
374 splicing kinetics on lncRNA chromatin release.
375 An important finding is that lncRNAs transcribed from active enhancers display increased degree of chromatin
376 dissociation. This implies that the commonly termed "enhancer-associated" lncRNAs (or elncRNAs[7], equivalent
377 to la-EPCs[3]) do not remain chromatin associated. Instead, chromatin dissociation is an important feature which
378 might underlie their function and impact enhancer activity. Again, it is noteworthy that single locus experimental
379 validation aiming to alter the degree of lncRNA chromatin association will allow to examine the effect on cognate
380 enhancer activity and target gene expression. So far, decrease in lncRNA chromatin tethering was achieved
381 transcriptome-wide by inhibiting U1 snRNP[14] (without examining the associated effects on putative *cis* targets),
382 but it remains to be experimentally analyzed what is the effect of enforced elncRNA chromatin retention on
383 cognate enhancer activity and target gene expression. This can be achieved either in modified cell lines by
384 CRISPR gene editing or by targeting functional transcription termination and polyadenylation sites with blocking

oligonucleotides for a short period of time to avoid secondary dampening effects on transcriptional activity. Modifying donor and acceptor splice sites of individual lncRNA loci by inserting point mutations should allow to experimentally validate the correlation between splicing efficiency and chromatin dissociation. It would also be highly relevant under such experimental conditions to examine alterations in local chromatin states and loop conformation, so as to characterize chromatin structure associated effects caused by enforced lncRNA chromatin tethering on enhancer functionality. Of great interest will be to draw conclusions on cognate enhancer activity after locus manipulation leading to altered locus-specific lncRNA chromatin-tethering without affecting splicing activity. This can be achieved for instance by interfering with 3' end formation leading to increased transcriptional readthrough and suppressing nascent RNA transcript release[39].

We note here that our approach to couple nascent RNA sequencing with chromatin fractionation at different pulse-chase time points would benefit by additionally applying long RNA sequencing of chromatin-associated and released nascent transcripts. The 3' ends of long reads represent the position of Pol II at full-length (non-fragmented) synthesized nascent RNA transcripts, and this technique was recently employed to corroborate that co-transcriptional splicing greatly enhances mammalian gene expression (Reimer et al., biorxiv 2020). Previous experimental and computational studies focused at understanding nuclear retention of lncRNAs[24,40]. In these predictive models, inefficient splicing was a major factor contributing to lncRNA nuclear retention[24]. Here, by combining chromatin fractionation with sequencing of nascent RNA from the chromatin-associated and nucleoplasmic fraction at different pulse-chase time points and by employing machine learning we show that splicing of the least efficiently processed intron per transcript may act as a 'bottleneck' for efficient nascent RNA transcript chromatin release. Other factors like U1 snRNP (SNRNP70) binding, coupled with inefficient splicing, contribute to lncRNA chromatin retention as it was recently demonstrated in mESC[14]. We additionally show that lncRNAs transcribed from active enhancers do not remain chromatin tethered but rather display increased chromatin dissociation efficiency. Essentially elncRNAs are enriched in fast-released lncRNA transcripts, thus increased chromatin dissociation efficiency in addition to splicing[3,7] may contribute to shaping enhancer activity and regulation of target gene expression. The latter may be accomplished upon chromatin dissociation of the nascent lncRNA transcript forming or affecting regulatory protein interactions targeting gene expression *in cis* within the spatial proximity of pre-established chromosomal loops.

Figure legends

Figure 1. Measuring chromatin association of nascent RNA transcripts

(A) Distribution of chromatin association ratios at different pulse-chase time points.

(B) Same as in (A) but split for fast, medium and slow-released transcripts.

(C) Loess curve of chromatin association drawn based on the raw ratios (upper panel) and after fit on an exponential decay (lower panel).

(D) Exponential decay fit of the chromatin association over time for two representative lncRNAs, A-ROD (fast-/efficiently released) and PVT1 (slow-released/chromatin retained).

(E) Metagene analysis of 'CHR0' strand-specific read coverage (chromatin-associated nascent RNA sequencing from time point zero) in a ±3 Kb window around TSS.

(F) Cumulative distribution function (CDF) curves of intron splicing efficiencies measured in all analyzed samples.

(G) Distribution boxplots of intron splicing efficiency dynamics (SED) measured in the chromatin-associated (CHR20-CHR0) and nucleoplasmic fraction (NP20-NP0). Co-transcriptional SED is significantly higher compared to post-transcriptional SED (p-value < 2.2e-16).

Figure 2. Different degrees of chromatin association correlate with distinct physical and functional characteristics of nascent RNA transcripts

(A) Transcript mean splicing efficiency. P-value < 2.2e-16 for fast- vs. slow-released mRNAs; non-significant (NS) for lncRNAs.

(B) Transcript minimum splicing efficiency (i.e. splicing efficiency of worst spliced intron). P-value < 2.2e-16 fast- vs. slow-released mRNAs, p-value 0.0408 fast- vs. slow-released lncRNAs.

(C) Metagene analysis of GRO-seq read coverage (only sense strand plotted) in a window -500 bp to +5 Kb around transcript end site (TES) (ContextMap extracted pA site).

(D) Metagene analysis of 'CHR0' sense strand-specific read coverage (chromatin-associated nascent RNA sequencing from time-point zero) in the window -1 Kb to +5 Kb around TES. The average read coverage per nucleotide position is normalized to the position with the maximum read coverage within each group, and plotted separately for lncRNAs (left panel) and mRNAs (middle panel). Right panel: Boxplot distribution (plotted in log scale) of transcriptional readthrough for the different groups, measured as the ratio of 'CHR0' sense strand-specific read coverage 5 Kb downstream to 1 Kb upstream of TES (p-value = 3.108e-12 fast vs. slow released lncRNAs and p-value < 2.2e-16 fast vs. slow released mRNAs).

