

A state model of transcription and co-transcriptional processing enables measurement of the underlying kinetics

Matthew Gamble

matthew.gamble@einsteinmed.edu

Albert Einstein College of Medicine

Joshua Saltzberg

Albert Einstein College of Medicine <https://orcid.org/0009-0004-0451-505X>

Alyssa Fronk

Envisagenics, Inc.,

Charles Query

Albert Einstein College of Medicine <https://orcid.org/0000-0002-7692-2496>

Kenny Ye

Albert Einstein College of Medicine <https://orcid.org/0000-0002-1466-855X>

Article

Keywords:

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10830564/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

A state model of transcription and co-transcriptional processing enables measurement of the underlying kinetics

Joshua D. Saltzberg¹, Alyssa D. Fronk², Charles C. Query³, Kenny Ye^{4,5}, Matthew J. Gamble^{1,3,*}

¹ Department of Molecular Pharmacology, Albert Einstein College of Medicine, Bronx, NY 10461

² Envisagenics Inc., Long Island City, NY 11101

³ Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461

⁴ Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461

⁵ Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, NY 10461

Correspondence to: matthew.gamble@einsteinmed.edu

Abstract

The production of mature mRNA isoforms is governed by the kinetics of transcription and co-transcriptional RNA processing, yet these processes remain difficult to quantify simultaneously. Here, we introduce SKaTER, a mathematical state model of transcription that jointly infers the kinetics of elongation, splicing, and cleavage from either DRB released nascent RNA sequencing or a combination of PRO-seq and TT-seq. Rate measurements are reproducible and highly concordant across cell lines. Cleavage rates are strongly associated with the presence of the canonical polyadenylation signals and can be impacted by RNA binding proteins such as SRSF9. Splicing rates are dominated by spatial factors, with neighboring introns and introns in the same TAD showing strong correlations. Elongation rates are heavily determined by local chromatin marks, with rates varying both with intensity and combinations of different marks. Our framework enables joint kinetic analysis of mRNA biogenesis revealing how sequence, nuclear organization and chromatin state shape co-transcriptional RNA processing.

Introduction

Transcription of mRNA is a complex process involving multiple steps including initiation at the transcription start site (TSS), promoter proximal pause clearance, splicing of introns, and cleavage at a polyadenylation site (PAS). This process can result in the formation of numerous alternative isoforms that increase the diversity of the proteome. Alternative isoform production is regulated by the underlying kinetics of transcription, splicing, and cleavage as described by the window of opportunity model (Bentley, 2014; Cáceres & Kornblihtt, 2002). In this model, transcriptional elongation rates determine when alternative splice and cleavage sites become eligible for processing. Slower polymerase elongation rates provide a longer window of opportunity for weaker upstream sites to be processed before stronger downstream sites can be transcribed and outcompete them (Bentley, 2014; Cáceres & Kornblihtt, 2002). Therefore, accurate measurement of transcriptional kinetics is necessary to understand alternative isoform production.

Numerous methods have been developed to measure the kinetics of polymerase elongation rates including single cell imaging (Darzacq et al., 2007), DRB-seq protocols (Fuchs et al., 2014; Singh & Padgett, 2009; Veloso et al., 2014), Pol II mapping (Cortazar et al., 2019; Danko et al., 2013; Jonkers et al., 2014), and steady state metrics incorporating TT-seq (Schwalb et al., 2016) like elongation index (Caizzi et al., 2021; Mimoso & Adelman, 2023; Stein et al., 2022; Žumer et al., 2021; Žumer et al., 2024). However, steady state experiments using TT-seq are distorted by the loss of intronic coverage due to splicing (Mimoso & Adelman, 2023; Schwalb et al., 2016) while other approaches are limited to measuring the pioneer (Danko et al., 2013) or terminal polymerase (Jonkers et al., 2014). Genome wide splicing rates measurements require

the assumption of constant elongation rates to estimate intronic half lives (Carrillo Oesterreich et al., 2016; Pai et al., 2017; Reimer et al., 2021; Wachutka et al., 2019). A unified approach capable of measuring all the kinetics of transcription simultaneously is needed to accurately measure each individual rate.

