

Genome-wide measurement of RNA dissociation from chromatin classifies transcripts by their dynamics and reveals rapid dissociation of enhancer lncRNAs

Highlights

- Chromatin dissociation dynamics of nascent RNA transcripts is measured with chrTT-seq
- Predictive modeling identifies distinctive features of lncRNA chromatin dissociation
- lncRNAs transcribed from enhancers display increased degrees of chromatin dissociation
- Fast chromatin-released lncRNAs display distinctive RBP-binding propensities

Authors

Evgenia Ntini, Stefan Budach,
Ulf A. Vang Ørom, Annalisa Marsico

Correspondence

evgenia.ntini@imbb.forth.gr (E.N.),
annalisa.marsico@
helmholtz-muenchen.de (A.M.)

In brief

Ntini et al. measure the chromatin dissociation dynamics of nascent RNA transcripts by coupling pulse-chase metabolic labeling with chromatin fractionation and deep sequencing. This study helps outline distinctive features of lncRNA chromatin dissociation dynamics and suggests the chromatin release of enhancer-transcribed lncRNAs.



Article

Genome-wide measurement of RNA dissociation from chromatin classifies transcripts by their dynamics and reveals rapid dissociation of enhancer lncRNAs

Evgenia Ntini,^{1,2,3,*} Stefan Budach,^{1,2} Ulf A. Vang Ørom,⁴ and Annalisa Marsico^{1,2,5,6,*}¹Max-Planck Institute for Molecular Genetics, 14195 Berlin, Germany²Freie Universität Berlin, 14195 Berlin, Germany³Institute of Molecular Biology and Biotechnology, IMBB-FORTH, 70013 Heraklio, Greece⁴Aarhus University, Department of Molecular Biology and Genetics, 8000 Aarhus, Denmark⁵Computational Health Center, Helmholtz Center Munich, Munich, Germany⁶Lead contact*Correspondence: evgenia.ntini@imbb.forth.gr (E.N.), annalisa.marsico@helmholtz-muenchen.de (A.M.)<https://doi.org/10.1016/j.cels.2023.09.005>

SUMMARY

Long non-coding RNAs (lncRNAs) are involved in gene expression regulation in *cis*. Although enriched in the cell chromatin fraction, to what degree this defines their regulatory potential remains unclear. Furthermore, the factors underlying lncRNA chromatin tethering, as well as the molecular basis of efficient lncRNA chromatin dissociation and its impact on enhancer activity and target gene expression, remain to be resolved. Here, we developed chrTT-seq, which combines the pulse-chase metabolic labeling of nascent RNA with chromatin fractionation and transient transcriptome sequencing to follow nascent RNA transcripts from their transcription on chromatin to release and allows the quantification of dissociation dynamics. By incorporating genomic, transcriptomic, and epigenetic metrics, as well as RNA-binding protein propensities, in machine learning models, we identify features that define transcript groups of different chromatin dissociation dynamics. Notably, lncRNAs transcribed from enhancers display reduced chromatin retention, suggesting that, in addition to splicing, their chromatin dissociation may shape enhancer activity.

INTRODUCTION

Long non-coding RNAs (lncRNAs) are a diverse group of molecules that exceed 200 nt in length and comprise the majority of the human transcriptome. Although lncRNAs have been linked to various biological processes and diseases, only a small fraction has been characterized at the molecular level.^{1–6} Although typically regarded as nuclear or chromatin enriched, lncRNAs demonstrate a dynamic and regulated distribution across all cell compartments and organelles.⁷ Understanding the principles of lncRNA subcellular localization, along with the compartment-specific networks of interactions of lncRNAs with proteins and other nucleic acids, can help elucidate mechanisms of lncRNA function.^{7–11} With the help of high-throughput techniques mapping the specific subcellular and subnuclear localization patterns of both coding and non-coding RNAs (ncRNAs) at a high resolution,^{7,12,13} recent studies have shed light on the sequences, as well as the *cis*- and *trans*-acting factors underlying lncRNA subcellular localization^{14–18}; however, these mainly focus on nuclear-versus cytoplasmic-enriched lncRNAs, whereas the molecular determinants of the chromatin versus nucleoplasmic (NP) enrichment of lncRNAs remain unclear.

Bidirectional transcription is a prominent characteristic of active enhancers, leading to the production of short-lived ncRNA transcripts termed eRNAs. eRNAs are short, non-spliced, potentially

terminated by the Integrator complex,¹⁹ and subjected to rapid exosome degradation.²⁰ This explains their observed chromatin enrichment and limited detection in whole-cell RNA sequencing (RNA-seq) data at steady state. eRNA production, measured by various techniques of nascent RNA-seq, along with DNase I hypersensitivity, distinct histone marks, (H3K27Ac and H3K4me1) and CBP/p300 binding, demarcate active enhancers.^{20–24} A small subset of bidirectionally transcribed enhancers produce a more stable and spliced lncRNA elongating in one direction,^{21,25,26} whereas a quarter of annotated lncRNAs overlap enhancer-like regions.²⁷ Enhancer-associated lncRNAs (elncRNAs) are linked to stronger enhancer activity, characterized by higher levels of nascent RNA transcription, H3K27Ac histone marks, and DNase accessibility, and play a crucial role in the regulation of gene expression in *cis* and in shaping local chromatin structure.^{21,25,26} elncRNAs are cell type specific and may shape transcriptional regulatory programs during development and cell differentiation, both in normal and pathological contexts. Thus, recent studies aim to uncover the mechanisms of elncRNA biogenesis and co-transcriptional processing.^{28–32} However, we still lack mechanistic insights into how elncRNAs regulate gene expression in *cis* and whether all elncRNAs follow a uniform mechanistic mode of action. It also remains unclear to what extent elncRNAs remain chromatin associated and whether their function depends on their chromatin association.



A substantial portion of lncRNAs are enriched in the chromatin fraction, presumably tethered at their sites of transcription through elongating (transcriptionally engaged) RNA polymerase II (RNA Pol II), and are involved in regulation of proximal gene expression in *cis*.^{2,33,34} In some cases, lncRNAs may function solely through the act of their transcription or splicing.^{2,3} It is plausible that lncRNAs that spend more time on chromatin underlie epigenetic functions by being involved in the regulation of three-dimensional genomic architecture and chromosomal organization, interacting with and recruiting chromatin modifiers and remodelers. Thus, the rate at which nascent RNA transcripts are released from chromatin impacts chromatin and gene function by determining the dynamic pool of lncRNAs that remains chromatin associated to mediate epigenetic functions. However, lncRNAs transcribed from the anchor points of chromosomal loops and enhancer-like regions show lower chromatin-to-NP enrichment at steady state.³⁵ This may indicate that the process of chromatin dissociation, which relies on co-transcriptional RNA maturation steps, could be important for the function of many enhancer-transcribed lncRNAs, acting in *cis* and within the spatial proximity of pre-established chromosomal loops.³⁶ In this line, we previously 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.³⁵ Enforcing *A-ROD* chromatin retention, by splice-inhibiting morpholinos or blocking 3' end formation, suppressed target gene expression, suggesting that chromatin dissociation is crucial for lncRNA-mediated gene expression regulation in *cis*.³⁵ Additional reports have shown that co-transcriptional splicing regulates the nuclear dynamics of nascent RNA transcripts.^{37–40} For instance, improperly processed transcripts susceptible to intron retention are subject to NP turnover.^{41,42} Compared with other lncRNAs that are not enhancer associated, elncRNAs display conserved splice sites and higher splicing density, which is associated with local changes in chromatin states and positively impacts their cognate enhancer activity.^{21,25} However, a link between elncRNA splicing and chromatin association (or dissociation) has not been clarified. Moreover, although recent bioinformatics analyses infer an impact of elncRNA processing on enhancer activity,^{21,25,26} the role of elncRNA chromatin (dis-) association has not been systematically examined.

Most of the existing models for predicting subcellular RNA localization^{43–46} (from either primary sequence or other genomic features) are based on steady-state expression levels, which ignore RNA processing dynamics that determine RNA fate. Although a few models focus on lncRNAs,^{17,47} they still rely solely on steady-state compartment-specific RNA-seq data. To fully understand lncRNA localization and its implication for functional potential, it is essential to model transcriptome-wide measurements of compartment-specific dynamics. In this study, we set out to understand how the rate at which nascent RNA transcripts are released from chromatin is regulated genome wide for both coding and non-coding transcripts. In particular, we aim at comprehensively characterizing, for the first time, the dynamic pool of lncRNAs that remain chromatin associated or are fast released from chromatin, their distinctive features compared with mRNAs, and their potential impact on gene regulation. To achieve this

goal, we have combined pulse-chase metabolic labeling with chromatin fractionation and transient transcriptome sequencing (TT-seq^{48,49}) to follow nascent RNAs from the point of their transcription to their chromatin release. This new method, termed chromatin-associated TT-seq (chrTT-seq), allows assessing co- and post-transcriptional RNA processing and quantitatively estimating chromatin dissociation rates of newly transcribed RNA transcripts. Using these data, we generate a first transcriptome-wide catalog of chromatin retention and dissociation dynamics for both mRNAs and lncRNAs in MCF-7 cells. Predictive models of chromatin (dis-) association rates, as measured by chrTT-seq, unravel the interplay among genomic, transcriptomic, and epigenetic features and RNA-binding protein (RBP) interactions that regulate chromatin dissociation dynamics and the subnuclear localization of all transcripts in general and of lncRNAs in particular. Our study identifies rules of lncRNA chromatin-association dynamics and unravels mechanistic aspects of a subgroup of lncRNAs, whose increased chromatin dissociation is linked to their regulatory functions genome wide, by shaping their cognate enhancer activity on target gene expression.

RESULTS

Modeling chromatin (dis-)association of nascent RNA transcripts

To quantify chromatin dissociation dynamics of coding and ncRNAs genome wide, we designed an experimental assay to follow nascent RNA transcripts from their synthesis to their post-transcriptional chromatin dissociation by employing nascent RNA-seq of both the chromatin-associated and chromatin-released (aka, NP) fractions. We performed 4-thiouridine (4-SU) metabolic labeling of MCF-7 cells for an 8-min pulse, followed by 5, 10, 15, and 20 min uridine chase (Figure 1A; STAR Methods). To additionally capture nascent RNA Pol II transcription in a high resolution and follow transcription dynamics, we fragmented RNA prior to the isolation of nascent RNA. Thus, in essence, we employed TT-seq⁴⁸ by coupling it for the first time with chromatin fractionation and pulse-chase metabolic labeling. We refer to this modified assay as chrTT-seq.

To model chromatin dissociation, we extracted read coverage from the last exon of each gene, selecting the longest transcript in the case of overlapping transcript isoforms (STAR Methods). Since we did not block new transcription initiation events during the pulse-chase experiment, extracting signal from the last exon minimizes read coverage due to new transcription initiation events during the pulse-chase time period. Thus, read coverage closer to the transcript 3' end better quantifies the dynamics of full-length transcripts (see also discussion). After spike-in normalization of read coverage for each library (STAR Methods), we observe that chromatin-associated read coverage decreases over time, whereas NP read coverage increases (Figure S1A). To quantify chromatin dissociation, we computed, for each transcript and each time point, the fraction of reads coming from the RNA-seq of chromatin-associated fraction versus the total number of reads ("CHR/(CHR + NP)"). As expected, we see an overall decrease in the transcript chromatin association over pulse-chase time (Figure 1B), indicating that the experiment captures the chromatin dissociation dynamics of nascent RNA transcripts. For each transcript, we fitted an exponential decay function to the

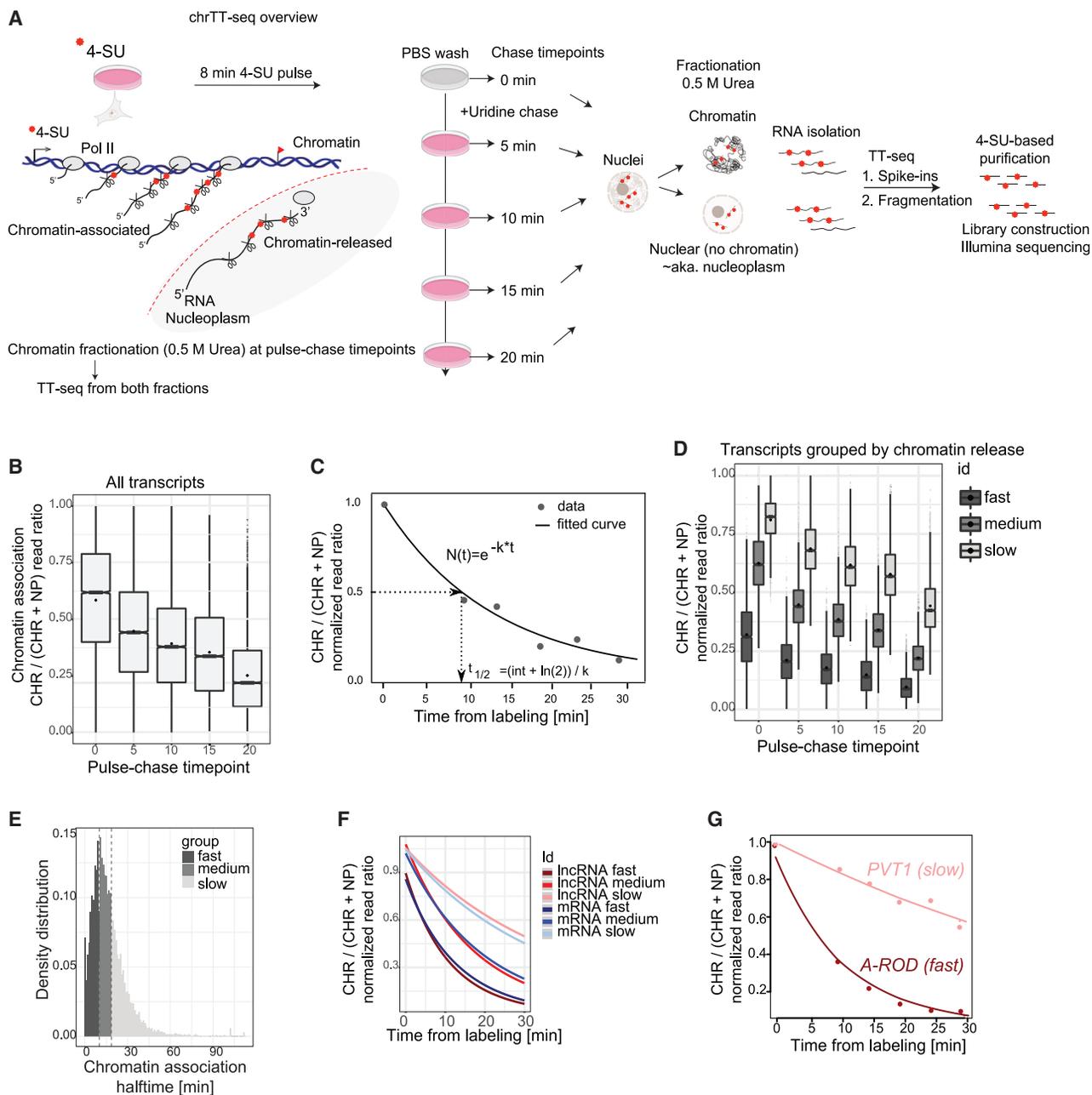


Figure 1. Measuring chromatin association of nascent RNA transcripts

(A) Schematic representation of chrTT-seq. Cells were labeled with 4-SU for 8 min, followed by chase with an excess of uridine for zero to 20 min. At each pulse-chase time point, isolated nuclei were lysed at 0.5 M urea, and TT-seq was performed from the chromatin-associated (“CHR”) and soluble chromatin-released (~nucleoplasmic [“NP”]) RNA fraction.