449 (E) Metagene analysis of histone marks average profiles around the TSS of different groups of nascent RNA
450 transcripts (left panels: split lncRNAs vs. mRNAs, middle panels: split distinct degrees of chromatin
451 association i.e. fast/medium/slow-released lncRNAs and mRNAs). Right panels: Boxplot distribution of
452 promoter-associated histone mark signal around TSS (p-value < 2.2e-16 fast vs. slow-released lncRNAs;
453 NS for mRNAs).
454 (F) Metagene analysis of DNase-seq signal around TSS (left, middle panels) and the respective boxplot
455 distribution (right panel, p-value < 2.2e-16 fast vs. slow-released lncRNAs; NS for mRNAs).
456 (G) Promoter-overlapping ChIA-PET maximum scores (fast vs. slow released lncRNAs p-value 4.489e-10).
457
458 **Figure 3. Enhancer-associated elncRNAs are enriched in fast-released transcripts**
459 (A) Profile of nascent RNA transcription (GRO-seq) over bidirectionally transcribed enhancers in MCF-7.
460 (B) Cumulative plots of interdistances of transcript TSS to closest enhancer midpoint.
461 (C) Distribution of TSS interdistances (log10 bp) to closest enhancer midpoint for elncRNAs, mRNAs and
462 lncRNAs not associated to active enhancers.
463 (D) elncRNAs show significantly higher ChIA-PET interaction scores compared to mRNAs (p-value 0.0004281)
464 and to lncRNAs not associated with active enhancers (p-value 0.0002974).
465 (E) elncRNAs show significantly lower chromatin association halftimes (p-value 0.02329 to mRNAs and
466 5.116e-06 to rest lncRNAs; p-value 7.236e-05 rest lncRNAs to mRNAs).
467
468 **Figure 4. Contribution of distinct features to modeling chromatin (dis-)association of nascent RNA**
469 **transcripts**
470 (A) Logistic regression to predict fast vs. slow-released (chromatin-retained) lncRNAs. A 10x cross-validation
471 was applied (best AUC 0.9275, average 0.8891). Coefficients bigger than 0.3 or smaller than -0.3 and with
472 a p-value < 0.001 are marked red.
473 (B) Same as in (A) but for mRNAs (best AUC 0.84, average 0.81).
474
475 **Supplementary Figure 1.**
476 (A) Nascent RNA sequencing read coverage over the last exon from all pulse-chase time points.
477 (B) Same as in (A) but after splitting in the 3 groups of fast, medium and slow-released transcripts.
478 (C) K-means clustering of all analyzed transcripts (n = 18,837) using the chromatin association ratios with k = 3
479 defines 3 clusters corresponding (from top to bottom) to 'slow', 'fast', and 'medium'-released transcripts.

480 (D) Boxplot distribution of chromatin-association halftimes (calculated by fitting the exponential decay curve)
481 for the k-means clustering-derived groups.
482 (E) Distribution of transcript length (fast vs. slow mRNAs p-value < 2.2e-16, fast vs. slow lncRNAs p-value
483 2.387e-12).
484 (F) Left panel: Distribution of number of exons per transcript (p-value 0.003545 fast vs. slow mRNAs, NS for
485 lncRNAs). Right panel: Distribution of exon density (nr of exons per Kb) (p-value = 0.0009458 fast vs. slow
486 lncRNAs, p-value < 2.2e-16 fast vs. slow mRNAs).
487 (G) Chromatin association halftime (extracted by fitting the chromatin association ratios on an exponential
488 decay curve at p-value <0.05; Methods) for the three groups of fast, medium and slow released.
489 (H) Chromatin association ratios at steady state (log2 CHR/NP) for the 3 groups.
490 (I) Distribution of distances to closest antisense PCG TSS for the 3 groups of fast, medium and slow-released
491 lncRNA transcripts.
492 (J) Cumulative distribution function curves of lncRNA interdistances to closest antisense PCG TSS for the 3
493 groups of fast, medium and slow-released lncRNA transcripts.
494 (K) Cumulative distribution function plots of intron splicing efficiencies from all time points and samples, at
495 chromatin (left panel) and nucleoplasm (middle panel), and the respective boxplot distributions (right
496 panel).
497 (L) Boxplot distribution of normalized post-transcriptional splicing efficiency at intron (left panel) and transcript
498 level (middle and right panels; extracted as the mean normalized post-transcriptional splicing efficiency per
499 transcript).
500 (M) UCSC screenshot from the GATA3- GATA3-AS1 locus.
501
502 **Supplementary Figure 2.**
503 (A) Transcript mean psi value (left) and median (right).
504 (B) Pausing index for lncRNAs and mRNAs (p-value < 2.2e-16) measured by extracting the ratio of P-Ser2 Pol
505 II coverage 500 nt downstream of TSS to gene body.
506 (C) Same as in (B) but split for fast, medium and slow-released transcripts (fast vs. slow lncRNAs p-value
507 2.384e-08, NS for mRNAs).
508 (D) Pausing index using strand-specific GRO-seq read coverage (fast vs. slow released lncRNAs p-value
509 1.117e-15, NS for mRNAs).
510 (E) Strand-specific GRO-seq read coverage 1 Kb downstream of TSS (fast vs. slow lncRNA p-value = 6.185e-
511 13, NS for mRNAs).

512 (F) Metagene analysis of GRO-seq strand-specific read coverage for the different groups of RNA transcripts,
513 ±3 Kb around TSS.
514 (G) Transcription bidirectionality score extracted using GRO-seq (log2 antisense/sense read coverage 1 Kb
515 around TSS; fast vs. slow lncRNA p-value < 2.2e-16, NS for mRNA).
516 (H) Distribution of interdistances of ContextMap extracted pA site to annotated transcript 3' ends.
517 (I) Metagene analysis of average P-Ser2 Pol II density -500 bp to +5 Kb around TES of lncRNAs and mRNAs
518 (left panel), and boxplot distribution of the corresponding transcription termination indices (extracted as the
519 density ratio of 2.5 Kb downstream of TES to gene body (Methods); right panel).
520 (J) Metagene analysis of average GRO-seq read coverage profile (only the sense strand plotted) around the
521 TES of grouped RNA transcripts; raw (upper panels), and after normalization of the average profile to value
522 at nucleotide position zero (TES).
523 (K) Transcription termination index (NS)
524 (L) Travel index (fast vs. slow mRNAs p-value < 2.2e-16; fast vs. slow lncRNAs p-value 0.01248).
525
526
527 **Supplementary Figure 3.**
528 (A) Metagene profiles for different groups of transcripts (Id* label under panel C) of the average H3K4me1 to
529 H3K4me3 ratio in a window ±3 Kb around TSS (left panel), and boxplot distribution of the overall H3K4me1
530 to H3K4me3 ratio 2 Kb downstream of TSS (right panel, p-value < 2.2e-16 fast vs. slow-released lncRNAs,
531 p-value = 0.002769 fast vs. slow-released mRNAs).
532 (B) Average H3K9me3 (left) and H3K27me3 profiles around TSS of different groups of transcripts (Id* label
533 under panel C).
534 (C) Average POL2RA profiles around TSS of different groups of transcripts.
535 (D) Average CTCF profiles around TSS of different groups of transcripts (first three panels), and boxplot
536 distribution of CTCF enrichment ±1 Kb around TSS (fourth panel, p-value < 2.2e-16 fast vs. slow-released
537 lncRNAs, p-value 6.032e-08 fast vs. slow released mRNAs).
538 (E) Average profiles of YY1 binding probability ±1 Kb around the TSS of different groups of transcripts (left
539 panel, color Id* label under panel C), and the respective boxplot distributions of YY1 binding probability
540 ±1 Kb around TSS (right panel, p-value < 2.2e-16 fast vs. slow-released lncRNAs, NS for mRNAs).
541 (F) Boxplot distributions of promoter ChIA-PET score, extracted as the sum of scores of the ChIA-PET nodes
542 overlapping the promoter (±2 Kb TSS) (p-value 2.375e-09 fast vs. slow released mRNAs and 3.039e-11
543 fast vs. slow-released lncRNAs).
544