In this work, we present a novel state model of transcription incorporating splicing, cleavage, and Pol II rates into a single mathematical framework called Splicing Kinetics and Transcriptional Elongation Rates (SKaTER). We employ this model to measure transcriptional kinetics from nascent transcription following a DRB block (SKaTER-seq) or the combination of TT-seq and PRO-seq (TT-SKaTER). We show that both SKaTER-seq and TT-SKaTER are able to accurately and reproducibly measure transcriptional kinetics across multiple different cell lines. We validate these kinetics by comparing our observed rates with known properties of transcription and identify novel regulatory factors underlying cleavage, splicing, and elongation rates. We find that cleavage rates are strongly associated with RNA motifs, splicing rates are heavily coordinated by spatial proximity, and elongation rates can be predicted based on the combination of a handful of chromatin marks. Our work establishes the role of multiple factors in regulating the kinetics of transcription, which could potentially serve to modulate alternative isoform formation.

Results

State Model of Transcription

To measure transcriptional kinetics, we developed a state model of transcription. In this model, each state describes a unique combination of Pol II position, splicing status of each intron, and cleavage status of each cleavage and polyadenylation site (CPS). In this framework, a simple

gene with only one intron and a single CPS can be described using seven states (Figure 1A).

Transcripts initiate at the transcription start site at rate:

$$(1) \quad dr(t) = \begin{cases} 0 & t \leq 0 \\ G\lambda_n - G(\lambda_n - \Lambda_0)e^{[-\Lambda_0 t]} & t > 0 \\ G\lambda_n & t \rightarrow \inf \end{cases}$$

where G is the depth normalized number of TSS and λ_n is the rate of initiation at active TSS. t is the amount of time since gene activation (DRB washout in SKaTER-seq or infinity to represent steady state in TT-SKaTER). Finally, Λ_0 is the rate of initial Pol II activation at the TSS. After initiation, transcripts move between states by splicing at an intron, cleaving at a CPS, or transcribing particular ‘critical points’ such as a CPS or 3’ splice site. Polymerase arrival at any particular position is modeled as a time delay, and splicing and cleavage are modeled to follow an exponential distribution. This produces a set of differential equations describing the flux of transcripts between different states as a function of time (Figure 1B).

We developed a set of general solutions to differential equations of this format to measure the total number of mRNAs in any particular state (Supplemental Text). With these general solutions, we can map these states back on to the gene to predict the coverage at any position in the gene (Figure 1C). The number of states required to model a gene grows geometrically with the number of introns and CPS within the gene. Therefore, we simplify the model by assuming that the rate of splicing and cleavage at each site is independent of the transcript state. We can divide any given gene into multiple smaller segments, termed events, by substituting all introns that contain no information about the interrogated position with constitutive exons. We can then assemble predictions for coverage for the gene by combining coverage from intervals of each event (Figure 1D).

Using our equations, we can model two different types of kinetic information. First is coverage of final positions of individual reads within a gene. Within introns, we observe a loss of coverage relative to the flanking exons. For example, we can model coverage for TT-seq (Schwalb et al., 2016) at different sets of polymerase elongation rates to show intronic coverage distributions are partially a function of elongation rates (Figure 1E) and splicing rates (Supplemental Figure 1A-C). This shows that varying transcriptional kinetics should result in changes in coverage and splice junction ratios. Real sequencing data shows properties consistent with our modeled data. For example, we see a dramatic depletion of intronic coverage in TT-seq in IMR90 cells (Figure 1F, Supplemental Figure 1G). Similarly, we observe a loss of coverage downstream of the CPS, as predicted by the state model (Figure 1G, Supplemental Figure 1H).

The second source of information is the splicing status of an intron as a function of final position of a fragment. For any given fragment, we know that the polymerase transcribing the fragment must have been past the final position of the read pair at time of sample collection. If the fragment spans either an intron–exon or exon–exon junction it informs on the splicing status of that RNA molecule. We can calculate the ratio of these junctions at any given position based on the set of states that contain either junction (Supplemental Figure 1D) and predict how these ratios should change with different kinetics (Supplemental Figure 1E-F). As predicted, TT-seq shows that the spliced fraction of junctional reads increases as the final nucleotide gets further from the 3' splice site as those mRNAs have had more time to be spliced on average than those that are close to the 3' splice site (Figure 1H). This effect is exacerbated by the context of the final nucleotide with positions inside exons showing higher spliced fractions than those in introns. This is expected as introns may have been spliced out before sample collection, while exons can be sampled from

any time during the experiment. However, later introns both splice faster (Bedi et al., 2021; Pai et al., 2017) and have more downstream exonic coverage, confounding these plots.