(B) Distribution of chromatin-association ratios (spike-in normalized CHR/(CHR + NP) read coverage) for all transcripts at different pulse-chase time points.

(C) Fit of an exponential decay curve to the chromatin-association ratios and estimation of chromatin-association halftime.

(D) Same as in (B) after split in fast-, medium-, and slow-released transcripts.

(E) Density distribution of chromatin-association halftimes for fast-, medium-, and slow-released transcripts.

(F) Loess curves of average chromatin association over time for the different transcript groups after fitting on an exponential decay curve.

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

CHR/(CHR + NP) ratios and extracted a “chromatin-association half-time,” that is, the time point when the chromatin-association read coverage ratio is reduced by 50% compared with the zero

time point (Figure 1C). By keeping only entries that fit the exponential decay curve at a $p < 0.05$, we estimated reliable chromatin-association half-times for 2,077 lncRNAs and 10,314 mRNAs (STAR

Methods). Based on the calculated chromatin-association half-time, we split the whole dataset into three equal-size quantiles corresponding to “fast,” “medium,” and “slow” chromatin-released transcripts (Figures 1D, 1E, and S1B). Transcripts were also clustered into three groups by k-means, which rendered a similar subgrouping (Figures S1C and S1D). Of note, during this short pulse-chase time frame (Figure 1A), only 222 transcripts showed a reduction in NP read coverage at the 20 min chase time point compared with the 0 min chase time point, indicating that the vast majority of transcripts are not being rapidly degraded or exported. Thus, our measurements of chromatin dissociation dynamics are not likely to be confounded by mechanisms of nuclear turnover, i.e., RNA degradation and export. Among the compared subgroups, chromatin fast-released mRNAs accumulated fewer NP reads, probably due to their efficient nuclear export (Figure S1E). This analysis suggests that transcripts with computed higher chromatin-association halftimes are not merely characterized as slow released due to a greater NP turnover (Figure S1E; supplemental information). Chromatin-association halftimes extracted with chrTT-seq reflect the chromatin-association ratios at steady state (Figures S1F and S1G), implying that the longer it takes for a transcript to be released from chromatin, the higher the chance for it to be enriched in the chromatin fraction at steady state. Albeit shorter in transcript length and with a smaller number of exons (Figures S1H and S1I; Derrien et al.⁵⁰), lncRNAs show, on average, slower chromatin dissociation (t test $p = 2.1e-5$; Figure S1J), mainly contributed by the slow-released subgroup (Figures 1F, S1D, and S1F). We identified *A-ROD* as a fast-released lncRNA, consistent with its efficient chromatin dissociation,³⁵ and *PVT1*⁵¹ as a slow-released lncRNA, in agreement with its observed chromatin retention at steady state (Figure 1G).

chrTT-seq captures transcriptional profiles and reveals major co-transcriptional RNA processing

chrTT-seq combines nascent RNA-seq with chromatin fractionation to track RNA Pol II transcriptional dynamics at high resolution. The application of a short metabolic pulse and RNA fragmentation prior to nascent RNA purification, as in the original TT-seq protocol,⁴⁸ combined with chromatin fractionation, further enriches nascent RNA reads.⁵² Nascent RNA-seq also captures promoter-associated divergent transcription producing short unstable antisense transcripts (PROMPTs).^{53,54} We noted that lncRNA loci produce higher upstream antisense transcription, compared with mRNAs, which extends beyond the typical PROMPT length (~200–600 nt) (Figure S2A). This is most probably because many lncRNAs arise upstream and antisense of protein-coding genes. Fast-released lncRNAs display stronger upstream antisense transcriptional signal, suggesting that they originate more often divergent to protein-coding genes. Indeed, fast-released lncRNAs display, on average, significantly smaller interdistance to the closest antisense protein-coding gene transcript start site (TSS) (t test $p < 2.2e-16$, Figure S2B). We observe that about half of the fast-released lncRNAs originate within less than 1 kb antisense to mRNA TSS (either upstream or internal antisense) (Figure S2C). An example is the fast-released lncRNA *GATA3-AS1* transcribed upstream and antisense of *GATA3* (Figure S2D). In agreement, lncRNAs annotated in ENCODE with the biotype “antisense” are enriched in

fast-released transcripts (odds ratio 1.47, $p = 4.2e-6$, Fisher’s exact test), whereas *de-novo*-assembled lncRNA transcripts from chromatin-associated RNA-seq data not overlapping ENCODE annotations (STAR Methods) are enriched in slow-released/chromatin-retained transcripts (odds ratio 2.07, $p = 1.9e-11$, Fisher’s exact test).

Nascent RNA-seq from the chromatin-associated and chromatin-released (aka, NP) fractions at different pulse-chase time points also allows for tracking the progress of co- and post-transcriptional splicing. To measure splicing, we used high-confidence introns (STAR Methods) and extracted splicing efficiency (SE) by calculating the ratio of split to non-split reads at the 3’ splice site (as in Schlackow et al.⁵⁵). By plotting the cumulative fraction of intron SEs 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 (Figures S2E and S2F). Moreover, co-transcriptional SE dynamics (SED) values of introns⁵⁶ (STAR Methods) were significantly higher compared with post-transcriptional SED (t test $p < 2.2e-16$, Figure S2G). These results are in agreement with recent reports that the majority of splicing occurs co-transcriptionally.⁵⁷ We also observe that introns of fast-released transcripts undergo the least additional post-transcriptional splicing upon chromatin dissociation (calculated as in STAR Methods and Figure S2H, left and middle), suggesting that most of their processing has been concluded co-transcriptionally while at chromatin. We observe that mRNAs show, overall, a higher degree of post-transcriptional processing than lncRNAs (t test $p = 1.8e-6$; Figure S2H, right). This is in agreement with recent findings that some post-transcriptional splicing can occur upon chromatin dissociation; after transcription is completed; and, potentially, while nascent RNA transcripts localize to speckles.¹²

Accurate prediction of transcript chromatin dissociation with machine learning models

We first set out to uncover factors that determine transcript chromatin dissociation dynamics in an RNA-biotype-agnostic manner by building a statistical model to predict chromatin dissociation dynamics based on genomic, transcriptomic, and epigenetic features (Table S1). We collected publicly available high-throughput datasets measuring chromatin modifications (histone marks), CTCF and YY1 binding, RNA Pol II occupancy, transcriptional pausing index (PI), and chromatin looping in MCF-7 cells (chromatin immunoprecipitation sequencing [ChIP-seq], GRO-seq, and ChIA-PET) (Figure 2A). We computed the enrichment of these features at promoters or gene body regions as appropriate (Table S1; STAR Methods). We also extracted the average and minimum splicing efficiency (SE) index per transcript (STAR Methods). We applied regularized logistic regression (elastic nets⁵⁸) to predict, from a total of 18 features, whether a transcript would be fast or slow released from chromatin (Figure 2A), yielding an average 10-fold misclassification error (MCE) of 26.9% (Figure 2B, left). We assessed the individual contribution of each feature to the classification accuracy by inspecting the values and signs of the regression coefficients (Figure 2B, right), where positive high values indicate features predictive of slow-released transcripts, whereas negative values correspond to features associated with fast-released transcripts (Figure 2B, right).

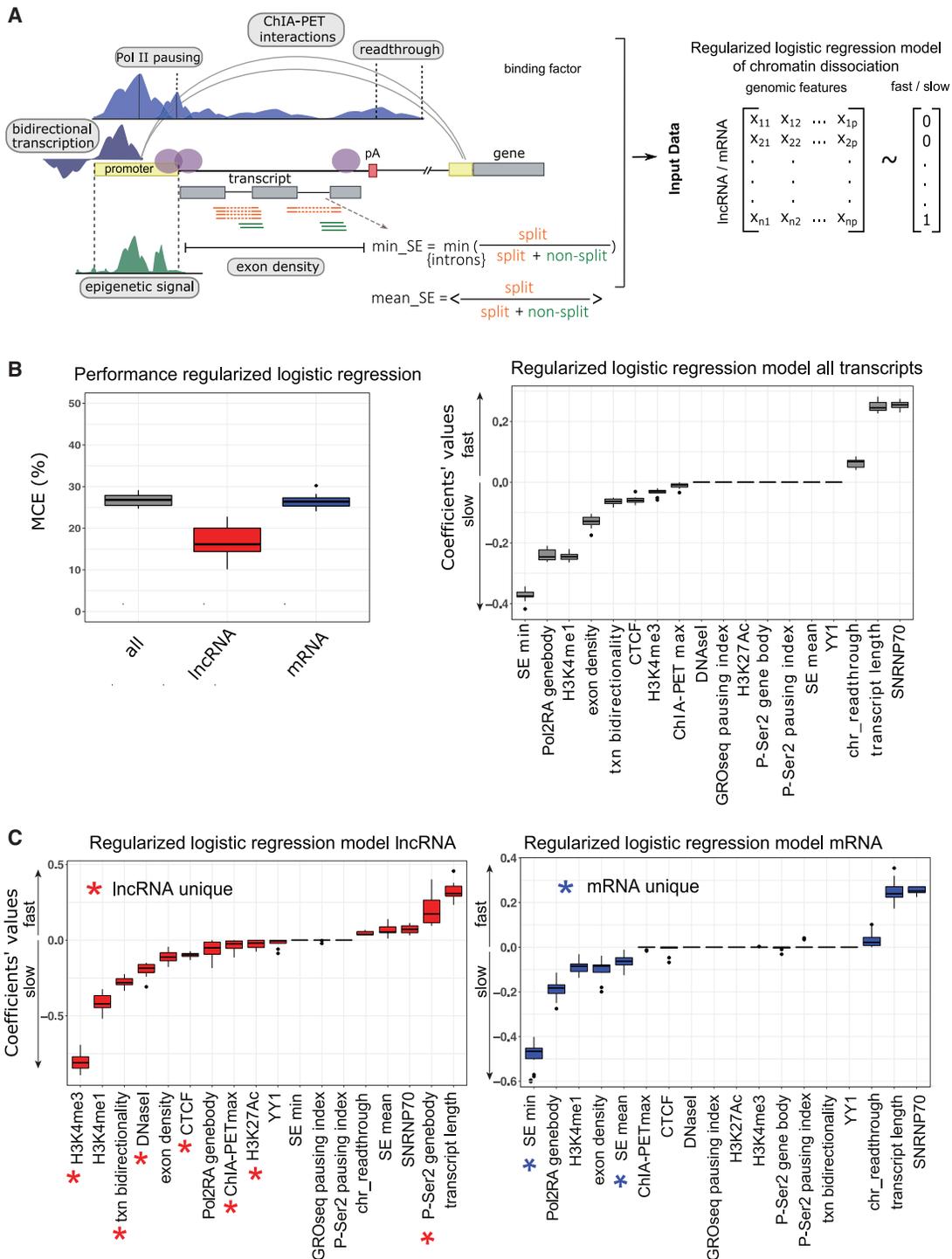


Figure 2. Contribution of distinct features to modeling chromatin (dis-)association of nascent RNA transcripts

(A) Schematic representation of the extracted features and modeling approach. Epigenetic and genomic input data for the model are collected, and feature matrices are computed for all transcripts (mRNAs and lncRNAs) with estimated chromatin-association half-lives. In the case of classification, a regularized logistic regression model (elastic net) model is trained to predict fast (class “0”) versus slow-released (class “1”) transcripts.

(B) Left: model performance for the regularized logistic regression model (elastic net misclassification error [MCE]) for all-transcript, lncRNA-, and mRNA-specific models. Right: feature average coefficient values (over 10× cross-validation) for the regularized logistic regression model (elastic nets) of all transcripts. Coefficient values >0 are associated with slow chromatin release and values <0 with fast chromatin release.

(C) Feature average coefficient values for lncRNA (left) and mRNA model (right). Highlighted with stars are features with non-zero coefficient values, which were important for either the lncRNA model only (red) or the mRNA model only (blue). Features with positive coefficient values contribute to chromatin tethering (slow chromatin release), whereas negative values contribute to efficient chromatin dissociation (fast chromatin release).

Our model corrects for transcripts length by using it as one of the predicting covariates in the elastic net. As expected, transcript length influences the dynamics of release from chromatin, since longer transcripts might take longer to conclude transcription; thus, they reside longer on chromatin, associated with their transcriptional template (Figure 2B). Besides transcript length, high small nuclear ribonucleoprotein 70 kDa (SNRNP70) enrichment over the transcription unit and RNA Pol II transcriptional read-through are the main determinants of slow-released transcripts. When it comes to fast-released transcripts, besides the exon density (which reflects overall splicing activity²¹), the H3K4me1 histone mark, RNA Pol II occupancy at the gene body, and a higher SE of the worst spliced intron (SE min) are among the most important predictive features. However, as 11% of the transcripts in our dataset are mono-exonic, by definition, a minimum and mean transcript SE (mean SE) could not be computed for those transcripts and were, therefore, imputed (STAR Methods). To better understand the impact of splicing-related features on chromatin dissociation dynamics, we excluded 870 transcripts with imputed values for those features from our dataset and rebuilt the model with multi-exonic transcripts only (Figure S3A). Although the most important predictive features of a model that excludes mono-exonic transcripts did not change, we found that a higher SE mean, in addition to high SE minimum, is important for predicting fast-released transcripts (Figure S3A), indicating that efficiently spliced transcripts have shorter chromatin-association halftimes.