545 **Supplementary Figure 4.**

546 (A) Linear regression models (lm) run with 10 x cross-validation to predict chromatin association halftime (as a
547 continuous value) of lncRNAs (left panels) and mRNAs (right panels) by incorporating several parameters
548 (significant parameters with a coefficient p-value < 0.001 are red-marked).

549 (B) Two-class random forest run with 10 x cross-validation to predict fast vs. slow released lncRNAs (upper
550 panels, best model accuracy 0.91, mean accuracy 0.86) and mRNAs (lower panels, best model accuracy
551 0.81, mean accuracy 0.77).

552 (C) Two-class random forest run with 10 x cross-validation to predict fast vs. slow released lncRNAs (upper
553 panels, best model accuracy 0.808, mean accuracy 0.769) and mRNAs (lower panels, best model
554 accuracy 0.795, mean accuracy 0.776) by incorporating 100 RBP whole transcript binding probabilities
555 (*pysster* predictions). Mean Decrease Accuracy and Mean Decrease Gini values of the top best 30 factors
556 are shown.

557 (D) Boxplot distribution of whole transcript binding probabilities (*pysster* predictions) for factors important either
558 for predicting fast vs. slow-released lncRNAs (upper panels), fast vs. slow-released mRNAs (middle
559 panels) or both types (lower panels). Student's *t.test* p-values are noted (red for fast vs. slow lncRNAs and
560 blue for fast vs. slow-released mRNAs).

561

562 **Acknowledgements**

563 We thank Rutger Gjaltema and Edda Schulz for helpful discussions and comments on the manuscript. This
564 work was supported by the DFG Grant MA 4454/3-1 to A.M.

565

566 **AUTHOR CONTRIBUTIONS**

567 E.N. and A.M. conceived and planned the study with input from UAVØ. E.N. designed the computational and
568 experimental pipelines, performed experiments and computational analyses. S.B. implemented the *pysster*
569 method and contributed to data analyses. A.M. supervised computational analyses. E.N. and A.M. wrote the
570 manuscript.

571

572 **DECLARATION OF INTERESTS**

573 The authors declare no competing interests.

574

575

576

Methods

KEY RESOURCES TABLE

Reagent or Resource	Source	Identifier
MCF-7 Pser2 Pol II ChIP-seq	(Menafra et al., Plos One 2014)	GEO: GSM1388130
MCF-7 GRO-seq	(Franco et al., Genome Res 2018)	GEO: GSM2545179, GSM2545180, GSM2545181
MCF-7 Pol2RA ChIP-seq	https://www.encodeproject.org/	ENCFF663QKE
MCF-7 nuclear polyA+ RNA-seq	https://www.encodeproject.org/	ENCSR000CTO
H3K4me3	https://www.encodeproject.org/	ENCSR985MIB (GEO: GSM945269, ENCFF797IUA.bigWig)
H3K4me1	https://www.encodeproject.org/	ENCSR493NBY (GEO: GSE86714, ENCFF275KBS.bigWig)
H3K27Ac	https://www.encodeproject.org/	ENCSR000EWR (GEO: GSM945854 , ENCFF515VXR.bigWig
H3K9me3	https://www.encodeproject.org/	ENCSR999WHE (GEO: GSE96517, ENCFF191LDZ.bigWig)
H3K27me3	https://www.encodeproject.org/	ENCSR000EWP (GEO: GSM970218, ENCFF081UQC.bigWig)
MCF-7 CTCF ChIP-seq	https://www.encodeproject.org/	ENCSR000AHD (GEO: GSM1010734, ENCFF991NDB.bigWig)
YY1 Avocado imputation (signal p-value)	https://www.encodeproject.org/	ENCSR678ZGZ (ENCFF065FZS.bigWig)
MCF-7 ChIA-PET	https://www.encodeproject.org/	GEO:GSM970209
FANTOM5/NET-CAGE enhancers	Hirabayashi et al., 2019	https://fantom.gsc.riken.jp/5/suppl/Hirabayashi_et_al_2019/
SNRNP70 ChIP-seq	https://www.encodeproject.org/	ENCFF346UDN
eCLIP	https://www.encodeproject.org/	ENCSR456FVU

Software and Algorithms	Source	Identifier
ContextMap	Bonfert et al., 2017	https://www.bio.ifi.lmu.de/software/contextmap
Bedtools	Quinlan and Hall, Bioinformatics 2010	https://bedtools.readthedocs.io/en/latest/

609 Pysster Budach and Marsico, Bioinformatics 2018 https://github.com/budach/pysster

610 STAR Dobin et a., Bioinformatics 2013 https://github.com/alexdobin/STAR/releases

611 UCSC tools (bigWigAverageOverBed)

612

613 **METHOD DETAILS**

614 **Extraction of transcript 3' end site (TES)**

615 We ran ContextMap v2.7.9 on paired-end MCF-7 nuclear polyA+ data (ENCODE) using Bowtie2 aligner and
616 Bowtie2-build-l indexer, with parameters -mismatches 3 -seed 30 -maxhits 10 --polyA -t 8 -Xms4000M -
617 Xmx30000M. This generated 39,991 ContextMap scored polyA sites. Nearby polyA sites were clustered with
618 bedtools cluster –s –d 10, keeping the one with maximum score. Annotated transcript 3' ends were assigned a
619 ContextMap polyA site by fetching the closest with bedtools closest –s.

620

621 **Enhancer-associated lncRNAs in MCF-7**

622 From the FANTOM5/NET-CAGE enhancers (n = 85,786) we extracted the ones that show evident bidirectional
623 transcription in MCF-7 using GRO-seq (GSE96859) (bigWigCoverageOverBed mean0 coverage > 0.1 for both
624 strands), resulting in 10,008 bidirectional actively transcribed enhancers. We then fetched closest transcript start
625 site (TSS) to enhancer midpoints using bedtools closest –s and defined lncRNAs with an interdistance <
626 2000 bp as elncRNAs (n = 248 out of the 2077 analyzed).