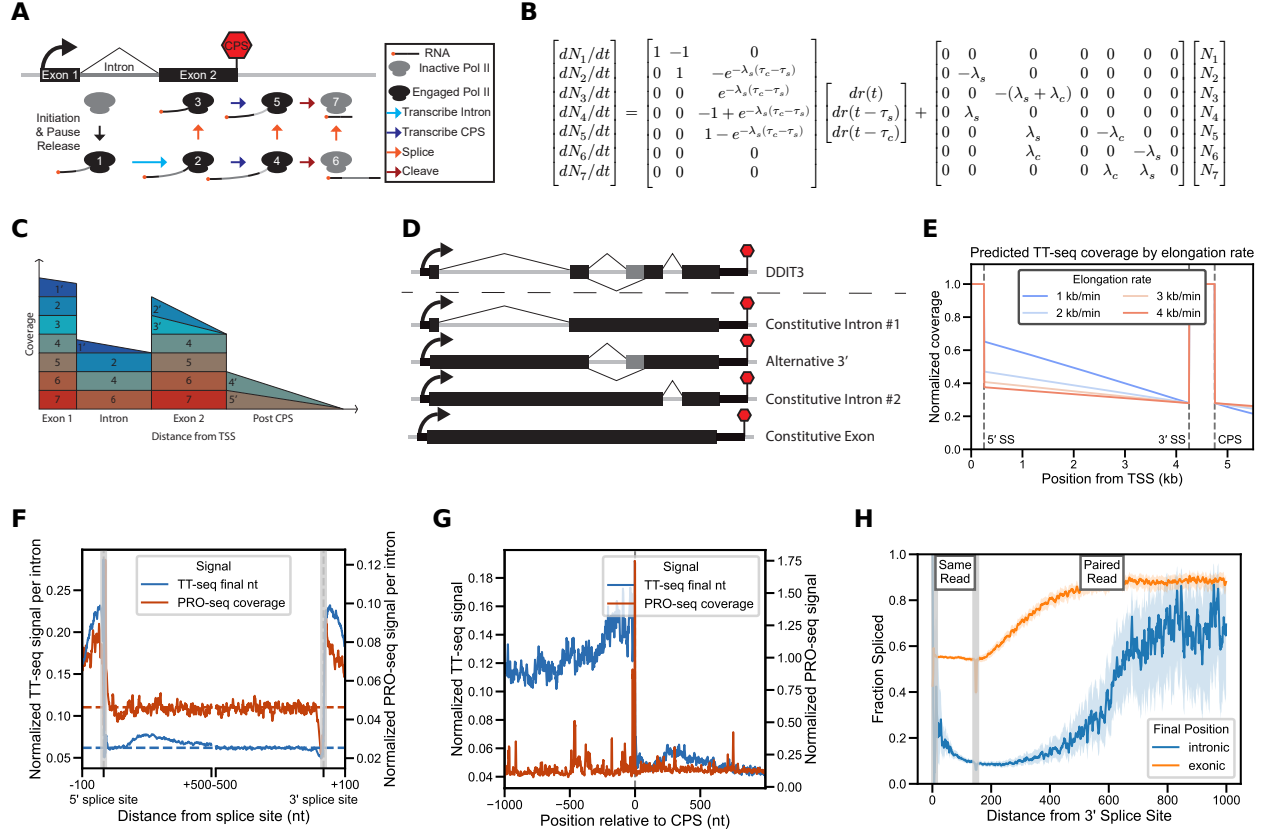


Figure 1. Development of a state model of nascent RNA transcription: (A) Schematic of state model of transcription. Inactive polymerases initiate at the TSS into the first state, and move through the different states by transcription, splicing, or cleavage. (B) State model defines the set of seven differential equations. N_x corresponds to the number of mRNAs in state x . λ_s and λ_c are the rate of splicing and cleavage, and τ_s and τ_c are the amount of time required to reach the 3' splice site or CPS respectively. (C) mRNAs are able to contribute to coverage in a state dependent manner. States that are only able to partially contribute to coverage at particular locations are accented. (D) Example using DDIT3 demonstrating how genes can be divided into multiple events by replacing non-conflicting splicing with constitutive introns. (E) Predicted TT-seq coverage at different elongation rates. Slower elongation causes increase in coverage at the 5' end of the intron. (F) Pileup of nascent coverage around the 5' and 3' splice sites of constitutive introns in IMR90 cells. Nascent coverage is notably lower than surrounding exonic coverage. There is a further loss of coverage within the intron from the 5' to the 3' end. This effect is due to splicing as PRO-seq, which is unaffected by splicing, does not show the same trend. Regions around the 5' and 3' splice sites have difficulty mapping and are masked. (G) Coverage pileup past CPS shows a dramatic drop in coverage in the TT-seq data, consistent with cleavage of nascent transcripts. This effect is not observed in PRO-seq, demonstrating that cleaved transcripts can contribute to exonic coverage in TT-seq. (H) Plot of fraction indicated as spliced by junctional reads in TT-seq based on the downstream position of the final nucleotide of the fragment. Positions within constitutive exons are in orange while constitutive introns are in blue. Transitions between single reads and paired ends cause mapping difficulties and are masked. Further, there are very few intronic reads close to the 3' splice site, so this region should be disregarded as well. Exonic final positions can be retained in the final transcript and have higher baseline fraction spliced than

intronic final positions, which only can be present in nascent transcripts. There is also a general positive trend where positions that are farther from the 3' splice site have higher fraction spliced than positions that are closer.

Measuring transcriptional kinetics from sequencing