In a second step, we investigated whether modeling lncRNA and mRNA chromatin-association dynamics separately yields more accurate predictions and reveals biotype-distinctive features. Although the average 25.6% MCE of a mRNA-specific elastic net model was comparable to that of the biotype-agnostic model (Figure 2B, left), a lncRNA-specific model of fast- versus slow-released transcripts yielded an average MCE of 16%, improving by more than 10% over the accuracy of a biotype-agnostic model (Figure 2B, left). This hints at the possibility that lncRNA chromatin dissociation dynamics are partly dictated by a set of unique genomic and functional features, which are less important for the chromatin dissociation of mRNAs. Our results were generally confirmed by a regularized linear regression model of the chromatin-association half-time (as a continuous value, Figures S3B–S3D) and by a two-class random forest (RF) model (Figure S4). We note that some outliers in the regularized linear regression models, with a computed half-time 1.5-fold higher than the predicted value, do not accumulate less nascent NP read coverage over time (Figure S3E), confirming that slow chromatin-released transcripts (either lncRNAs or mRNAs) with greater chromatin-association halftimes are not seemingly characterized as such due to faster NP losses (Figure S1E). In addition, RF feature importance analysis largely confirmed the important features identified by the logistic regression models in determining chromatin dissociation dynamics (Figure S4). All in all, our analysis confidently predicts chromatin dissociation dynamics for both lncRNAs and mRNAs.

Common features characterizing lncRNA and mRNA chromatin dissociation dynamics

Following the values of the coefficients from the logistic regression, which pinpoint the positive and negative feature contribu-

tions to the prediction of fast- versus slow- chromatin-released transcripts, we examined the identified important features more in detail from a mechanistic perspective.

Splicing activity affects the chromatin dissociation of both mRNAs and lncRNAs

lncRNAs show, on average, significantly lower co-transcriptional SE than mRNAs (t test $p < 2.2e-16$, Figure S5A, left), which is in agreement with previous studies using either steady-state or nascent RNA-seq data.^{55,59} Our biotype-agnostic model identifies the minimum SE per transcript (minimum SE) as the most predictive feature of fast chromatin dissociation (Figure 2B). This is almost entirely driven by the mRNA class (Figure 2C, right). To unravel whether a similar linkage between minimum SE and chromatin dissociation dynamics holds for lncRNAs, we assessed the relationship for multi-exonic lncRNAs, with a minimum of three exons (STAR Methods). Similar to mRNAs, we find that multi-exonic fast-released lncRNAs are associated with higher values of SE minimum, compared with their slow-released counterparts (Figure S5B). Together with exon density, which reflects splicing activity, these results imply that the splicing of a slowly or inefficiently processed intron, which is potentially a rate-limiting step for full transcript maturation,¹⁷ could act as a kinetic “bottleneck” for both nascent mRNA and multi-exonic lncRNA chromatin dissociation. We also find that lncRNAs show, on average, higher alternative splicing than mRNAs (Figure S5C), in agreement with previous findings.⁵⁹ In addition, chromatin-retained lncRNAs undergo significantly higher alternative splicing compared with fast-released transcripts, suggesting that extended times of chromatin association and/or suboptimal constitutive intron splicing could create space for stochastic splice site selection (Figure S5C).

Transcriptional activity affects chromatin dissociation dynamics

We then examined how chromatin dissociation dynamics is affected by transcriptional activity per se. For this purpose, we had included in the model several features relevant to transcriptional activity, such as total RNA Pol II occupancy at the gene body (“Pol2RA genebody”), transcriptionally engaged RNA Pol II (phosphorylated at Ser2 [P-Ser2] RNA Pol II), and promoter-proximal PI measured by P-Ser2 RNA Pol II and GRO-seq (STAR Methods). That total RNA Pol II occupancy (Pol2RA genebody) is a feature contributing to the fast release of both mRNAs and lncRNAs indicates that fast-released transcripts are produced from loci that are overall more transcriptionally active. It was also shown before that promoters of lncRNAs show distinct transcriptional burst kinetics compared with mRNAs (lower burst frequencies¹³) and that mRNAs display higher promoter-proximal RNA Pol II pausing than lncRNAs.^{17,55} Although RNA Pol II PI was not among the most important predictors of chromatin dissociation (Figure 2B), metagene profiles at promoters show an association between the fast release of lncRNAs, but not mRNAs, and PI (Figures S5D–S5F). In agreement, we find significantly higher levels of transcriptionally engaged RNA Pol II over the first kb downstream of TSS for fast- versus slow-released lncRNAs, but not mRNAs (Figures S5G and S5H). Thus, within lncRNAs, promoters of fast-released transcripts are more transcriptionally active and display a higher degree of RNA Pol II pausing compared with chromatin-retained lncRNA transcripts, suggesting that fast-released lncRNAs employ mRNA-like transcriptional characteristics.

Chromatin-retained transcripts exhibit higher U1 snRNP binding

Our model recovered the U1 SNRNP70 as enriched at loci associated with chromatin slow-released mRNAs and lncRNAs. This is in agreement with Yin et al.⁶⁰ suggesting U1-mediated chromatin retention of inefficiently processed transcripts, most probably due to unresolved spliceosomes.⁶⁰ That P-Ser2 RNA Pol II over gene body is also a strong predictor of slow-released lncRNAs (Figure 2B) points to slow-released lncRNAs being tethered to chromatin through transcriptionally engaged RNA Pol II, as previously suggested,³³ and shows that transcriptional activity could contribute to U1 snRNP-mediated tethering of inefficiently processed transcripts.⁶⁰ A similar SNRNP70 enrichment at slow-released mRNAs suggests that their chromatin association could be partly achieved through persistent U1 snRNP binding to inefficiently processed transcripts.

Readthrough beyond TESs is associated with chromatin-retained transcripts

Chromatin dissociation of nascent RNA transcripts is coupled with transcription termination and 3' end formation.^{37,61} We performed metagene analysis around the transcript 3' end site (TES) using ChIP-seq signal from transcriptionally engaged P-Ser2 RNA Pol II or strand-specific GRO-seq read coverage (STAR Methods). P-Ser2 RNA Pol II metagene profiles around TES resemble the ones obtained by mNET-seq,⁵⁵ revealing polyadenylation-associated RNA Pol II pausing in a 2 kb window downstream of the TES of mRNAs, but not lncRNAs (Figure S5I). In conjunction, mRNAs display significantly higher transcription termination index than lncRNAs, as previously reported⁵⁵ (Figures S5I–S5K). By measuring transcriptional readthrough and extracting an RNA Pol II travel index (STAR Methods), we note that RNA Pol II of slow-released transcripts tends to travel further beyond the polyadenylation-associated pausing site (Figures S5L and S5M), suggesting that ongoing transcription may contribute to chromatin tethering and slow release of nascent RNA transcript. Taken together with the observed low splicing activity (~exon density) at the loci of slow-released transcripts (Figures 2B and S1I), these results are in agreement with a cross talk among splicing, transcription, and transcription termination^{61,62} and with recent findings that inefficient splicing associates with readthrough transcription.⁵⁷

Distinctive features of lncRNA chromatin dissociation dynamics

Fast-released lncRNAs display significantly higher antisense (divergent) transcription

Our model highlights bidirectional transcription at promoters as an important feature for predicting efficiently chromatin-dissociated lncRNAs (Figure 2B). Extracting the bidirectionality score⁶³ either from GRO-seq (“txn bidirectionality” variable; STAR Methods) or from the chrTT-seq data at chase time point zero (“CHRO”) confirms that fast-released lncRNAs display significantly higher antisense (divergent) transcription (Figures S6A and S2A), and this is due to their enrichment in originating near and antisense of protein-coding gene TSS (Figures S2B and S2C).

Chromatin states demarcate the dissociation dynamics of fast-released lncRNAs

Our models suggest that distinct degrees of nascent lncRNA transcript chromatin association relate to distinct chromatin

states (Figures 2B and S4). Metagene analysis plots of histone marks associated with transcriptional activity (H3K4me3, H3K4me1, and H3K27Ac) confirm significant differences in the promoter regions of fast, medium, and slow-released lncRNAs, but not mRNAs (Figure S6B). By extracting the ratio H3K4me1 to H3K4me3 around the TSS, we observe that fast-released lncRNAs resemble mRNAs in terms of promoter activity (Figure S6C), whereas chromatin-retained lncRNAs display, on average, a higher H3K4me1 to H3K4me3 ratio (i.e., lower levels of H3K4me3 mark in accordance with the model; Figure 2B) and, on average, higher signals of repressive histone marks H3K9me3 and H3K27me3 (Figure S6D). We observe that fast-released lncRNAs are transcribed from regions with significantly greater chromatin accessibility (Figures S6E and S6F) and display significantly higher CTCF and YY1 binding (Figures S6G and S6H), which are factors associated with chromatin looping, promoting enhancer-promoter interactions.⁶⁴

elncRNAs do not remain chromatin associated

Promoters of fast-released lncRNAs display significantly higher ChIA-PET scores (Figures 2B and S6I), indicating that they tend to be transcribed from the anchor points of chromosomal loops. In general, lncRNAs transcribed from enhancer-like regions display on average higher promoter-associated ChIA-PET scores.³⁵ Although recent studies have started shedding light on the biogenesis and co-transcriptional processing mechanisms of enhancer-transcribed lncRNAs,^{28,32} the role of their chromatin dissociation dynamics remains unclear. We, therefore, examined the association of distinct degrees of lncRNA chromatin dissociation with enhancer activity in more detail. For this purpose, we used FANTOM5-^{20,24} and NET-CAGE²²-defined human “permissive” enhancers that are transcriptionally active in MCF-7 cells (STAR Methods), ending up with 10,008 high-confidence bidirectionally transcribed enhancers (Figure 3A). About 2.5% of bidirectionally transcribed enhancers had a lncRNA TSS within a genomic distance of 2 kb, consistent with previous reports.^{21,25} Thus, these lncRNAs can be regarded as elncRNAs²⁵ in MCF-7 cells, and their cognate enhancers as “lncRNA-associated eRNA-producing centers,” (la-EPCs).²¹ Notably, elncRNAs are enriched in fast-released lncRNAs (odds ratio ~1.86, $p = 0.0013$, Fisher’s exact test), whereas mRNAs at an interdistance <2 kb to the closest enhancer midpoint are not enriched in fast-released mRNAs (odds ratio 0.89, $p = 0.38$, Fisher’s exact test) (Figures 3B and S6J). This suggests that transcribed enhancers are more likely to be associated with a fast-released lncRNA, rather than a chromatin-retained one. In addition, elncRNAs (defined at an interdistance <2 kb to enhancer midpoint, Figure 3C) show significantly higher association with the anchor points of chromatin loops (Figure 3D). Although lncRNAs as a class display higher chromatin-association halftimes than mRNAs, elncRNAs escape this rule by showing significantly lower chromatin-association halftimes (Figure 3E), consistent with elncRNAs being enriched in fast-released transcripts. In conclusion, we show that enhancer-associated or rather enhancer-transcribed lncRNAs (elncRNAs) show increased degrees of chromatin dissociation, in addition to increased splicing.^{21,25,26}

Effect of lncRNA chromatin dissociation on the cis-regulation of target genes

We previously showed that lncRNAs transcribed from the anchor points of chromosomal loops precede their ChIA-PET-interacting

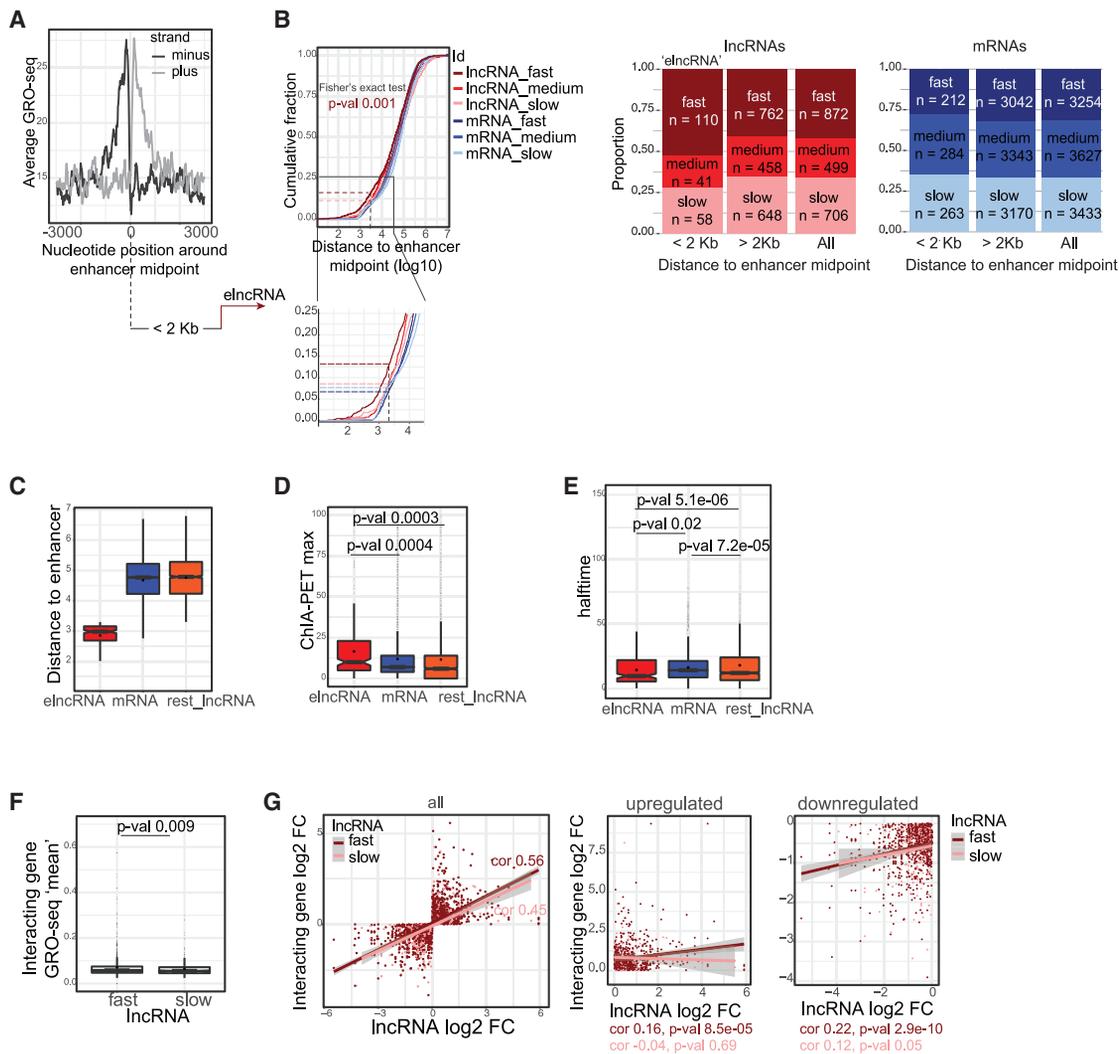


Figure 3. Enhancer-associated lncRNAs are chromatin dissociated

(A) Profile of nascent RNA transcription (GRO-seq)⁶⁵ over bidirectionally transcribed enhancers in MCF-7 (n = 10,008). eIncRNAs are defined at an interdistance less than 2 kb to enhancer midpoint.