627

628 **SNRNP70 occupancy over transcription units**

629 As a proxy we used SNRNP70 ChIP-seq from HepG2 and by intersecting the intervals corresponding to full-
630 length transcripts with ChIP-seq narrow peaks (ENCFF346UDN) we extracted a mean binding score per
631 transcription unit.

632

633 **Nascent RNA sequencing combined with pulse-chase and chromatin fractionation**

634 MCF-7 cells were seeded in P10 (6 plates per time point) and grown to ~80% confluency in 5% FCS, then
635 labeled for 8 min with 1mM 4-thio-Uridine (4-SU). Cells were either immediately harvested (lifted intact in ice-
636 cold PBS) or washed twice in PBS and chase was applied for 5, 10, 15, 20 min in 10 mM uridine diluted in
637 growth medium. Chromatin fractionation was performed as in ref[9]. Briefly cells were lysed in 400 ul lysis buffer
638 0.15% NP-40 and lysate was loaded on 800 ul sucrose buffer for brief centrifugation. Pelleted nuclei were
639 washed in ice-cold PBS, resuspended in 200 ul glycerol buffer and lysed in 0.6 M urea to fractionate chromatin
640 from the nucleoplasmic fraction. RNA from the chromatin and nucleoplasmic fraction was extracted with acidic

641 phenol (pH 4.5) and acidic phenol/chloroform. 3 ug of RNA were fragmented with 0.15 M NaOH final
642 concentration for 25 min on ice. Prior the RNA fragmentation, 0.15 ng of the 4-SU-labeled and unlabeled spike-
643 ins mix (as in the TT-seq protocol[15]) had been added to the 3 ug of RNA. The fragmentation reaction was
644 stopped in 10 mM Tris pH 7.4, purified with RNeasy MinElute Spin columns and eluted in 45 ul TE buffer (Tris
645 10 mM pH 7.4, 1mM EDTA). 5 ul Biotin-HPDP/DMF 1 mg/ml were added (i.e. final concentration 0.1 mg/ml) and
646 incubated for 2 hours at room temperature. Further steps of RNA purification, binding to T1 Dynabeads,
647 washing and elution were done according to the A. Regev protocol[41] (using 5 ug T1 Dynabeads for 2 ug 4-SU-
648 biotinylated RNA), leading to library construction for Illumina sequencing.
649
650 **Mapping and spike-ins normalization**
651 Reads were mapped to GRCh38 (gencode.v23.primary_assembly.annotation) and to ERCC92 sequences using
652 STAR 2.5.4a with standard parameters. Only reads mapped to a single genomic location were kept (score 255).
653 Three labeled (ERCC00043, ERCC00092, ERCC00136) and three unlabeled spike-ins (ERCC00002,
654 ERCC00145, ERCC00170) had been added to each RNA sample. For each sample a 'sizeFactor' was
655 extracted for spike-ins normalization as follows: each of the three labeled spike-ins read counts were normalized
656 to the sum of the respective spike-in counts across the ten labeled samples (CHR 0, 5, 10, 15, 20 min and NP
657 0, 5, 10, 15, 20 min), and then the median value from the three normalized labeled spike-ins was extracted per
658 sample ('smoothened median' = sizeFactor). For each labeled sample the cross-contamination value 'epsilon'
659 was calculated as the sum of unlabeled spike-in read counts (U) to the sum of U plus the sum of labeled spike-
660 in read counts (L): epsilon = cross-contamination = U/ (L+U). Strand-specific read counts over features were
661 normalized to sizeFactor and feature length and multiplied by (1-epsilon). Fitted_counts = measured_counts/
662 sizeFactor(labeled_sample)/ feature_length * (1-epsilon), or Fitted_counts = measured_counts/
663 sizeFactor(labeled_sample)/ feature_length * L/ (L+U).
664
665 **Transcript dataset**
666 We used GENCODE V29 lncRNA annotation (n = 8,992) supplemented with novel (non-overlapping GENCODE
667 V29 lncRNAs annotation) lncRNA transcripts from *de novo* transcript assembly (n = 10,606) on chromatin-
668 associated RNA-seq in MCF-7 (described in ref[9]; those are lacking protein-coding potential, are not overlapping
669 protein coding genes, and have at least 1 splice junction). From this initial set we kept 3,671 lncRNAs with non-
670 zero read coverage in all 12 sequenced samples. We also used 15,166 mRNA transcripts with non-zero read
671 coverage in all 12 samples.
672

673 **Modeling chromatin dissociation**

674 Strand-specific read counts over the last exon of the 18,837 transcripts were normalized to spike-ins and feature
675 length (as described in the Methods section 'Mapping and spike-ins normalization'). For each pulse-chase time
676 point we extracted a ratio of chromatin (CHR) to chromatin plus nucleoplasmic (NP) normalized read coverage
677 (CHR/ (CHR+NP)). We fit those ratios on an exponential decay using R function lm (log (x) ~ time), for
678 timepoints [0, 8, 13, 18, 23, 28] (ratio set to 1 at timepoint 0), which returns *intercept*, *k* and *p-value* of
679 exponential decay fit. We kept 12,391 entries that fit the curve with a p-value <0.05 (of which 2077 lncRNAs,
680 and 10,314 mRNAs). We defined a 'chromatin association halftime' as -(*intercept* + log (2)) / *k*. Based on the
681 halftime values, we split the dataset in three equal-size quantiles corresponding to 'fast', 'medium' and 'slow'
682 released nascent RNA transcripts.

683

684 **Splicing efficiency, SED and degree of post-transcriptional splicing**

685 We measured intron splicing efficiency (SE or *thita* value) as in ref[20] by extracting the ratio of split to split plus
686 non-split reads overlapping 3' splice sites of introns with at least one split and one non-spit read at the 3' splice
687 site (n = 154,467 high-confidence introns). We measured alternative splicing as in ref[21] by extracting the ratio
688 (*psi* value) of alternative split to constitutive split reads covering the high-confidence introns. We extracted co-
689 and post-transcriptional splicing efficiency dynamics (SED) as in ref[21], by subtracting the difference of splicing
690 efficiency at 20 min pulse-chase from the splicing efficiency at 0 min and normalizing this to the splicing
691 efficiency at 0 min [*SED = (SE_20min + 0.001 – SE_0min) / (SE_0min + 0.001)*]. We extracted the extent of
692 post-transcriptional splicing relative to co-transcriptional as the difference of chromatin-associated splicing
693 efficiency from the nucleoplasmic splicing efficiency, normalized to chromatin. This was done at intron and
694 transcript level (mean value of the transcript's high-confidence introns).