To measure kinetics using this model, we need two pieces of information. First is a source of nascent RNA sequencing that can provide coverage and junctional information. Second is a method of estimating the amount of time it takes Pol II to move between different locations within a given gene. A feature of our model is it can operate using various types of data including SKaTER-seq data or the combination of TT-seq and PRO-seq (TT-SKaTER) (Figure 2A). In SKaTER-seq, cells are treated with DRB for three hours to block transcription at promoter proximal pause (Darzacq et al., 2007; Singh & Padgett, 2009). This synchronizes the cell population and allows transcription to move down each individual gene in the genome in a discrete wave following DRB washout (Singh & Padgett, 2009). We collect samples at 5-minute intervals for the first 35 minutes following washout to capture a series of snapshots of this wave of transcription. This is long enough for the transcription wave to reach the end of most genes.

Alternatively, we can collect nascent transcriptional information using TT-seq, which uses a 4sU labeling and enrichment to isolate nascent transcripts (Schwalb et al., 2016). To estimate the amount of time separating any two nucleotides, we use PRO-seq to capture a snapshot of the locations of engaged polymerase within the gene (Kwak et al., 2013). We model the relative amount of time between any two nucleotides to be proportional to the number of PRO-seq reads separating them.

To fit the set of kinetic parameters that best describes the coverage of a given gene, we employ a differential evolution optimization algorithm (Storn & Price, 1996) (Figure 2B-C). For example, we plot observed coverage of final nucleotides from SKaTER-seq for RAC1 at each of

the time points collected against predicted coverage using our best fit kinetics (Figure 2D) and repeat this process to fit the same gene using TT-SKaTER (Figure 2E).

To validate our model, we developed an independent stochastic modeling algorithm to randomly generate TT-seq reads for a simple gene containing a single intron using a given set of input transcriptional kinetics. As modeled sequencing depth increases, the optimizer converges to the input kinetics (Figure 2F), regardless of intron length (Supplemental Figure 2A). For elongation rates, we find that TT-SKaTER has a dependence on both sequencing depth and the length of the longest intron or included post CPS region, but is always able to accurately fit rates at TT-seq depths over two fragments per nucleotide (Figure 2G, Supplemental Figure 2B-F). To explore the model's accuracy for alternative splicing, we replaced the constitutive intron with alternative 5' splice site, alternative 3' splice site or a cassette exon and found the model was able to accurately measure all transcriptional processing kinetics regardless of architecture (Figure 2H). These results show that TT-SKaTER is highly accurate at high sequencing depths and is a reasonable approximation at lower depths.