(B) Cumulative plots of interdistances of transcript TSS to closest enhancer midpoint. Fast-released lncRNAs (dark red line) are significantly enriched at interdistances <2 kb to enhancer midpoint (odds ratio 1.86, Fisher’s exact test p = 0.0013). NS for mRNAs. The vertical dashed line marks the 2 kb interdistance cutoff. A zoomed-in version of a predefined portion of the main plot is depicted in the lower panel. Proportion graphs with transcript counts are provided in the right panel.

(C) Distribution of TSS interdistances (log₁₀ bp) to the closest enhancer midpoint for eIncRNAs, mRNAs and lncRNAs not associated with active (bidirectionally transcribed) enhancers (“rest” lncRNAs). eIncRNAs are defined at <2 kb from enhancer midpoint.

(D) eIncRNAs show significantly higher ChIA-PET scores compared with mRNAs (p = 0.0004) and to lncRNAs not associated with active enhancers (p = 0.0003).

(E) eIncRNAs show significantly lower chromatin-association half-times (t test p = 0.02 versus mRNAs and p = 5.1e−6 versus all other [rest] lncRNAs).

(F) Average GRO-seq-measured expression of RNA Pol II ChIA-PET-identified interacting genes, interacting with either fast- or slow- chromatin-released lncRNAs.

(G) Correlation plots between RNA Pol II ChIA-PET identified interacting genes and their analyzed lncRNAs, of log₂ fold changes in expression upon 40 min E2 treatment over control in MCF-7 cells. Left, all interacting gene-lncRNA pairs; middle and left panels, either up- or down-regulated.

genes in expression changes in response to estradiol (E2) treatment.³⁵ To unveil the role of efficient chromatin dissociation of eIncRNAs and/or lncRNAs transcribed from the anchor points of chromosomal loops on target gene expression, we measured the expression of target genes identified by ChIA-PET (as in Ntini et al.³⁵) in control and 40 min E2 treatment using available GRO-seq data^{65,66} in MCF-7 cells. As expected, the interacting target genes of fast- and slow-released lncRNAs tend to show

only negligible differences in expression in basal condition (control, EtOH treatment), suggesting that other features and/or locus-specific factors contribute to the regulation of target gene expression at steady state (Figure 3F). Although there is a positive correlation in expression changes between the interacting lncRNAs and target genes (cor = 0.54), the correlation is higher for fast- (cor = 0.57) than for slow-released lncRNAs (cor = 0.45) (Figure 3G), suggesting that lncRNA chromatin dissociation may

have a positive effect on the regulation of target gene expression in *cis*, underlying transcriptional responses.

Distinct RBPs dictate lncRNA and mRNA chromatin dissociation dynamics

We then asked whether RBPs, which control every aspect of RNA metabolism, play a role in mRNA and lncRNA chromatin dissociation dynamics and whether transcripts of different degrees of chromatin association would interact with distinct RBP complexes.

For this purpose, we used the ENCODE-available eCLIP data^{67,68} from HepG2 and K562 cells as a proxy dataset. As most lncRNAs are expressed in a cell-type-specific manner, we trained the *pysster*⁶⁹ algorithm, a convolutional neural network model for the classification of biological sequence, on transcripts' RBP binding sequences from the ENCODE eCLIP data. We then used the trained models, one for each RBP, to predict sequence-based RBP-binding propensities across full transcripts in MCF-7 cells (see [STAR Methods](#) for details and [Figures 4A and S7](#)). Utilizing the predicted RBP-binding probabilities as input features, we trained two separate RF models, one for lncRNAs and one for mRNAs, to classify fast versus slow chromatin-released transcripts. Because of the high number of training features and their high correlations (i.e., groups of proteins exhibiting similar binding patterns^{67,70}), we opted for the usage of a non-linear model, such as RF, which predicted chromatin dissociation dynamics with a 10-fold cross-validation mean accuracy of 0.77 and 0.78 for lncRNAs and mRNAs, respectively ([Figure 4B](#)).

Applying the lncRNA-specific RF RBP model to predict fast-versus slow- chromatin-released mRNAs results in a lower accuracy of ~ 0.54 . Similarly, applying the mRNA-specific RBP model to predict fast versus slow chromatin-released lncRNAs results in a lower accuracy of ~ 0.56 . This suggests that the performance of the RF models is biotype- (either lncRNA or mRNA) specific and that a subset of distinct RBPs might be implicated in defining fast- versus slow- lncRNAs or mRNAs. RBPs with high binding probabilities, found to be important for chromatin association of both lncRNAs and mRNAs, include factors with additional DNA binding activity (localizing to chromatin), similar to the KH-domain-containing factors KHSRP and KHDRBS1, FUBP3 and SUGP2 ([Figures 4B and S8](#)). These factors display increased binding probabilities for chromatin-retained transcripts, either lncRNAs or mRNAs ([Figure S8B](#)). CSTF2 involved in 3' end formation⁷⁴ is enriched in slow-released transcripts, perhaps reflecting persistent binding and unresolved RNA-protein complexes in the case of inefficient transcription termination and 3' end formation. The exosome component EXOSC5 is also enriched in slow-released transcripts, implying chromatin-associated clearance of inefficiently processed nascent RNA transcripts.⁷⁵ Among biotype-specific RBPs, DROSHA is a candidate significantly enriched in fast-released lncRNAs ([Figure S8B](#)). DROSHA was found to underlie pA-signal-independent transcription termination and 3' end formation of lncRNAs serving as miRNA hosts.⁷⁶ Yet, an increased RNA-binding probability of DROSHA, specifically in fast-released lncRNAs, could also suggest their post-transcriptional processing. Although we do not find any statistically significant enrichment of lncRNA miRNA hosts in the fast-released lncRNAs, a closer examination would be required to resolve microprocessor involvement in

lncRNA transcription termination (and 3' end formation) as an implicated mechanism. Additional lncRNA-specific factors with increased RNA-binding probabilities for fast-released lncRNAs are the nuclear cap-binding protein (NCBP)2, splicing factor NONO, helicase DDX3X also involved in splicing and RNA export, and XRN2 and CSTF2T, involved in transcription termination and 3' end formation.⁷⁷ Because NONO, XRN2, and CSTF2T have also DNA binding activity and localize to chromatin,⁷⁸ this suggests that their predicted binding could be co-transcriptional, and their activity may contribute to promoting chromatin dissociation of nascent lncRNA transcripts.

XRN2 promotes the chromatin dissociation of nascent lncRNAs

To examine the role of XRN2 in chromatin dissociation of nascent lncRNA transcripts, we used nascent RNA-seq data from HCT116 cells under normal and XRN2-depleted conditions⁷² generated via POINT-seq technology.⁷³ It is established that in Xrn2-dependent transcription termination, upon CPSF3-mediated co-transcriptional endonucleolytic cleavage at polyadenylation sites (PASs), the 5'-3' exonuclease activity of Xrn2 attacks the unprotected 5' end of the 3' RNA flanking fragment and pursues the transcriptionally engaged RNA Pol II, in the so-called "torpedo" model.⁷⁹ Overall, mRNAs employ Xrn2-dependent transcription termination⁷²; thus, it was previously shown that there is a substantial increase of readthrough transcriptional signal (measured by the termination index⁷³) upon 2 h of auxin-inducible XRN2 degradation in HCT116 cells.⁷³ To analyze the effect of XRN2 on lncRNA transcriptional readthrough, a subset of 1,908 (out of 2,077) lncRNAs expressed in HCT116, was analyzed (after removing lncRNAs overlapping up to 2 kb downstream of TES with other highly expressed transcription units).

Assuming that chromatin dissociation dynamics are similar in the two cancer cell lines, we used chromatin dissociation rates calculated in MCF-7 as a proxy for the same transcripts in HCT116 to examine the relationship between XRN2-mediated termination and chromatin dissociation dynamics. Of note, there is a high correlation in expression levels and SEs ($R = 0.8$ and $R = 0.85$, respectively) between the two cell lines motivating this assumption ([Figures S9A and S9B](#)). Fast-released lncRNAs show significantly higher changes in the termination index upon XRN2 depletion compared with their slow-released counterparts ($p = 9.5e-6$, t test, [Figure 4C](#)), whereas there are no significant differences in XRN2-mediated effect on readthrough transcriptional activity among distinct groups of mRNAs ([Figure 4C](#)).

This is in agreement with the RBP RF model suggesting XRN2 as a factor promoting efficient chromatin dissociation of lncRNAs, with enriched XRN2 binding probabilities among fast-released lncRNA transcripts ([Figure S8B](#)). Metagene analysis, profiling POINT-seq data⁷³ around the TES of lncRNAs and mRNAs in HCT116 cells, shows that fast-released lncRNAs have, on average, higher XRN2 depletion-to-control signal ratio (+auxin [IAA] 2 to 0 h), suggesting that efficiently chromatin-dissociated lncRNAs are characterized by a higher XRN2 dependency in transcription termination ([Figures 4D and S9C](#)). This is also supported by the finding that transcriptional readthrough is predictive of slow-released transcripts ([Figures 2B, S4B, and S5L](#)).

Taken together, these results suggest that slow-released lncRNAs do not utilize Xrn2-dependent transcription termination.

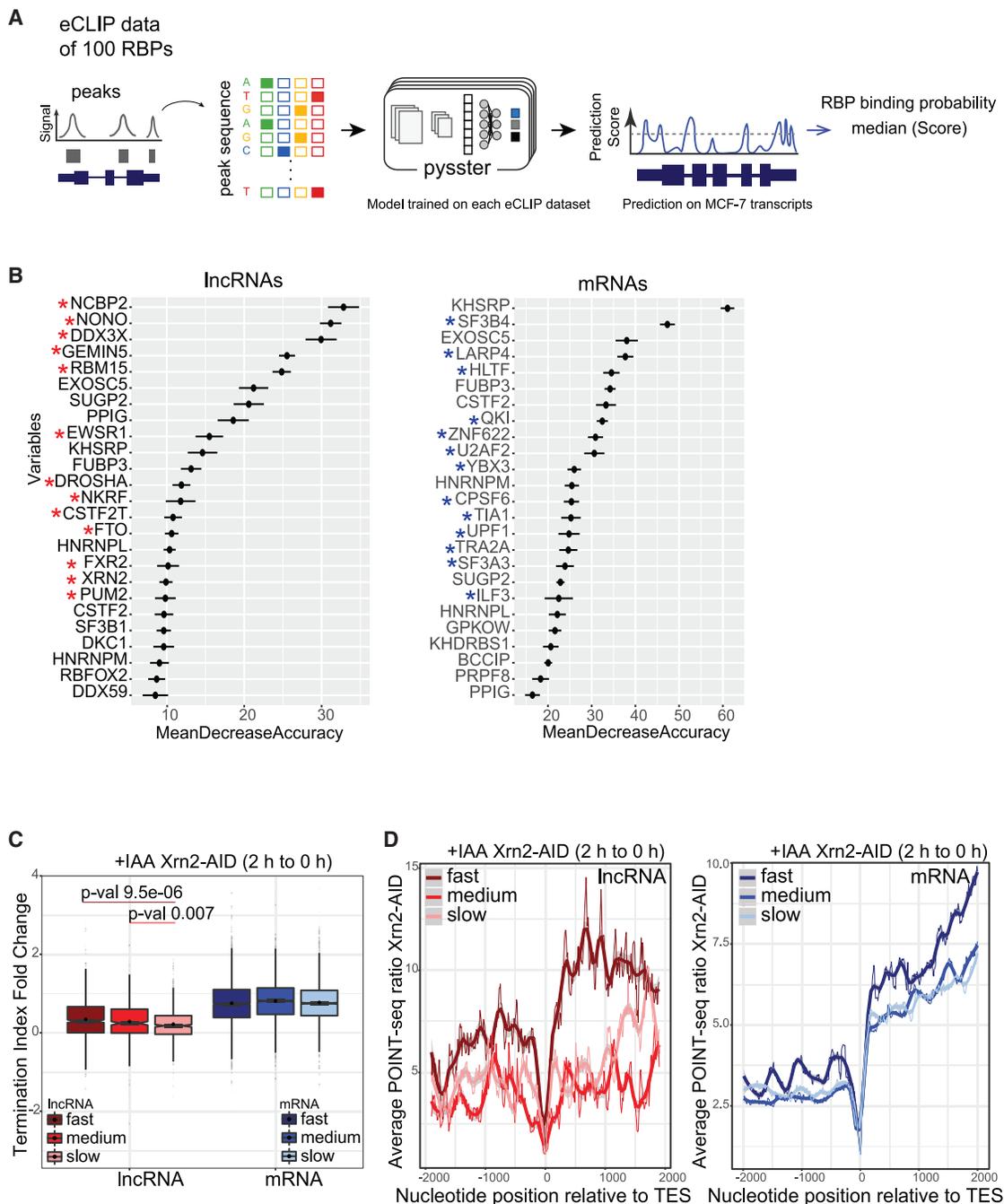


Figure 4. Modeling chromatin dissociation using RBP-binding probabilities

(A) Schematic representation of extracting RNA-binding probability per transcript per RBP using ENCODE eCLIP data and pysster⁶⁹; transcript binding probability for a given RBP is the median of prediction scores from positions above a certain cutoff (STAR Methods). Scheme is modified from Horlacher et al.⁷¹

(B) Two-class random forest model run with 10× cross-validation to predict fast- versus slow-released IncRNAs (left, best model accuracy 0.81, mean accuracy 0.77) and mRNA (right, best model accuracy 0.795, mean accuracy 0.78) by incorporating 100 RBP whole transcript binding probabilities (pysster⁶⁹ predictions). Mean decrease accuracy values of the top best 25 factors are shown. Factors specifically important for IncRNAs (found in the top 20 most important for IncRNAs and not among the top 30 for mRNAs) are marked with a red star; factors specifically important for mRNAs, respectively, are marked with a blue star.

(C) Boxplot distribution of transcription termination index fold-change upon 2 h IAA-induced Xrn2 depletion⁷² to control. Termination index was measured using POINT-seq read coverage as in Sousa-Luís et al.⁷³

(D) Metagene analysis depicting average POINT-seq read coverage ratios upon Xrn2 depletion-to-control, per nucleotide position ±2 kb around TES for the different chromatin dissociation classes of IncRNAs and mRNAs. Read ratios were extracted in 10 bp bins in the interval ±2 kb around TES and averaged across the analyzed loci.

We therefore propose a model where Xrn2-dependent degradation of the 3' flanking RNA at lncRNA loci could promote chromatin dissociation of the upstream nascent lncRNA transcript (Figure 5). In summary, fast-released lncRNAs employ mRNA processing mechanisms, such as Xrn2-dependent transcription termination, pointing to their functional regulatory potential.