695

696 **Transcriptional indices (TSS-proximal pausing index and termination index)**

697 We assessed transcriptional pausing index by extracting the ratio of strand-specific GRO-seq read coverage or
698 P-Ser2 Pol II ChIP-seq density in the window 500 nt downstream of TSS to the gene body. Gene body was
699 defined as the middle 50% of the interval TSS+500 to TES, as in ref[20]. Transcription termination index was
700 measured as in in ref[20] by extracting the length-normalized ratio of strand-specific GRO-seq read coverage (or
701 Pol II ChIP-seq read density) in the window 2.5 Kb downstream of TES to gene body. Travel index was
702 extracted as the ratio of read coverage in the interval [2.5 to 5 Kb] downstream of TES to the first 2.5 Kb
703 downstream of TES.

704

705 **Machine learning models**

706 Logistic regression to predict fast versus slow-released nascent RNA transcripts was ran on standardized
707 parameters (R function *stdize()* of the package '*pls*') using the R function *glm()* and ten-times cross-validation.
708 Linear regression to model chromatin-association halftime as a continuous value was ran on standardized
709 parameters using R function *lm()* and ten-times cross-validation. Random forest to predict fast versus slow-
710 released nascent RNA transcripts was ran with R function *randomForest()* and ten-times cross-validation,
711 setting number of trees 1000 (*ntree* = 1000) and dataset-specific best *mtry* parameter. Best *mtry* was found
712 using the function train() of the package 'caret' with a grid-search and ten-times cross-validation.

713

714 **RBP predictions (build *pysster* models and prediction scan summary)**

715 To train *pysster* models we used ENCODE available eCLIP data from HepG2 cell line for 100 RNA-binding
716 proteins (2 biological replicates). eCLIP peaks found in both biological replicates and with a log-fold enrichment
717 > 2 over the input control were selected (5' end of peaks are used as binding sites from now on). *Pysster* was
718 used to train a multi-class convolutional neural network (CNN) classifier. We trained one model for each RBP,
719 and each model was trained on 3 classes:

720 - class 1: sequences of length 400 centered at a binding site of the protein of interest

721 - class 2: randomly sampled sequences of length 400 from lncRNAs (lncRNA models) or mRNAs (mRNA
722 models) that contain at least one binding site of the protein of interest (sequences were sampled such that they
723 don't overlap with class 1 though)

724 - class 3: sequences of length 400 centered at randomly selected binding sites of all other proteins to reduce
725 the impact of eCLIP bias signal (no overlap with class 1 again)

726 In addition to the sequences itself, the CNNs also use the following additional data as input: (1) is sequence
727 position 200 located in an exon or intron? (zero/one encoded), (2) distance of sequence position 200 to the
728 TSS/TTS (normalized to the transcript length such that zero indicates overlap with the TTS and one overlap with
729 the TSS). For each model a hyperparameter grid search was performed: 3 convolutional layers, kernels of
730 length 12, 18 or 24 and 150 or 300 kernels per layer (all other *pysster* parameters were left at their defaults). A
731 trained RBP model could then be applied to a transcript of interest as follows: using a sliding window approach
732 (window size 400, step size 1) the score of belonging to class 1 was predicted for all bases of a transcript. All
733 predictions > 0.66 were selected and their median was computed.

734

735

736

References

1. Lai, F., Gardini, A., Zhang, A. & Shiekhattar, R. Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**, 399-403 (2015).
2. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461 (2014).
3. Gil, N. & Ulitsky, I. Production of Spliced Long Noncoding RNAs Specifies Regions with Increased Enhancer Activity. *Cell Syst* **7**, 537-547 e3 (2018).
4. Hirabayashi, S. et al. NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. *Nat Genet* **51**, 1369-1379 (2019).
5. Young, R.S., Kumar, Y., Bickmore, W.A. & Taylor, M.S. Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome Biol* **18**, 242 (2017).
6. Arner, E. et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010-4 (2015).
7. Tan, J.Y., Biasini, A., Young, R.S. & Marques, A.C. Splicing of enhancer-associated lincRNAs contributes to enhancer activity. *Life Sci Alliance* **3**(2020).
8. Vucicevic, D., Corradin, O., Ntini, E., Scacheri, P.C. & Orom, U.A. Long ncRNA expression associates with tissue-specific enhancers. *Cell Cycle* **14**, 253-60 (2015).
9. Ntini, E. et al. Long ncRNA A-ROD activates its target gene DKK1 at its release from chromatin. *Nat Commun* **9**, 1636 (2018).
10. Werner, M.S. & Ruthenburg, A.J. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *Cell Rep* **12**, 1089-98 (2015).
11. Werner, M.S. et al. Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. *Nat Struct Mol Biol* **24**, 596-603 (2017).
12. Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet* **21**, 102-117 (2020).
13. Ntini, E. & Marsico, A. Functional impacts of non-coding RNA processing on enhancer activity and target gene expression. *J Mol Cell Biol* **11**, 868-879 (2019).
14. Yin, Y. et al. U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* **580**, 147-150 (2020).
15. Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225-8 (2016).
16. Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775-89 (2012).
17. Drexler, H.L., Choquet, K. & Churchman, L.S. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Mol Cell* **77**, 985-998 e8 (2020).
18. Ntini, E. et al. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* **20**, 923-8 (2013).
19. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. & Sharp, P.A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360-3 (2013).
20. Schlackow, M. et al. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell* **65**, 25-38 (2017).
21. Louloupi, A., Ntini, E., Conrad, T. & Orom, U.A.V. Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. *Cell Rep* **23**, 3429-3437 (2018).
22. Mele, M. et al. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res* **27**, 27-37 (2017).
23. Franco, H.L. et al. Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. *Genome Res* **28**, 159-170 (2018).
24. Zuckerman, B. & Ulitsky, I. Predictive models of subcellular localization of long RNAs. *RNA* **25**, 557-572 (2019).
25. Bonfert, T. & Friedel, C.C. Prediction of Poly(A) Sites by Poly(A) Read Mapping. *PLoS One* **12**, e0170914 (2017).
26. Saldi, T., Cortazar, M.A., Sheridan, R.M. & Bentley, D.L. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J Mol Biol* **428**, 2623-2635 (2016).
27. Rigo, F. & Martinson, H.G. Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. *RNA* **15**, 823-36 (2009).
28. Schreiber, J., Bilmes, J. & Noble, W.S. Completing the ENCODE3 compendium yields accurate imputations across a variety of assays and human biosamples. *Genome Biol* **21**, 82 (2020).
29. Weintraub, A.S. et al. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588 e28 (2017).
30. Van Nostrand, E.L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508-14 (2016).
31. Budach, S. & Marsico, A. pysster: classification of biological sequences by learning sequence and structure motifs with convolutional neural networks. *Bioinformatics* **34**, 3035-3037 (2018).
32. Misra, A. & Green, M.R. From polyadenylation to splicing: Dual role for mRNA 3' end formation factors. *RNA Biol* **13**, 259-64 (2016).