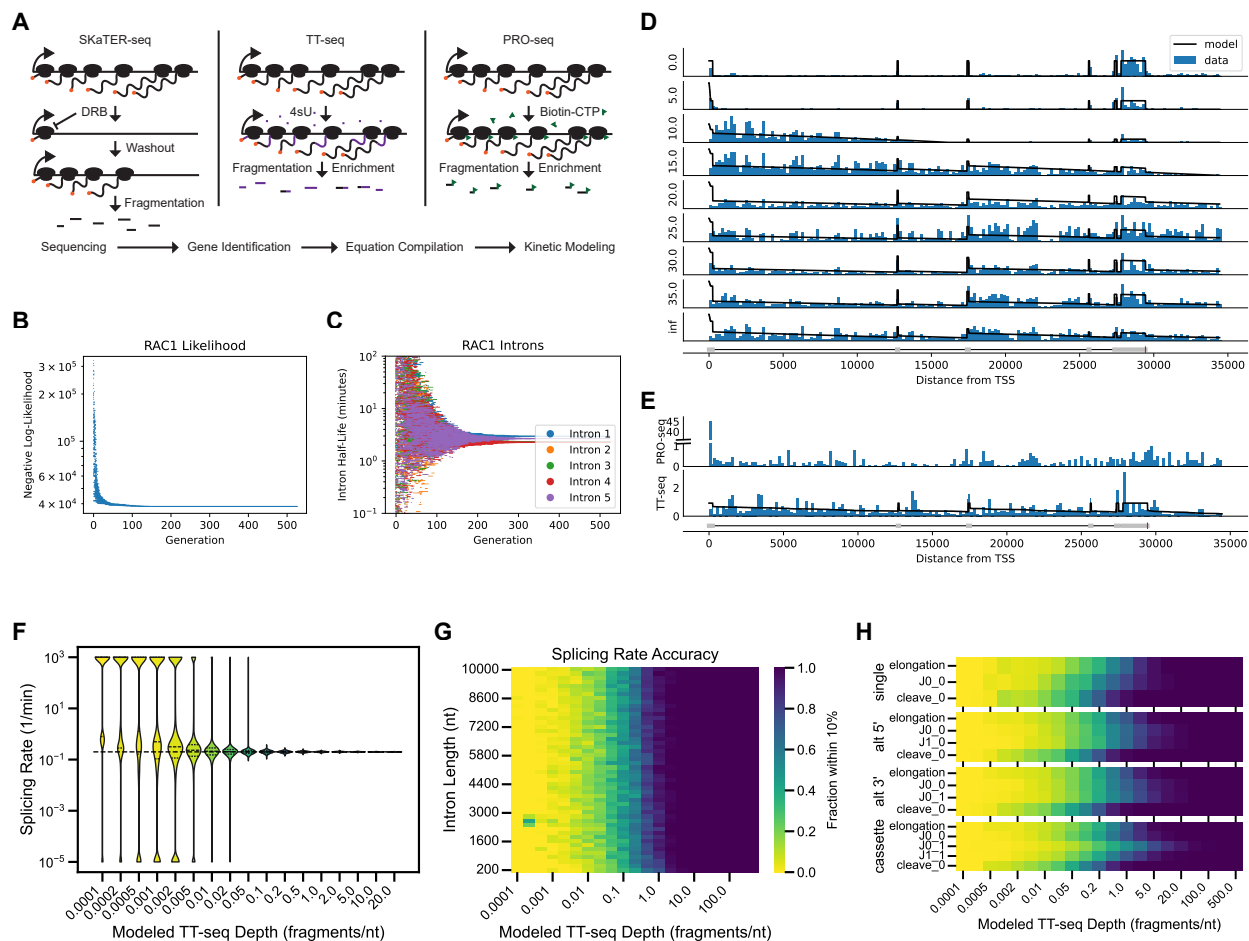


Figure 2. SKaTER and TT-SKaTER measure transcriptional kinetics: (A). Schematics of data sources for SKaTER-seq and TT-SKaTER. (Left) In SKaTER-seq, cells are treated with DRB then are collected at multiple time points following washout. (Middle) TT-seq steady state cells are labeled with 4sU to enrich for only nascent transcripts. (RIGHT) PRO-seq uses biotinylated CTP is to accurately map polymerase positions. SKaTER-seq data or a combination of TT-seq and PRO-seq datasets were provided to our model for kinetic analysis. (B) Kinetics are fit using a differential evolution optimization algorithm where we generate a set of parameters and calculate a likelihood of the parameters explaining the observed coverage. With increasing generation of the optimization, the algorithm gradually converges to the lowest negative log-likelihood. (C) Parameter populations start at a wide array of potential values but gradually converge to single values at high generations. (D) Sample fit of SKaTER-seq data from RAC1. Observed coverage for each time point (DRB blocked 0 minutes to steady state unblocked (inf)) with modeled fit overlaid in blue. (E) Sample fit of RAC1 using TT-SKaTER. Polymerase positions from PRO-seq provide estimated distances between different nucleotides while TT-seq provides coverage information. (F) Estimated splicing rates of stochastically modeled TT-seq data for sample gene with a single constitutive intron at from 0.0001–20 fragments/nt. Input splicing rate is 0.2 min^{-1} . Color indicates the fraction of time the estimated rates are within 10% of the input rates with blue indicating rates are always within this window and yellow indicating rates are never within the window. With increasing TT-seq sequencing depth, the distribution of parameters centers on the correct value before collapsing to the input splicing rate. (G) Elongation rate accuracy for