DISCUSSION

lncRNAs constitute a large heterogeneous class with broad functional potential in regulation of gene expression, RNA processing, and chromatin states.^{2,80} Functional potential of lncRNAs is shaped by their subcellular localization, where they can form compartment-specific interactions with distinct RBPs and posit local specificity. Previous computational work generated predictive models of lncRNA subcellular localization (nuclear versus cytoplasmic enrichment) using steady-state RNA-seq and showed that inefficient splicing and intron retention is a major predictor of nuclear localization.¹⁷ However, what determines the observed lncRNA chromatin enrichment at steady state (referred to as chromatin retention or chromatin tethering), whether the same factors dictate localization dynamics of both lncRNAs and mRNAs, and whether chromatin (dis-) association dynamics are related to lncRNA functional potential are still open questions. Previous studies suggest that lncRNAs may remain tethered to chromatin via ongoing RNA Pol II transcription^{7,33} (because inhibiting RNA Pol II transcription elongation abolished lncRNA chromatin tethering³³), whereas the function of chromatin-bound lncRNAs acting in *cis* in regulation of proximal gene expression and local chromatin structure is coupled to their ongoing transcription.^{3,34,81} On the other hand, enhanced processing and chromatin dissociation of enhancer-transcribed lncRNAs suggests that they could act in *cis* within pre-established chromosomal proximity.^{32,35,36,82,83}

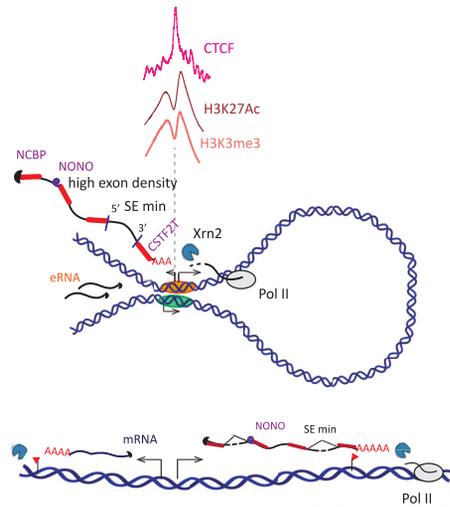
In this study we present the first machine learning framework to predict the chromatin (dis-) association dynamics of both lncRNAs and mRNAs from a large number of genomic features, as well as RBP data. Predictive models were trained on transcript dynamics measured by chrTT-seq, a new method in which we combine chromatin fractionation with sequencing of nascent RNA from the chromatin-associated and NP fraction at different pulse-chase time points. Compared with previous approaches, this allows us to characterize the dynamics of coding and non-coding nascent RNA transcripts at a higher resolution compared with steady state. By employing chrTT-seq we compute the chromatin dissociation rates of newly synthesized transcripts genome wide and uncover the most important features that dictate their retention to chromatin or fast release.

Previous work⁶⁰ implicated persistent U1 snRNP binding as a means of lncRNA chromatin tethering, which relies on U1 site enrichment in lncRNA exons, depletion of 3' splice sites and/or inefficient splicing, and U1 snRNP70 protein interactions with transcriptionally engaged RNA Pol II. A previous study indicated that the overall lower SE of lncRNAs (compared with mRNAs) is not due to defects in the U1-PAS axis, which is very similar to mRNAs.⁵⁹ In agreement, we also find U1 snRNP70 binding as a predictive feature of lncRNA and mRNA chromatin retention (Figure 2B). In Yin et al.,⁶⁰ U1 inhibition dampened the chromatin association of both well and

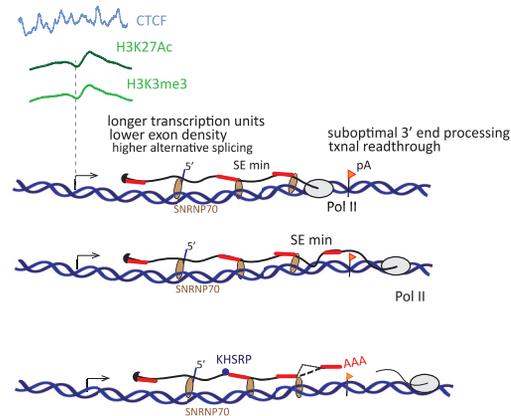
poorly spliced lncRNAs, suggesting that a kinetic effect due to delayed release of unspliced (or inefficiently/poorly spliced) nascent RNA cannot be the major determinant for lncRNA chromatin retention. However, the SE of the worst spliced intron per transcript is important for chromatin dissociation, suggesting that, for certain transcripts, inefficient processing of a poorly spliced intron may act as a bottleneck for chromatin release. Future experimental examination by modulating splicing at specific splice sites will help to validate the impact of co-transcriptional splicing on lncRNA chromatin release. Notably, compared with the transcript's average SE, the exon density, which reflects overall splicing activity, is a strong and high-confidence predictor of efficient chromatin dissociation of both nascent lncRNA and mRNA transcripts (Figures 2 and S1). Thus, we suggest that it is an increased propensity for splicing events and splicing signals within a certain transcript length that promotes chromatin dissociation, rather than the efficiency of splicing per se at the transcript's intron 3' splice sites. The latter is measured as the ratio of split to non-split reads spanning exon junctions, whereas splicing activity, assessed by extracting the transcript's exon density, may be determined by the overall interactions of the nascent RNA transcript with RBPs during co-transcriptional processing.

A finding of this study is that lncRNAs transcribed from active enhancers display increased degree of chromatin dissociation. This implies that the commonly termed *elncRNAs*²⁵ (equivalent to *la-EPCs*²¹) do not remain chromatin associated. Instead, chromatin dissociation is an important feature, which might underlie their functional potential and impact cognate enhancer activity. By using available GRO-seq data⁶⁶ in MCF-7, we show that there is a higher correlation in transcriptional responses to estradiol between fast-released lncRNAs and their ChIA-PET-identified target genes, compared with slow-released lncRNAs. Although our analysis is correlative in nature, and no causal mechanisms can be inferred at this stage, it underpins a possible role for lncRNA efficient chromatin dissociation in enhancing gene expression in *cis*. Ultimately, a future detailed experimental validation of single-lncRNA loci, aiming at altering the degree of lncRNA chromatin association, will facilitate the interpretation of these findings and shed light on the relationship among lncRNAs, chromatin dissociation dynamics, and its effect on cognate enhancer activity and target gene expression. So far, a decrease in lncRNA chromatin tethering was achieved transcriptome wide by inhibiting U1 snRNP⁶⁰ (without examining any effected alterations on putative *cis* targets), but it remains to be experimentally analyzed what is the effect of enforced *elncRNA* chromatin retention on cognate enhancer activity and target gene expression. For instance, impeding the co-transcriptional splicing of the *elncRNA A-ROD* with splice-inhibiting morpholinos had a repressive transcriptional effect on its target gene *DKK1*.³⁵ Modifying splice sites of other lncRNAs identified in this study by point mutations can validate the link between splicing activity and chromatin dissociation dynamics and assess to which extent efficient chromatin dissociation of *elncRNAs* results from increased splicing. It is also critical to disentangle the effect of *elncRNA* chromatin dissociation from splicing activity^{26,28,32} on cognate enhancer function by modulating the chromatin dissociation rate of an *elncRNA* without affecting their splicing.

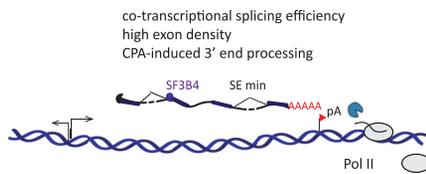
A lncRNAs, fast-released



B lncRNAs, slow-released (chromatin-retained)



C mRNAs, fast-released



D mRNAs, slow-released

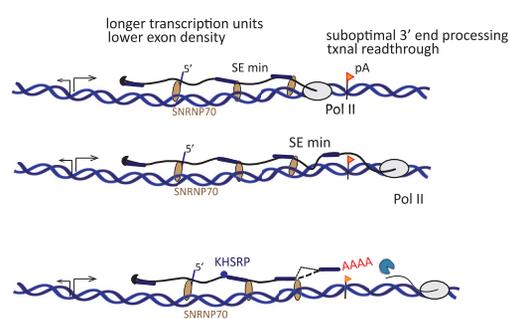


Figure 5. Summary of the mechanistic model

(A–D) Schematic depiction of the modeling, highlighting distinctive features between fast- versus slow-released transcripts, either lncRNAs (A and B) or mRNAs (C and D), not necessarily class (lncRNA or mRNA)-specific.

(A) Fast-released lncRNAs are enriched among lncRNAs transcribed from enhancer-like regions and the anchor points of chromosomal loops, with *A-ROD* being a representative example.³⁵ Thus, they show higher ChIA-PET scores, chromatin accessibility (DNase I), CTCF binding, H3K3me3, and H3K27Ac characteristics of enhancers. Fast-released lncRNAs showing high promoter-associated transcription bidirectionality may also be transcribed upstream antisense of protein-coding genes, with *GATA3-AS1* being a representative example (Figure S2). High splicing activity across the locus, measured by a significantly higher exon density, underlies efficient chromatin dissociation. For multi-exonic transcripts, splicing efficiency of a poorly processed intron (SE minimum) may act as a kinetic bottleneck during chromatin release. Analysis of RNA probabilities and incorporation of POINT-seq data^{72,73} suggests that Xrn2-dependent degradation of 3' RNA flanking fragments, generated by co-transcriptional cleavage at pA sites, promotes chromatin dissociation of the upstream nascent lncRNA transcript. Fast-released lncRNAs also show high binding probabilities for NCBP4, NONO, and CSTF2T.

(B) Slow-released lncRNAs are defined by lower promoter activity (marked by a lower H3K4me3 and H3K427Ac), increased chromatin-association half-lives due to extended transcription elongation across longer transcription units, and persistent SNRNP70 binding at U1 sites (presumably unresolved spliceosomes), most probably due to uneven distribution between 5' and 3' splice sites and/or suboptimal/weak 3' splice sites. The model cannot distinguish with certainty between slow-released lncRNAs with greater co-transcriptional chromatin-association half-lives and lncRNAs that remain tethered to chromatin post-transcriptionally, after their transcription has been concluded and at least some degree of splicing and A-tailing has occurred. In that case, interactions with RBPs, such as KHSRP, HNRNPL, and CSTF2 (which also bind DNA and localize on chromatin), as well as persistent SNRNP70 binding, may contribute to chromatin tethering. In the case of slow-released/chromatin-retained lncRNAs, cleavage and polyadenylation (CPA)- and Xrn2-independent modes of transcription termination and 3' end processing may be employed.

(C) Histone marks, chromatin states, and transcriptional activity are not strong distinctive features for fast- versus slow-released mRNAs. In general mRNAs as a class employ CPA-dependent XRN2-mediated modes of transcription termination and 3' end processing; thus, Xrn2 binding is not a strong predictive feature for fast-released mRNAs, which are defined by optimal co-transcriptional splicing efficiencies, high splicing activity across the locus measured by high exon density and show high binding probability of the splicing factor SF3B4.

(D) Same as in (C), also in the case of mRNAs, the model cannot distinguish with certainty between slow-released mRNAs with greater chromatin-association half-lives, due to extended transcription elongation times across long transcription units and, to some degree, polyadenylated (and intron-retaining) mRNAs that remain attached to chromatin post-transcriptionally. Slow-released mRNAs show suboptimal co-transcriptional splicing (lower SE and exon density) and 3' end processing, leading to transcriptional readthrough. Impeded/suboptimal cross talk between co-transcriptional splicing and 3' end processing was previously shown to lead to mRNA readthrough transcripts.⁵⁷ As in (B), high SNRNP70 binding is a predictive feature of slow-released mRNAs as well.

When it comes to the question of whether lncRNAs and mRNAs share the same molecular mechanisms of chromatin dissociation or whether distinct features dictate their dynamics, our models indicate that slow-released transcripts are retained by common mechanisms and share common features, such as longer transcription units, suboptimal co-transcriptional processing, transcriptional readthrough, and SNRNP70 binding. However, it seems that lncRNAs employ molecular features that will allow their efficient chromatin dissociation. In that sense, fast-released lncRNAs tend to resemble mRNAs as a class, by displaying an increased transcriptional activity, employing mRNA promoter characteristics, such as the H3K4me3 mark of active transcription, co-transcriptional splicing and exon density, and mechanisms of transcription termination and 3' end processing commonly utilized by mRNAs. Thus, although as a class lncRNAs are processed less efficiently compared with mRNAs and may depend on polyadenylation-signal-independent modes of 3' end formation and transcription termination,^{28,32,84} fast-released lncRNAs harness mRNA processing mechanisms, such as Xrn2-dependent transcription termination (Figure 5). This indicates potential regulatory roles and interactions upon chromatin dissociation, not necessarily in *trans* or after diffusion in the nucleus, but most probably on chromatin target sites defined by close three-dimensional proximity to their transcription sites.^{35,82,83} In addition, by leveraging predictive models of RBP binding, we pinpoint at several RBPs with a predicted important role in chromatin (dis-) association dynamics, some of which with a class-specific importance, suggesting that RBPs represents a critical regulatory layer when it comes to RNA cellular flows and to lncRNA activity. Experimental examination through RBP knockdown would validate these predictions and indicate specific candidate involvement in promoting chromatin release or tethering.

Regarding some technical points, we speculate that irregular transcriptional bursting at some loci may perplex the estimation of chromatin dissociation dynamics due to new transcription initiation events during the pulse-chase experiment. Therefore, to estimate chromatin-association halftimes, we modeled chromatin dissociation dynamics based on the last exon read coverage to minimize this effect. Of note, estimation of chromatin dissociation rates based on full-length transcript exonic read coverage renders a good correlation in the obtained half-time values ($\text{cor} = 0.84$, Figure S10A) and similar performance and feature importance of all elastic net models (Figures S10B and S10C). However, the feature coefficients of the lncRNA model for slow-released lncRNAs are dampened and result in unstable estimations (i.e., high error bars), indicating that this class of transcripts might be the most affected by the choice of “full length” over “last exon” modeling (Figure S10D). Characterizing how transcriptional bursting affects chromatin dissociation dynamics per se should be addressed experimentally in future work, for instance, by combining chrTT-seq with 4-SU-DRB-seq.⁸⁵ Furthermore, there might be additional molecular features acting in a cohort to define chromatin dissociation dynamics of newly transcribed RNA transcripts that have not been taken into account in our models. For instance, co-transcriptionally formed R-loops and epitranscriptomic marks (such as the N6-methyladenosine

RNA modification) deposited near the nascent transcript 3' end have been reported to underlie transcription termination efficiency^{86–88}; thus, their effect on chromatin dissociation dynamics remains to be determined.