800 33. Dhir, A., Dhir, S., Proudfoot, N.J. & Jopling, C.L. Microprocessor mediates transcriptional termination of
801 long noncoding RNA transcripts hosting microRNAs. *Nat Struct Mol Biol* **22**, 319-27 (2015).
802 34. Eaton, J.D. & West, S. An end in sight? Xrn2 and transcriptional termination by RNA polymerase II.
803 *Transcription* **9**, 321-326 (2018).
804 35. Kopp, F. & Mendell, J.T. Functional Classification and Experimental Dissection of Long Noncoding
805 RNAs. *Cell* **172**, 393-407 (2018).
806 36. Engreitz, J.M. et al. Local regulation of gene expression by lncRNA promoters, transcription and
807 splicing. *Nature* **539**, 452-455 (2016).
808 37. Stojic, L. et al. Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its
809 transcriptional and product-related functions. *Nat Commun* **7**, 10406 (2016).
810 38. Cho, S.W. et al. Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell*
811 **173**, 1398-1412 e22 (2018).
812 39. Ntini, E. & Vang Orom, U.A. Targeting Polyadenylation for Retention of RNA at Chromatin. *Methods Mol*
813 *Biol* **2161**, 51-58 (2020).
814 40. Lubelsky, Y. & Ulitsky, I. Sequences enriched in Alu repeats drive nuclear localization of long RNAs in
815 human cells. *Nature* **555**, 107-111 (2018).
816 41. Rabani, M. et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation
817 dynamics in mammalian cells. *Nat Biotechnol* **29**, 436-42 (2011).
818

819 Preprints:

820 Tan and Marques, bioRxiv 2020. The activity of human enhancers is modulated by the splicing of their
821 associated lncRNAs. doi: https://doi.org/10.1101/2020.04.17.045971

822 Reimer, Mimoso, Adelman and Neugebauer, bioRxiv 2020. Rapid and Efficient Co-Transcriptional Splicing
823 Enhances Mammalian Gene Expression. doi: https://doi.org/10.1101/2020.02.11.944595

824 Coté et al., bioRxiv 2020. The spatial distributions of pre-mRNAs suggest post-transcriptional splicing of specific
825 introns within endogenous genes. doi: https://doi.org/10.1101/2020.04.06.028092

826 Johnsson et al., bioRxiv 2020. Transcriptional kinetics and molecular functions of long non-coding RNAs. doi:
827 https://doi.org/10.1101/2020.05.05.079251

Figure 1

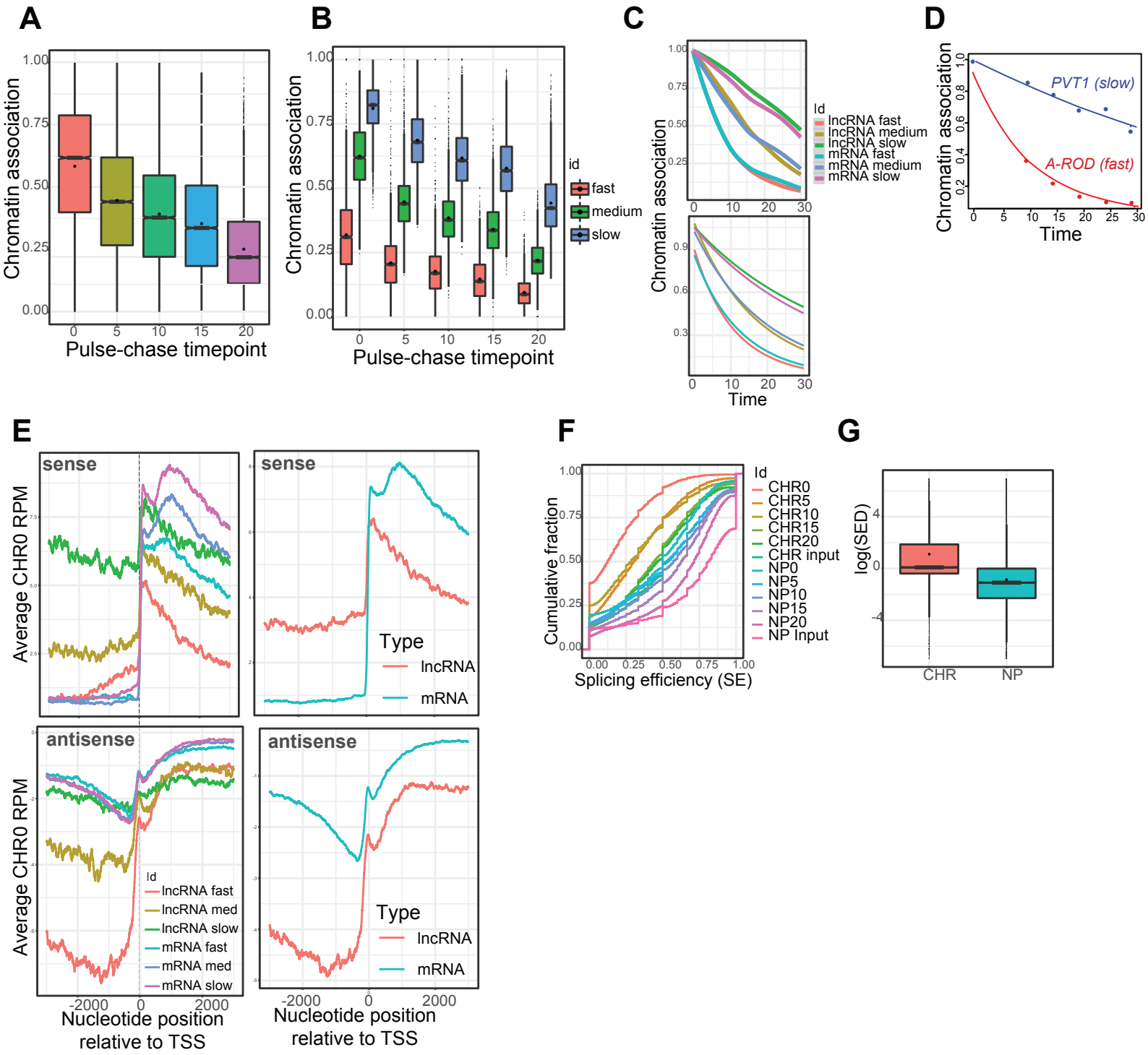

A
Chromatin association
Pulse-chase timepoint
B
Chromatin association
Pulse-chase timepoint
id
fast
medium
slow
C
Chromatin association
Time
Id
lncRNA fast
lncRNA medium
lncRNA slow
mRNA fast
mRNA medium
mRNA slow
D
Chromatin association
Time
PVT1 (slow)
A-ROD (fast)
E
sense
antisense
Average CHR0 RPM
Nucleotide position relative to TSS
Type
lncRNA
mRNA
Id
lncRNA fast
lncRNA med
lncRNA slow
mRNA fast
mRNA med
mRNA slow
F
Cumulative fraction
Splicing efficiency (SE)
Id
CHR0
CHR5
CHR10
CHR15
CHR20
CHR input
NP0
NP5
NP10
NP15
NP20
NP Input
G
log(SED)
CHR
NP