Our chrTT-seq approach, employing short-read sequencing of fragmented, labeled nascent RNA, allows to measure RNA Pol II transcription dynamics in high resolution, as in the original TT-seq⁴⁸ protocol, while further enriching for nascent RNA molecules associated to the chromatin-residing transcriptional template. It therefore allows addressing the question of how transcriptional activity per se associates with distinct degrees of chromatin dissociation at given loci. In a recent study, Reimer et al. employed long-read sequencing of total chromatin-associated RNA to demonstrate that inefficient splicing associates with readthrough transcripts at single-molecule level.⁵⁷ A modified version of our approach, coupling nascent RNA-seq with chromatin fractionation at different pulse-chase time points, could leverage long-read sequencing of chromatin-associated and chromatin-released nascent, full-length transcripts to gain deeper mechanistic insights into the cross talk between co-transcriptional RNA processing/splicing dynamics of alternative splicing isoforms and their associated chromatin dissociation dynamics. In summary, the predicted models and findings of this study will help to enhance our understanding of lncRNA function in both normal and pathological contexts. Understanding subcellular and subnuclear lncRNA localization and the factors that determine its dynamic changes opens new avenues for targeted interference of lncRNA-mediated cellular processes and for designing effective RNA-based therapeutic strategies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - chrTT-seq
 - Mapping and spike-ins normalization
 - Transcript dataset
 - Modeling chromatin dissociation
 - Nucleoplasmic turnover
 - Definition of model features
 - Splicing efficiency, SED and degree of post-transcriptional splicing
 - Transcriptional indices (TSS-proximal pausing index and termination index)
 - SNRNP70 occupancy over transcription units
 - Machine learning models of chromatin (dis-)association based on genomic, transcriptomic and epigenetic features
 - Extraction of transcript 3' end site (TES)
 - Enhancer-associated lncRNAs in MCF-7
 - Prediction of RBP binding

- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL INFORMATION

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cels.2023.09.005>.

ACKNOWLEDGMENTS

We thank Rutger Gjaltema and Edda Schulz for helpful discussions and comments on the manuscript and the Max Planck Institute of Molecular Genetics for providing parts of the resources to conduct this study. E.N. acknowledges a Fondation Santé Research Grant. This work was supported by the DFG grant MA 4454/3-1 to A.M.

AUTHOR CONTRIBUTIONS

E.N. and A.M. conceived and planned the study with input from U.A.V.Ø. E.N. designed the computational and experimental pipelines and performed experiments and computational analyses. S.B. implemented the *pysster* method and contributed to data analyses. A.M. supervised and contributed to computational analyses. E.N. and A.M. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: October 30, 2022

Revised: May 24, 2023

Accepted: September 20, 2023

Published: October 18, 2023

REFERENCES

- Mattick, J.S., Amaral, P.P., Carninci, P., Carpenter, S., Chang, H.Y., Chen, L.L., Chen, R., Dean, C., Dinger, M.E., Fitzgerald, K.A., et al. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* *24*, 430–447. <https://doi.org/10.1038/s41580-022-00566-8>.
- Gil, N., and Ulitsky, I. (2020). Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* *21*, 102–117. <https://doi.org/10.1038/s41576-019-0184-5>.
- Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* *539*, 452–455. <https://doi.org/10.1038/nature20149>.
- Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* *22*, 96–118. <https://doi.org/10.1038/s41580-020-00315-9>.
- Andergassen, D., and Rinn, J.L. (2022). From genotype to phenotype: genetics of mammalian long non-coding RNAs in vivo. *Nat. Rev. Genet.* *23*, 229–243. <https://doi.org/10.1038/s41576-021-00427-8>.
- Ponting, C.P., and Haerty, W. (2022). Genome-wide analysis of human long noncoding RNAs: A provocative review. *Annu. Rev. Genomics Hum. Genet.* *23*, 153–172. <https://doi.org/10.1146/annurev-genom-112921-123710>.
- Carlevaro-Fita, J., and Johnson, R. (2019). Global positioning system: understanding long noncoding RNAs through subcellular localization. *Mol. Cell* *73*, 869–883. <https://doi.org/10.1016/j.molcel.2019.02.008>.
- Chen, L.L. (2016). Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* *41*, 761–772. <https://doi.org/10.1016/j.tibs.2016.07.003>.
- Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* *21*, 198–206. <https://doi.org/10.1038/nsmb.2764>.
- Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* *33*, 717–726. <https://doi.org/10.1016/j.molcel.2009.01.026>.
- Yao, R.W., Wang, Y., and Chen, L.L. (2019). Cellular functions of long non-coding RNAs. *Nat. Cell Biol.* *21*, 542–551. <https://doi.org/10.1038/s41556-019-0311-8>.
- Coté, A., Coté, C., Bayatpour, S., Drexler, H.L., Alexander, K.A., Chen, F., Wassie, A.T., Boyden, E.S., Berger, S., Churchman, L.S., and Raj, A. (2021). pre-mRNA spatial distributions suggest that splicing can occur post-transcriptionally. <https://doi.org/10.1101/2020.04.06.028092>.
- Johnsson, P., Ziegenhain, C., Hartmanis, L., Hendriks, G.J., Hagemann-Jensen, M., Reinius, B., and Sandberg, R. (2022). Transcriptional kinetics and molecular functions of long noncoding RNAs. *Nat. Genet.* *54*, 306–317. <https://doi.org/10.1038/s41588-022-01014-1>.
- Shukla, C.J., McCorkindale, A.L., Gerhardinger, C., Korthauer, K.D., Cabili, M.N., Shechner, D.M., Irizarry, R.A., Maass, P.G., and Rinn, J.L. (2018). High-throughput identification of RNA nuclear enrichment sequences. *EMBO J.* *37*, e98452. <https://doi.org/10.15252/emboj.201798452>.
- Lubelsky, Y., and Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* *555*, 107–111. <https://doi.org/10.1038/nature25757>.
- Lubelsky, Y., Zuckerman, B., and Ulitsky, I. (2021). High-resolution mapping of function and protein binding in an RNA nuclear enrichment sequence. *EMBO J.* *40*, e106357. <https://doi.org/10.15252/emboj.2020106357>.
- Zuckerman, B., and Ulitsky, I. (2019). Predictive models of subcellular localization of long RNAs. *RNA* *25*, 557–572. <https://doi.org/10.1261/rna.068288.118>.
- Goering, R., Arora, A., Pockalny, M.C., and Taliaferro, J.M. (2023). RNA localization mechanisms transcend cell morphology. *eLife* *12*, e80040. <https://doi.org/10.7554/eLife.80040>.
- Lai, F., Gardini, A., Zhang, A., and Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* *525*, 399–403. <https://doi.org/10.1038/nature14906>.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmid, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* *507*, 455–461. <https://doi.org/10.1038/nature12787>.
- Gil, N., and Ulitsky, I. (2018). Production of spliced long noncoding RNAs specifies regions with increased enhancer activity. *Cell Syst.* *7*, 537–547.e3. <https://doi.org/10.1016/j.cels.2018.10.009>.
- Hirabayashi, S., Bhagat, S., Matsuki, Y., Takegami, Y., Uehata, T., Kanemaru, A., Itoh, M., Shirakawa, K., Takaori-Kondo, A., Takeuchi, O., et al. (2019). NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. *Nat. Genet.* *51*, 1369–1379. <https://doi.org/10.1038/s41588-019-0485-9>.
- Young, R.S., Kumar, Y., Bickmore, W.A., and Taylor, M.S. (2017). Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome Biol.* *18*, 242. <https://doi.org/10.1186/s13059-017-1379-8>.
- Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Rönnerblad, M., Hrydzusko, O., Vitezic, M., et al. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* *347*, 1010–1014. <https://doi.org/10.1126/science.1259418>.
- Tan, J.Y., Biasini, A., Young, R.S., and Marques, A.C. (2020). Splicing of enhancer-associated lincRNAs contributes to enhancer activity. *Life Sci. Alliance* *3*, 3. <https://doi.org/10.26508/lsa.202000663>.

26. Tan, J.Y., and Marques, A.C. (2022). The activity of human enhancers is modulated by the splicing of their associated lncRNAs. *PLoS Comput. Biol.* *18*, e1009722. <https://doi.org/10.1371/journal.pcbi.1009722>.
27. Vučićević, D., Corradin, O., Ntini, E., Scacheri, P.C., and Ørom, U.A. (2015). Long ncRNA expression associates with tissue-specific enhancers. *Cell Cycle* *14*, 253–260. <https://doi.org/10.4161/15384101.2014.977641>.
28. Austenaa, L.M.I., Piccolo, V., Russo, M., Prosperini, E., Polletti, S., Polizzese, D., Ghisletti, S., Barozzi, I., Diaferia, G.R., and Natoli, G. (2021). A first exon termination checkpoint preferentially suppresses extragenic transcription. *Nat. Struct. Mol. Biol.* *28*, 337–346. <https://doi.org/10.1038/s41594-021-00572-y>.
29. Estell, C., Davidson, L., Steketee, P.C., Monier, A., and West, S. (2021). ZC3H4 restricts non-coding transcription in human cells. *eLife* *10*, e67305. <https://doi.org/10.7554/eLife.67305>.
30. Cortazar, M.A., Erickson, B., Fong, N., Pradhan, S.J., Ntini, E., and Bentley, D.L. (2022). Xrn2 substrate mapping identifies torpedo loading sites and extensive premature termination of RNA Pol II transcription. *Genes Dev.* *36*, 1062–1078. <https://doi.org/10.1101/gad.350004.122>.
31. Vlaming, H., Mimoso, C.A., Field, A.R., Martin, B.J.E., and Adelman, K. (2022). Screening thousands of transcribed coding and non-coding regions reveals sequence determinants of RNA polymerase II elongation potential. *Nat. Struct. Mol. Biol.* *29*, 613–620. <https://doi.org/10.1038/s41594-022-00785-9>.
32. Gil, N., and Ulitsky, I. (2021). Inefficient splicing curbs noncoding RNA transcription. *Nat. Struct. Mol. Biol.* *28*, 327–328. <https://doi.org/10.1038/s41594-021-00582-w>.
33. Werner, M.S., and Ruthenburg, A.J. (2015). Nuclear fractionation reveals thousands of chromatin-tethered noncoding RNAs adjacent to active genes. *Cell Rep.* *12*, 1089–1098. <https://doi.org/10.1016/j.celrep.2015.07.033>.
34. Werner, M.S., Sullivan, M.A., Shah, R.N., Nadadur, R.D., Grzybowski, A.T., Galat, V., Moskowitz, I.P., and Ruthenburg, A.J. (2017). Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. *Nat. Struct. Mol. Biol.* *24*, 596–603. <https://doi.org/10.1038/nsmb.3424>.
35. Ntini, E., Louloui, A., Liz, J., Muino, J.M., Marsico, A., and Ørom, U.A.V. (2018). Long ncRNA A-ROD activates its target gene DKK1 at its release from chromatin. *Nat. Commun.* *9*, 1636. <https://doi.org/10.1038/s41467-018-04100-3>.
36. Ntini, E., and Marsico, A. (2019). Functional impacts of non-coding RNA processing on enhancer activity and target gene expression. *J. Mol. Cell Biol.* *11*, 868–879. <https://doi.org/10.1093/jmcb/mjz047>.
37. Pandya-Jones, A., Bhatt, D.M., Lin, C.H., Tong, A.J., Smale, S.T., and Black, D.L. (2013). Splicing kinetics and transcript release from the chromatin compartment limit the rate of lipid A-induced gene expression. *RNA* *19*, 811–827. <https://doi.org/10.1261/ma.039081.113>.
38. Mauger, O., Lemoine, F., and Scheiffele, P. (2016). Targeted intron retention and excision for rapid gene regulation in response to neuronal activity. *Neuron* *92*, 1266–1278. <https://doi.org/10.1016/j.neuron.2016.11.032>.
39. Yeom, K.H., Pan, Z., Lin, C.H., Lim, H.Y., Xiao, W., Xing, Y., and Black, D.L. (2021). Tracking pre-mRNA maturation across subcellular compartments identifies developmental gene regulation through intron retention and nuclear anchoring. *Genome Res.* *31*, 1106–1119. <https://doi.org/10.1101/gr.273904.120>.
40. Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M., and Larson, D.R. (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell* *184*, 2878–2895.e20. <https://doi.org/10.1016/j.cell.2021.04.012>.
41. Pendleton, K.E., Park, S.K., Hunter, O.V., Bresson, S.M., and Conrad, N.K. (2018). Balance between MAT2A intron retention and splicing is determined cotranscriptionally. *RNA* *24*, 778–786. <https://doi.org/10.1261/ma.064899.117>.
42. Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., and Jensen, T.H. (2016). Identification of a nuclear exosome decay pathway for processed transcripts. *Mol. Cell* *64*, 520–533. <https://doi.org/10.1016/j.molcel.2016.09.025>.
43. Li, J., Zhang, L., He, S., Guo, F., and Zou, Q. (2021). SubLocEP: a novel ensemble predictor of subcellular localization of eukaryotic mRNA based on machine learning. *Brief. Bioinform.* *22*, bbaa401. <https://doi.org/10.1093/bib/bbaa401>.
44. Yan, Z., Lécuyer, E., and Blanchette, M. (2019). Prediction of mRNA subcellular localization using deep recurrent neural networks. *Bioinformatics* *35*, i333–i342. <https://doi.org/10.1093/bioinformatics/btz337>.
45. Musleh, S., Islam, M.T., Qureshi, R., Alajez, N.M., and Alam, T. (2023). MSLP: mRNA subcellular localization predictor based on machine learning techniques. *BMC Bioinformatics* *24*, 109. <https://doi.org/10.1186/s12859-023-05232-0>.
46. Wang, S., Shen, Z., Liu, T., Long, W., Jiang, L., and Peng, S. (2023). DeepmRNALoc: A novel predictor of eukaryotic mRNA subcellular localization based on deep learning. *Molecules* *28*, 2284. <https://doi.org/10.3390/molecules28052284>.
47. Gudenias, B.L., and Wang, L. (2018). Prediction of lncRNA subcellular localization with deep learning from sequence features. *Sci. Rep.* *8*, 16385. <https://doi.org/10.1038/s41598-018-34708-w>.
48. Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science* *352*, 1225–1228. <https://doi.org/10.1126/science.aad9841>.
49. Michel, M., Demel, C., Zacher, B., Schwalb, B., Krebs, S., Blum, H., Gagneur, J., and Cramer, P. (2017). TT-seq captures enhancer landscapes immediately after T-cell stimulation. *Mol. Syst. Biol.* *13*, 920. <https://doi.org/10.15252/msb.20167507>.
50. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* *22*, 1775–1789. <https://doi.org/10.1101/gr.132159.111>.
51. Olivero, C.E., Martínez-Teroba, E., Zimmer, J., Liao, C., Tesfaye, E., Hooshdaran, N., Schofield, J.A., Bendor, J., Fang, D., Simon, M.D., et al. (2020). p53 activates the long noncoding RNA Pvt1b to inhibit Myc and suppress tumorigenesis. *Mol. Cell* *77*, 761–774.e8. <https://doi.org/10.1016/j.molcel.2019.12.014>.
52. Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* *77*, 985–998.e8. <https://doi.org/10.1016/j.molcel.2019.11.017>.
53. Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R., et al. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* *20*, 923–928. <https://doi.org/10.1038/nsmb.2640>.
54. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* *499*, 360–363. <https://doi.org/10.1038/nature12349>.
55. Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol. Cell* *65*, 25–38. <https://doi.org/10.1016/j.molcel.2016.11.029>.
56. Louloui, A., Ntini, E., Conrad, T., and Ørom, U.A.V. (2018). Transient N6-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency. *Cell Rep.* *23*, 3429–3437. <https://doi.org/10.1016/j.celrep.2018.05.077>.
57. Reimer, K.A., Mimoso, C.A., Adelman, K., and Neugebauer, K.M. (2021). Co-translational splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell* *81*, 998–1012.e7. <https://doi.org/10.1016/j.molcel.2020.12.018>.

58. Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* 67, 301–320.
59. Melé, M., Mattioli, K., Mallard, W., Shechner, D.M., Gerhardinger, C., and Rinn, J.L. (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* 27, 27–37. <https://doi.org/10.1101/gr.214205.116>.
60. Yin, Y., Lu, J.Y., Zhang, X., Shao, W., Xu, Y., Li, P., Hong, Y., Cui, L., Shan, G., Tian, B., et al. (2020). U1 snRNP regulates chromatin retention of non-coding RNAs. *Nature* 580, 147–150. <https://doi.org/10.1038/s41586-020-2105-3>.
61. Rigo, F., and Martinson, H.G. (2009). Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. *RNA* 15, 823–836. <https://doi.org/10.1261/rna.1409209>.
62. Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA polymerase II transcription elongation with Pre-mRNA splicing. *J. Mol. Biol.* 428, 2623–2635. <https://doi.org/10.1016/j.jmb.2016.04.017>.
63. Chen, Y., Pai, A.A., Herudek, J., Lubas, M., Meola, N., Järvelin, A.I., Andersson, R., Pelechano, V., Steinmetz, L.M., Jensen, T.H., et al. (2016). Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. *Nat. Genet.* 48, 984–994. <https://doi.org/10.1038/ng.3616>.
64. Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. *Cell* 171, 1573–1588.e28. <https://doi.org/10.1016/j.cell.2017.11.008>.
65. Franco, H.L., Nagari, A., Malladi, V.S., Li, W., Xi, Y., Richardson, D., Allton, K.L., Tanaka, K., Li, J., Murakami, S., et al. (2018). Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. *Genome Res.* 28, 159–170. <https://doi.org/10.1101/gr.226019.117>.
66. Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.* 23, 1210–1223. <https://doi.org/10.1101/gr.152306.112>.
67. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature* 583, 711–719. <https://doi.org/10.1038/s41586-020-2077-3>.
68. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* 13, 508–514. <https://doi.org/10.1038/nmeth.3810>.
69. Budach, S., and Marsico, A. (2018). pysstter: classification of biological sequences by learning sequence and structure motifs with convolutional neural networks. *Bioinformatics* 34, 3035–3037. <https://doi.org/10.1093/bioinformatics/bty222>.
70. Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L., Pratt, G.A., et al. (2018). Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell* 70, 854–867.e9. <https://doi.org/10.1016/j.molcel.2018.05.001>.
71. Hortalcher, M., Oleshko, S., Hu, Y., Ghanbari, M., Cantini, G., Schinke, P., Vergara, E.E., Bittner, F., Mueller, N.S., Ohler, U., et al. (2023). A computational map of the human-SARS-CoV-2 protein-RNA interactome predicted at single-nucleotide resolution. *NAR Genom. Bioinform.* 5, lqad010. <https://doi.org/10.1093/nargab/lqad010>.
72. Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.* 32, 127–139. <https://doi.org/10.1101/gad.308528.117>.
73. Sousa-Luis, R., Dujardin, G., Zukher, I., Kimura, H., Weldon, C., Carmo-Fonseca, M., Proudfoot, N.J., and Nojima, T. (2021). POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. *Mol. Cell* 81, 1935–1950.e6. <https://doi.org/10.1016/j.molcel.2021.02.034>.
74. Misra, A., and Green, M.R. (2016). From polyadenylation to splicing: dual role for mRNA 3' end formation factors. *RNA Biol.* 13, 259–264. <https://doi.org/10.1080/15476286.2015.1112490>.
75. Planells, J., Jordán-Pla, A., Jain, S., Guadalupe, J.J., Proux-Wera, E., Euler, A.v., Pelechano, V., and Visa, N. (2023). The exosome degrades chromatin-associated RNAs genome-wide and maintains chromatin homeostasis. <https://doi.org/10.1101/2023.04.10.536209>.
76. Dhir, A., Dhir, S., Proudfoot, N.J., and Jopling, C.L. (2015). Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. *Nat. Struct. Mol. Biol.* 22, 319–327. <https://doi.org/10.1038/nsmb.2982>.
77. Eaton, J.D., and West, S. (2018). An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. *Transcription* 9, 321–326. <https://doi.org/10.1080/21541264.2018.1498708>.
78. Van Nostrand, E.L., Pratt, G.A., Yee, B.A., Wheeler, E.C., Blue, S.M., Mueller, J., Park, S.S., Garcia, K.E., Gelboin-Burkhart, C., Nguyen, T.B., et al. (2020). Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.* 21, 90. <https://doi.org/10.1186/s13059-020-01982-9>.
79. Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.* 34, 132–145. <https://doi.org/10.1101/gad.332833.119>.
80. Kopp, F., and Mendell, J.T. (2018). Functional classification and experimental dissection of long noncoding RNAs. *Cell* 172, 393–407. <https://doi.org/10.1016/j.cell.2018.01.011>.
81. Stojic, L., Niemczyk, M., Orjalo, A., Ito, Y., Ruijter, A.E., Uribe-Lewis, S., Joseph, N., Weston, S., Menon, S., Odom, D.T., et al. (2016). Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions. *Nat. Commun.* 7, 10406. <https://doi.org/10.1038/ncomms10406>.
82. Stojic, L., Lun, A.T.L., Mascacchi, P., Ernst, C., Redmond, A.M., Mangei, J., Barr, A.R., Bousgouni, V., Bakal, C., Marioni, J.C., et al. (2020). A high-content RNAi screen reveals multiple roles for long noncoding RNAs in cell division. *Nat. Commun.* 11, 1851. <https://doi.org/10.1038/s41467-020-14978-7>.
83. Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat. Rev. Mol. Cell Biol.* 17, 756–770. <https://doi.org/10.1038/nrm.2016.126>.
84. Nojima, T., and Proudfoot, N.J. (2022). Mechanisms of lincRNA biogenesis as revealed by nascent transcriptomics. *Nat. Rev. Mol. Cell Biol.* 23, 389–406. <https://doi.org/10.1038/s41580-021-00447-6>.
85. Beckman, W.F., Jiménez, M.Á.L., Moerland, P.D., Westerhoff, H.V., and Verschure, P.J. (2021). 4sUDRB-sequencing for genome-wide transcription bursting quantification in breast cancer cells. <https://doi.org/10.1101/2020.12.23.424175>.
86. Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* 42, 794–805. <https://doi.org/10.1016/j.molcel.2011.04.026>.
87. Hasanova, Z., Klapstova, V., Porrua, O., Stefl, R., and Sebesta, M. (2023). Human senataxin is a bona fide R-loop resolving enzyme and transcription termination factor. *Nucleic Acids Res.* 51, 2818–2837. <https://doi.org/10.1093/nar/gkad092>.
88. Yang, X., Liu, Q.L., Xu, W., Zhang, Y.C., Yang, Y., Ju, L.F., Chen, J., Chen, Y.S., Li, K., Ren, J., et al. (2019). m(6)A promotes R-loop formation to facilitate transcription termination. *Cell Res.* 29, 1035–1038. <https://doi.org/10.1038/s41422-019-0235-7>.
89. Menafrá, R., Brinkman, A.B., Matarese, F., Franci, G., Bartels, S.J., Nguyen, L., Shimbo, T., Wade, P.A., Hubner, N.C., and Stunnenberg, H.G. (2014). Genome-wide binding of MBD2 reveals strong preference

- for highly methylated loci. *PLoS One* 9, e99603. <https://doi.org/10.1371/journal.pone.0099603>.
90. Bonfert, T., and Friedel, C.C. (2017). Prediction of poly(A) sites by poly(A) read mapping. *PLoS One* 12, e0170914. <https://doi.org/10.1371/journal.pone.0170914>.
91. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
92. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
93. Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* 29, 436–442. <https://doi.org/10.1038/nbt.1861>.
94. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and analyzed data	This paper	GEO: GSE218726
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
MCF-7 P-Ser2 Pol II ChIP-seq	Menafra et al. ⁸⁹	GEO: GSM1388130
MCF-7 GRO-seq	Franco et al. ⁶⁵ ; Hah et al. ⁶⁶	GEO: GSM2545179, GSM2545180, GSM2545181, GSE43836
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, ENCFF7971UA.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. ²²	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
POINT-seq	Sousa-Luis et al. ⁷³	GEO: GSM4826619, GSM4826622
Experimental models: Cell lines		
Human: MCF-7 cells	ATCC	HTB-22
Software and algorithms		
ContextMap	Bonfert and Friedel ⁹⁰	https://www.bio.ifi.lmu.de/software/contextmap
Bedtools	Quinlan and Hall ⁹¹	https://bedtools.readthedocs.io/en/latest/
Pysster	Budach and Marsico ⁶⁹	https://github.com/budach/pysster
STAR	Dobin et al. ⁹²	https://github.com/alexdobin/STAR/releases
UCSC tools (bigWigAverageOverBed)	UCSC Genome Browser	https://genome.ucsc.edu/util.html
Code for chrTT-seq data analysis, chromatin dissociation modeling and machine learning	This paper	github.com/evgnt/chrTT-seq/
Code for chrTT-seq data analysis, chromatin dissociation modeling and machine learning (scripts for analysis of chrTT-seq data, computation of transcript half-life, linear and logistic regression models of chromatin dissociation from genomic features, random forest models of chromatin dissociation from RBP binding propensities, as well as auxiliary scripts to reproduce the main plots of the paper)	This paper	https://zenodo.org/record/8274758

RESOURCE AVAILABILITY**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Annalisa Marsico (annalisa.marsico@helmholtz-muenchen.de).

Materials availability

This study did not generate new materials.

Data and code availability

- All RNA-seq raw and processed data from this study have been deposited at GEO and are publicly available as of the date of publication. Accession number is listed in the [key resources table](#).
- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the [key resources table](#).
- All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).
- All code used for analysis has been deposited at Zenodo <https://zenodo.org/record/8274758>
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

METHOD DETAILS**chrTT-seq**

MCF-7 cells were seeded in P10 (6 plates per time point) and grown to ~80% confluency in 5% FCS, then labeled for 8 min with 1mM 4-thio-Uridine (4-SU). Cells were either immediately harvested (lifted intact in ice-cold PBS) or washed twice in PBS and chase was applied for 5, 10, 15, 20 min in 10 mM uridine diluted in growth medium. Chromatin fractionation was performed as in ref³⁵. Briefly, cells were lysed in 400 ul lysis buffer 0.15% NP-40 and lysate was loaded on 800 ul sucrose buffer for brief centrifugation. Pelleted nuclei were washed in ice-cold PBS, resuspended in 200 ul glycerol buffer and lysed in 0.6 M urea to fractionate chromatin from the nucleoplasmic fraction. RNA from the chromatin and nucleoplasmic fraction was extracted with acidic phenol (pH 4.5) and acidic phenol/chloroform. 3 ug of RNA were fragmented with 0.15 M NaOH final concentration for 25 min on ice. Prior to the RNA fragmentation, 0.15 ng of the 4-SU-labeled and unlabeled spike-ins mix (as in the TT-seq protocol⁴⁸) were added to the 3 ug of RNA. The fragmentation reaction was stopped in 10 mM Tris pH 7.4, purified with RNeasy MinElute Spin columns and eluted in 45 ul TE buffer (Tris 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 incubated for 2 hours at room temperature. Further steps of RNA purification, binding to T1 Dynabeads, washing and elution were done according to the Rabani et al. protocol⁹³ (using 5 ug T1 Dynabeads for 2 ug 4-SU-biotinylated RNA), leading to library construction for Illumina sequencing.

Mapping and spike-ins normalization

Reads from each library were aligned to GRCh38 (gencode.v23.primary_assembly.annotation) and to ERCC92 sequences using STAR⁹² 2.5.4a with default parameters. Only reads mapped to a single genomic location were retained for further analysis (STAR assigned score 255). Strand-specific reads counts for each transcript were computed with bedtools Version 2.27.0 (*coverage* function).

Three labeled (ERCC00043, ERCC00092, ERCC00136) and three unlabeled spike-ins (ERCC00002, ERCC00145, ERCC00170) had been added to each RNA sample. We computed for each sample j the sequencing depth σ_j (also so-called 'size factor') and the cross-contamination rate ϵ_j based on the spike-in counts. We used a similar approach to the one described in Schwalb et al.,⁴⁸ but adapted it to our experimental setting. Briefly, we used a statistical model that describes the observed read counts in a TT-seq sample by the length of the feature (spike-in / transcript n), I_n , and the feature-specific unlabeled and labeled RNA amounts that we want to estimate, α_j and β_j .

$$k_{nj} = I_n \sigma_j (\alpha_{nj} + \epsilon_j \beta_{nj})$$

For labeled samples $\alpha_{nj} = 0$, therefore $\beta_{nj} = k_{nj} / I_n \sigma_j \epsilon_j$

For each sample the size factor σ_j was computed as follows: each of the three labeled ($i \in L$) spike-in read count in sample j (S_{ij}) was normalized to the sum of the respective spike in counts across the ten labeled samples (CHR 0, 5, 10, 15, 20 min and NP 0, 5, 10, 15, 20 min). Then, for each sample j , the median from the the three scaled labeled spike ins was extracted ('smoothed median'):

$$\sigma_j = \text{median}_L \left(s_{ij} / \sum_{i=1}^{10} S_{ij} \right)$$

For each labeled sample (L) the cross-contamination rate ϵ_j was computed as the sum of unlabeled (U) spike in read counts over the sum of unlabeled (U) and labeled (L) spike in read counts:

$$\epsilon_j = \frac{\sum_{k \in U} S_{kj}}{\sum_{k \in U} S_{kj} + \sum_{i \in L} S_{ij}}$$

where i is the index over the labeled spike ins and k is the label over the unlabeled spike ins. Labeled samples with minimal or no contamination will have a value of ϵ_j close to zero, while highly contaminated samples will have a value of ϵ_j close to one. Normalized strand-specific read counts over features $n \in N$ were then computed, in each sample j , as:

$$\beta_{nj} = k_{nj} / I_n \sigma_j \epsilon_j$$

Transcript dataset

We used GENCODE V29 lncRNA annotation ($n = 8,992$) supplemented with lncRNA transcripts non-overlapping GENCODE V29 lncRNAs annotation, from *de novo* transcript assembly on chromatin-associated RNA-seq in MCF-7 ($n = 10,606$; described in ref³⁵). Specifically, newly identified lncRNA transcripts are lacking protein-coding potential, are not overlapping protein-coding genes, and have at least one splice junction. From this initial set we kept 3,671 lncRNAs with non-zero read coverage in all 12 sequenced samples. We also used 15,166 mRNA transcripts with non-zero read coverage in all 12 samples.