Figure 2

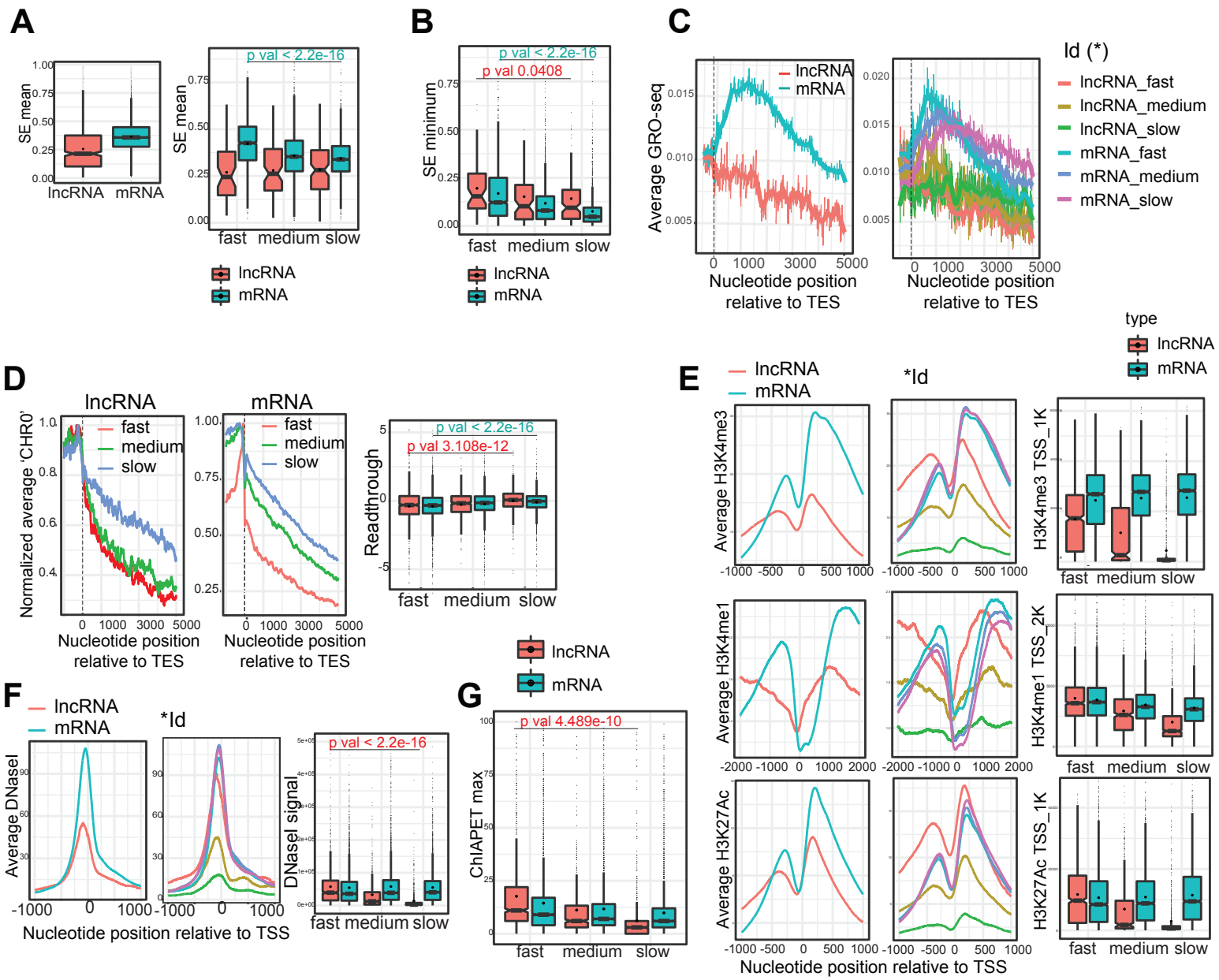

A
SE mean
lncRNA
mRNA
p val < 2.2e-16
fast
medium
slow
B
SE minimum
p val < 2.2e-16
p val 0.0408
C
Average GRO-seq
Nucleotide position relative to TES
Id (*)
lncRNA_fast
lncRNA_medium
lncRNA_slow
mRNA_fast
mRNA_medium
mRNA_slow
D
Normalized average 'CHR0'
Readthrough
p val 3.108e-12
E
*Id
type
Average H3K4me3
H3K4me3 TSS_1K
Average H3K4me1
H3K4me1 TSS_2K
Average H3K27Ac
H3K27Ac TSS_1K
Nucleotide position relative to TSS
F
Average DNaseI
DNaseI signal
G
ChIAPET max
p val 4.489e-10

Figure 3

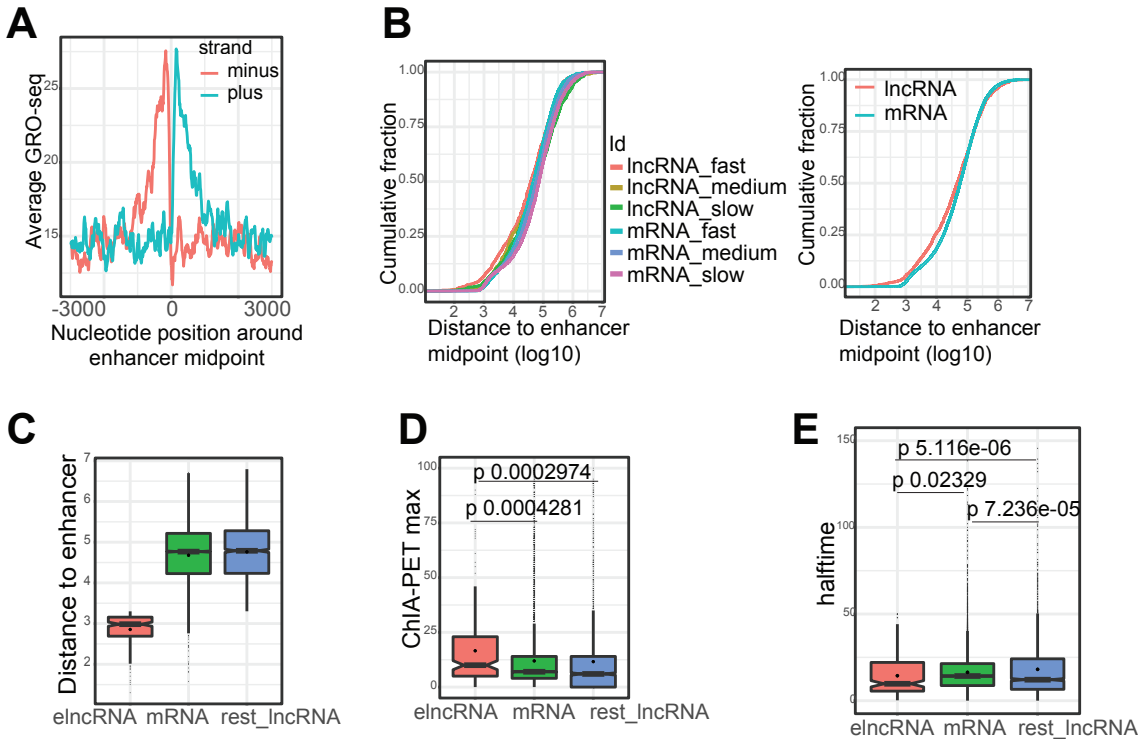
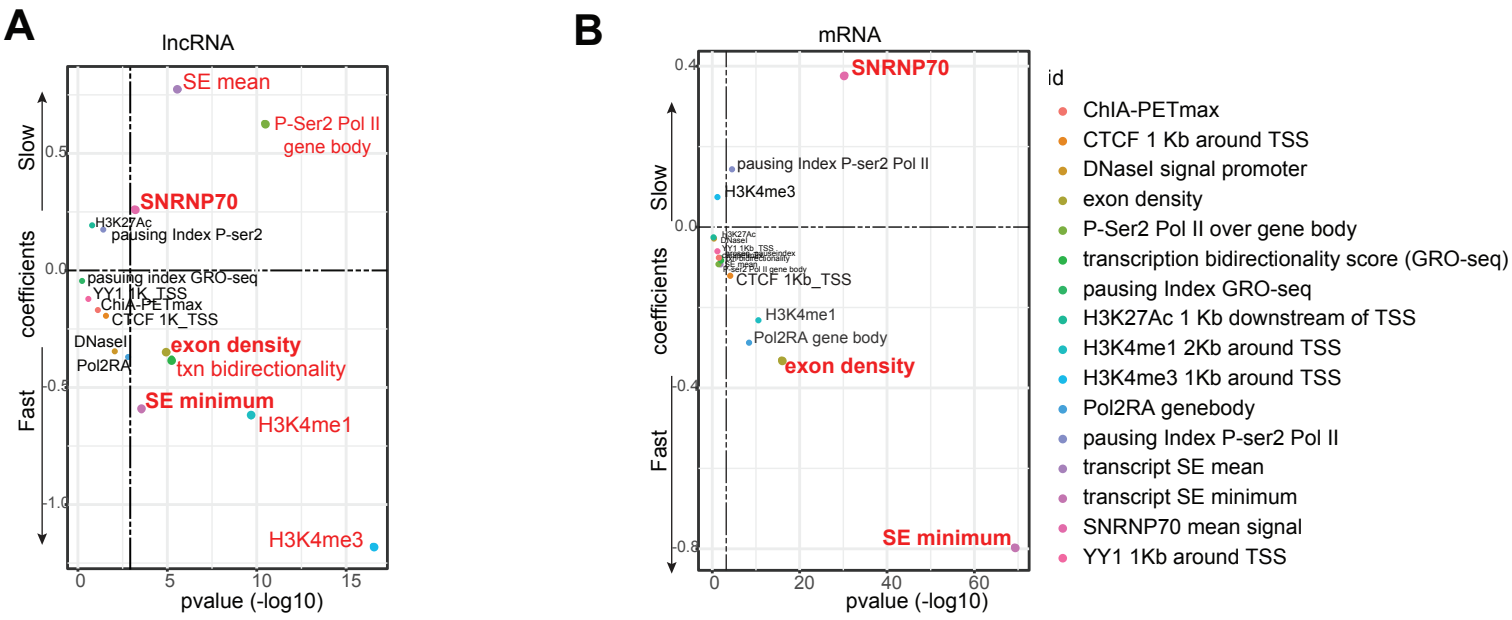

A
strand
minus
plus
Average GRO-seq
25
20
15
-3000
0
3000
Nucleotide position around
enhancer midpoint
B
Cumulative fraction
1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
Distance to enhancer
midpoint (log10)
Id
lncRNA_fast
lncRNA_medium
lncRNA_slow
mRNA_fast
mRNA_medium
mRNA_slow
lncRNA
mRNA
C
Distance to enhancer
elncRNA
mRNA
rest_lncRNA
D
ChIA-PET max
p 0.0002974
p 0.0004281
E
halftime
p 5.116e-06
p 0.02329
p 7.236e-05

Figure 4


A
lncRNA
SE mean
P-Ser2 Pol II
gene body
SNRNP70
H3K27Ac
pausing Index P-ser2
pasuing index GRO-seq
YY1 1K_TSS
ChIA-PETmax
CTCF 1K_TSS
DNaseI
Pol2RA
exon density
txn bidirectionality
SE minimum
H3K4me1
H3K4me3
Slow
Fast
coefficients
pvalue (-log10)
B
mRNA
SNRNP70
pausing Index P-ser2 Pol II
H3K4me3
CTCF 1Kb_TSS
H3K4me1
Pol2RA gene body
exon density
SE minimum
Slow
Fast
coefficients
pvalue (-log10)
id
ChIA-PETmax
CTCF 1 Kb around TSS
DNaseI signal promoter
exon density
P-Ser2 Pol II over gene body
transcription bidirectionality score (GRO-seq)
pausing Index GRO-seq
H3K27Ac 1 Kb downstream of TSS
H3K4me1 2Kb around TSS
H3K4me3 1Kb around TSS
Pol2RA genebody
pausing Index P-ser2 Pol II
transcript SE mean
transcript SE minimum
SNRNP70 mean signal
YY1 1Kb around TSS