Modeling chromatin dissociation

Strand-specific read counts over the last exon of the 18,837 transcripts (15,166 mRNAs and 3,671 lncRNAs) were normalized to spike-ins and feature length (as described in the STAR Methods section ‘mapping and spike-ins normalization’). For each pulse-chase time point we extracted the ratio of chromatin (CHR) versus total (chromatin plus nucleoplasmic (NP)) normalized read coverage (CHR/ (CHR+NP)). For each transcript, we expect this quantity to decay over time, as it dissociates from chromatin and translocates to the nucleoplasm with dynamics which are different from transcript to transcript. We fit those ratios on an exponential decay function of the form:

$$x(t) = x_0 e^{-kt}$$

using the R function $lm(\ln(x) \sim t)$, for time-points [0, 8, 13, 18, 23, 28].

We set x_0 to 1, as we assume that at time point 0 all transcripts are purely chromatin associated (at time point 0, 4-thio-uridine is starting to get incorporated into newly transcribed RNA).

This returns *intercept*, exponent k and *p-value* of exponential decay fit. We kept 12,391 entries that fit the curve with a p -value < 0.05 , of which 2077 were lncRNAs and 10,314 mRNAs. We defined a ‘chromatin association half-time’ as:

$$t_{1/2} = -(\text{intercept} + \ln(2)) / k$$

representing the time required for the decaying quantity (i.e. transcript chromatin-associated ratio) to fall to one half of its initial value (at $t = 0$).

Based on the half-time values of the whole transcript dataset, we split the whole transcript dataset ($n = 12,391$) into three equal-size quantiles corresponding to ‘fast’, ‘medium’ and ‘slow’ released nascent RNA transcripts.

Nucleoplasmic turnover

We computed the nucleoplasmic ‘turnover’ for each transcript (due to RNA degradation or export) as the ratio $(NP_{20} - NP_0) / NP_0$, where NP_0 and NP_{20} represent nascent RNA read coverage from a transcript’s last exon in the nucleoplasmic fraction (chromatin-released), measured at 0 minutes and 20 minutes from the start of chase, respectively. Please note that chase time-point zero, corresponds to the end of 8 minutes of 4SU labeling; at this point, 4SU is removed/washed away and chase starts with an excess of uridine for 0-20 min. Transcripts with a rapid nucleoplasmic turnover are expected to show small or negative values of $(NP_{20} - NP_0) / NP_0$, whereas transcripts subject to lower nucleoplasmic turnover (degradation and/or export), measured during 20 minutes chase, are expected to show relatively higher positive values of this ratio, as they accumulate more nascent nucleoplasmic reads from the flow of chromatin release.

Definition of model features

The 18 molecular features used to train predictive models of both lncRNA and mRNA (dis-)association dynamics are listed and described in detail in Table S1. Briefly, the list includes 10 epigenetic features extracted from publicly available ChIP-seq data in MCF-7 (histone marks, chromatin factors and Pol II-related features), four transcriptomic features related to either transcription *per se*, and extracted by publicly available GRO-seq data in MCF-7, or splicing (i.e. splicing efficiency, average and minimum values per transcript), extracted from our chrTT-seq raw data, and finally two genomic features including ChIA-PET interactions strength and exon density, as described in Table S1.

Splicing efficiency, SED and degree of post-transcriptional splicing

We measured intron splicing efficiency (SE or θ value) as in Schlackow et al.⁵⁵ by extracting the ratio of split to split plus non-split reads overlapping 3' splice sites of introns with at least one split and one non-split read at the 3' splice site ($n = 154,467$ high-confidence introns). For monoexonic transcripts, the minimum (SE min) and mean (SE mean) splicing efficiency values were imputed by applying k-nearest neighbor imputation with $k = 5$. We measured alternative splicing as in Louloupi et al.⁵⁶ by extracting the ratio (ρ value) of alternative split to constitutive split reads covering the high-confidence introns. We extracted co- and post-transcriptional splicing efficiency dynamics (SED) as in Louloupi et al.,⁵⁶ by subtracting the difference of splicing efficiency at 20 min pulse-chase time point from the splicing efficiency at 0 min and normalizing this to the splicing efficiency at 0 min [$SED = (SE_{20min} + 0.001 - SE_{0min}) / (SE_{0min} + 0.001)$]. We extracted the extent of post-transcriptional splicing relative to co-transcriptional as the difference of chromatin-associated splicing efficiency from the nucleoplasmic splicing efficiency, normalized to chromatin. This was done at intron and transcript level (mean value of the transcript's high-confidence introns).

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

We assessed transcriptional pausing index by extracting the ratio of strand-specific GRO-seq read coverage or P-Ser2 Pol II ChIP-seq density in the window 500 nt downstream of TSS to the gene body. Gene body was defined as the middle 50% of the interval TSS+500 to TES, as in Schlackow et al.⁵⁵ Transcription termination index was measured as in Schlackow et al.⁵⁵ by extracting the length-normalized ratio of strand-specific GRO-seq read coverage (or Pol II ChIP-seq read density) in the window 2.5 Kb downstream of TES to the gene body. Travel index was extracted as the ratio of read coverage in the interval [2.5 to 5 Kb] downstream of TES to the first 2.5 Kb downstream of TES. Transcription bidirectionality score was measured as in Chen et al.⁶³ by extracting the \log_2 antisense/sense ratio of GRO-seq read coverage 1 Kb around TSS.

SNRNP70 occupancy over transcription units

As a proxy we used SNRNP70 ChIP-seq from HepG2 and by intersecting the intervals corresponding to full-length transcripts with ChIP-seq narrow peaks (ENCF346UDN) we extracted a mean binding score per transcription unit, by averaging the signal over the peaks.

Machine learning models of chromatin (dis-)association based on genomic, transcriptomic and epigenetic features Regularized logistic regression

We developed a logistic regression model for predicting dissociation dynamics of transcripts in a biotype-agnostic manner, as well as individually for lncRNAs and mRNAs based on genomic and functional features. We used regularized logistic regression to determine the most important subset of variables for prediction, with soft constraints on non-zero coefficients. This helps assigning similar weights for correlated variables, which was particularly useful given the presence of few sets of correlated variables in our data (Figure S4A). We used Elastic Nets⁵⁸ as implemented in the *glmnet*³⁴ package for R. The objective function for regularized logistic regression is the penalized negative binomial log-likelihood:

$$-\left[\frac{1}{N} \sum_{i=1}^N y_i (\beta_0 + x_i^T \beta) - \log(1 + e^{(\beta_0 + x_i^T \beta)}) \right]$$

subjected to the constraint $(1 - \alpha) \|\beta\|_1 + \alpha \|\beta\|_2 \leq t$, where $\|\beta\|_1 = \sum_{i=1}^p |\beta_i|$ and $\|\beta\|_2 = \sqrt{\sum_{i=1}^p \beta_i^2}$ with $p =$ number of features,

$N =$ number of samples (lncRNAs or mRNAs) and $\alpha \in [0, 1]$. The first constraint is based on the L1-norm and forces unimportant coefficients to shrink to 0, thereby promoting sparsity (LASSO-type), while the second constraint is based on L2-norm and favors similar values for the coefficients (Ridge-type). The α parameter specifies the contribution of each constraint. Throughout the paper we chose $\alpha = 0.5$ in order to balance the contributions of the two penalizations. The t parameter was optimized for each model by the *cv.glmnet* function in a 10-fold cross-validation procedure on the training data, to minimize the misclassification error (MCE). The model training was repeated 10 times and each time 10% of the data was left out for testing purposes, while the other 90% was used for model optimization, as described above. The regression coefficients correspond to the average values over the 10 training steps, and the final model's MCE is the average over 10 test sets.

Regularized linear regression

For comparison, we build regularized linear regression models for all transcripts, as well as lncRNAs and mRNAs separately, to predict the chromatin-association half-time as a continuous value from the same set of features. A similar procedure, as described for regularized logistic regression, was employed here, where instead the sum of squared errors (SSE) was minimized and subjected to the same constraints. In a similar way, the value of t parameter was chosen with the *cv.glmnet* R function in a 10-fold cross-validation procedure to minimize the SSE. As a performance measure for linear regression we report the best Pearson correlation coefficient (R) value over the 10 test sets.

Random Forests

Random Forest (RF) classification models were also built to confirm predictions of fast versus slow-released transcripts with the *randomForest* R package. The number of trees (*n*tree parameter) was fixed to 1000, while the number of variables for each tree (*m*try parameter) was optimized using the *train()* function of the *caret* R package with a grid search in a ten-fold cross validation on the

training data. The best *mtry* was selected, such that it maximizes the model accuracy. Model training was repeated ten times and each time 10% of the data was left out for testing, while the other 90% was used for model optimization, as described above. The importance of each feature was computed as ‘mean decrease in Gini index’. Variables with large positive importance values correspond to features which are crucial for model classification, whereas variables with values close to zero or negative correspond to unimportant features.

Extraction of transcript 3' end site (TES)

We ran ContextMap³⁰ v2.7.9 on paired-end MCF-7 nuclear polyA+ data (ENCODE) using Bowtie2 aligner and Bowtie2-build-1 indexer, with parameters `-mismatches 3 -seed 30 -maxhits 10 -polyA -t 8 -Xms4000M -Xmx30000M`. This generated 39,991 ContextMap scored polyA sites. Nearby polyA sites were clustered with bedtools `cluster -s -d 10`, keeping the one with maximum score. Annotated transcript 3' ends were assigned a ContextMap polyA site by fetching the closest with bedtools `closest -s`.

Enhancer-associated lncRNAs in MCF-7

From the FANTOM5/NET-CAGE enhancers (n = 85,786) we extracted the ones that show evident bidirectional transcription in MCF-7 using GRO-seq (GSE96859), by calculating average coverage over all bases, with non-covered bases counting as zeroes; `bigWigCoverageOverBed mean0 coverage > 0.1` for both strands, resulting in 10,008 bidirectional actively transcribed enhancers. We then fetched the closest transcript start site (TSS) to enhancer midpoints using bedtools `closest -s` and defined lncRNAs with a distance < 2000 bp from a transcribed enhancer as elncRNAs (n = 247 out of the 2077 analyzed).

Prediction of RBP binding

RBP datasets and data pre-processing

To predict RBP binding patterns on MCF-7 lncRNA and mRNA transcripts, we used the machine learning tool pysster⁶⁹ as RBP eCLIP data for MCF-7 is not available in the ENCODE database. We trained pysster models for each RBP using eCLIP data from HepG2 and K562 and applied them to predict RBP binding on MCF-7 transcripts. We didn't use HepG2 or K562 RBP binding sites as proxies as the pool of expressed lncRNAs differs between cell types, resulting in substantial loss of information. When considering lncRNAs only, the percentage of commonly expressed transcripts between HepG2 and MCF-7 was only 19.5% (405 out of 2077 lncRNAs). As of September 2018, 161 eCLIP ENCODE datasets for RBPs were available. 100 RBPs with good quality scores were selected for pysster models, as summarized in Table S2. If an RBP was available for both HepG2 and K562, the higher quality cell line was used, or HepG2 by default. Two bed files containing narrow peaks were downloaded for each RBP, and peaks overlapping with the other replicate and with log-fold enrichment over the input control sample greater than or equal to two were selected. The 5' end of each peak was considered the binding site,⁷⁰ and binding locations were summarized using GENCODE gene annotations version 24 for the GRCh38 human genome assembly to identify transcripts overlapping with RBP binding sites.

Pysster training on ENCODE eCLIP data

We trained a Convolutional Neural Network (CNN) using pysster to classify binding sites specific to an RBP. The model was trained in a three-class classification setting to enable learning sequence preferences specific to the RBP of interest and avoid learning technical biases. Class 1 contained sequences of length 400 centered at the 5' end of the RBP's binding sites extracted from eCLIP peaks. Class 2 consisted of sequences of length 400 randomly sampled from lncRNAs or mRNA transcripts binding to the RBP of interest, but not overlapping with the peak sequences in class 1. Class 3 consisted of sequences of length 400 centered at the 5' end of eCLIP binding sites of any of the 99 RBPs in our dataset, distinct from the RBP of interest. The model also incorporated the genomic location of a binding site by adding three neurons to the first dense layer of the CNN. Categorical features, such as exons/introns, were zero/one encoded using as many neurons as the number of categories. Proximity to the TSS/TTS was encoded as a continuous feature and computed as the distance of the 5' end of the eCLIP peak to the TSS, normalized by the transcript's length. RBP pysster models were trained on imbalanced data in a ratio 1:5:5 between the positive class and the two background sets. Input sequences were randomly split into 70% training, 15% validation and 15% held-out set. For each RBP model an hyper-parameter search was performed to optimize the number and lengths of kernels.

Pysster RBP binding site predictions on MCF-7 transcripts

To obtain single-nucleotide binding site probabilities of an RBP along a full transcript we employed a sliding window approach (window size 400, step-size of one) to scan the model predictions over a full-length lncRNA or mRNA. The predicted probability of a sequence to belong to the positive class (i.e. to be bound by the RBP of interest) was assigned to the central nucleotide of each window. All positions with prediction scores higher than 0.66 (twice the value of the random probability to belong to the positive class) were selected as ‘high confidence predictions’ and their median binding probability was computed.

Modeling chromatin (dis-)association from RBP binding with Random Forests

A random Forest (RF) classification model to predict fast versus slow-released nascent RNA transcripts from RBP binding alone was implemented separately for both lncRNAs and mRNAs with the randomForest R package. RF models were trained on 100 features, each representing the median binding strength (median probability) of a single RBP according to pysster. RF Model training, testing and feature importance analysis were done as described earlier.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses used are indicated in the [method details](#) section; statistical correlations were computed by Spearman correlation tests unless otherwise indicated, comparisons between groups were computed by two-sided Student's test unless otherwise indicated, and comparisons of proportions of between two groups were computed by Fisher's exact tests. All p values are indicated in the figures; statistical tests used are indicated in the figure legends. Dots in the boxplots indicate mean (average) values of distributions, and lines represent the median.

ADDITIONAL INFORMATION

[Supplemental information](#) is available for this paper. Correspondence and requests for materials should be addressed to [evgenia.ntini@imbb.forth.gr](#) or [annalisa.marsico@helmholtz-muenchen.de](#) ([lead contact](#)).