simple gene containing a single intron lengths from 200–9800 nt at modeled at TT-seq depths from 0.0001–500 fragments/nt with. Splicing and cleavage rates were modeled at 0.2 min^{-1} and 0.3 min^{-1} respectively, and elongation rates were modeled at 1, 2, and 4 kb/min. We observe increasing accuracy with higher sequencing depths and longer introns. (H) Kinetic parameter accuracy for alternative events plotted against modeled sequencing depth on the x-axis. All genes were modeled using a 2100 nt intron with a 100 nucleotide alternative segment. Splicing rates for each intron was modeled at 0.03, 0.2, and 0.9 min^{-1} with fixed cleavage and elongation rates of 0.3 min^{-1} and 2 kb/min respectively. We find that TT-SKaTER is able to accurately estimate all kinetics at high sequencing depths.

SKaTER can reliably measure kinetics across multiple cell lines

To validate our methods, we collected two replicates of SKaTER-seq and two replicates of TT-seq and PRO-seq in IMR90 lung fibroblasts. Each of these datasets were fed into their respective algorithms to estimate transcriptional kinetics. SKaTER-seq measured mean elongation rates of 3.5 kb/min while TT-SKaTER rates were slightly slower at 2.2 kb/min (Figure 3A). This is consistent with previously reported elongation rates of 1–6 kb/min with an average of 2–3 kb/min (Danko et al., 2013; Gressel et al., 2017; Jonkers et al., 2014). The discrepancy between the two methods can be explained by insufficient sequencing depth as TT-seq had higher resolution at lower elongation rates while SKaTER-seq is more accurate at higher rates. We observe high correlations between replicates for both methods (Supplemental Figure 3A-B). To compare the rates predicted by the two methods, we reran the TT-SKaTER samples merging the two replicates and compared elongation rates in highly fit windows to SKaTER-seq elongation rates and found a similarly strong correlation (Supplemental Figure 3C).

We repeated the analysis using splicing rates where we find median intron half-lives of 1.7 minutes using SKaTER-seq and 2.8 minutes using TT-SKaTER (Figure 3B). These rates are similar to those previously reported genome-wide in *Drosophila* (Pai et al., 2017). We find high correlations between replicates (Supplemental Figure 3D-E) indicating the data is highly reproducible. However, we observe only a moderate correlation between the two methods (Supplemental Figure 3F). For cleavage rates, we find median half-lives of 2.0 minutes for SKaTER-seq and 2.5 minutes for TT-SKaTER (Figure 3C) and correlated the replicates (Supplemental Figure 3G-H). Like splicing rates, we observed only a moderate correlation between methods (Supplemental Figure 3I). For both cleavage and splicing rates, SKaTER-seq

measurements are confounded by the presence of mature mRNAs contaminating the sequencing data. While we try to account for this effect using a DRB blocked time point, it still introduces a large amount of variability and can inflate splicing and cleavage rates. Overall SKaTER-seq is more accurate at determining elongation rates, while TT-SKaTER improves the estimation of splicing and cleavage rates

To expand this dataset, we collected another two replicates of SKaTER-seq in A549 cells and publicly available TT-seq and PRO-seq samples from an additional five cell lines. We find widespread strong correlations for all kinetics between most of the different lines, with the strongest correlations between datasets from the same cell lines (Figure 3D-F). Taken together, this data indicates that SKaTER-seq and TT-SKaTER are able to reproducibly estimate transcriptional kinetics, and these rates are highly associated between different cell lines. This strong association potentially indicates the presence of shared regulatory factors across the different cell lines.

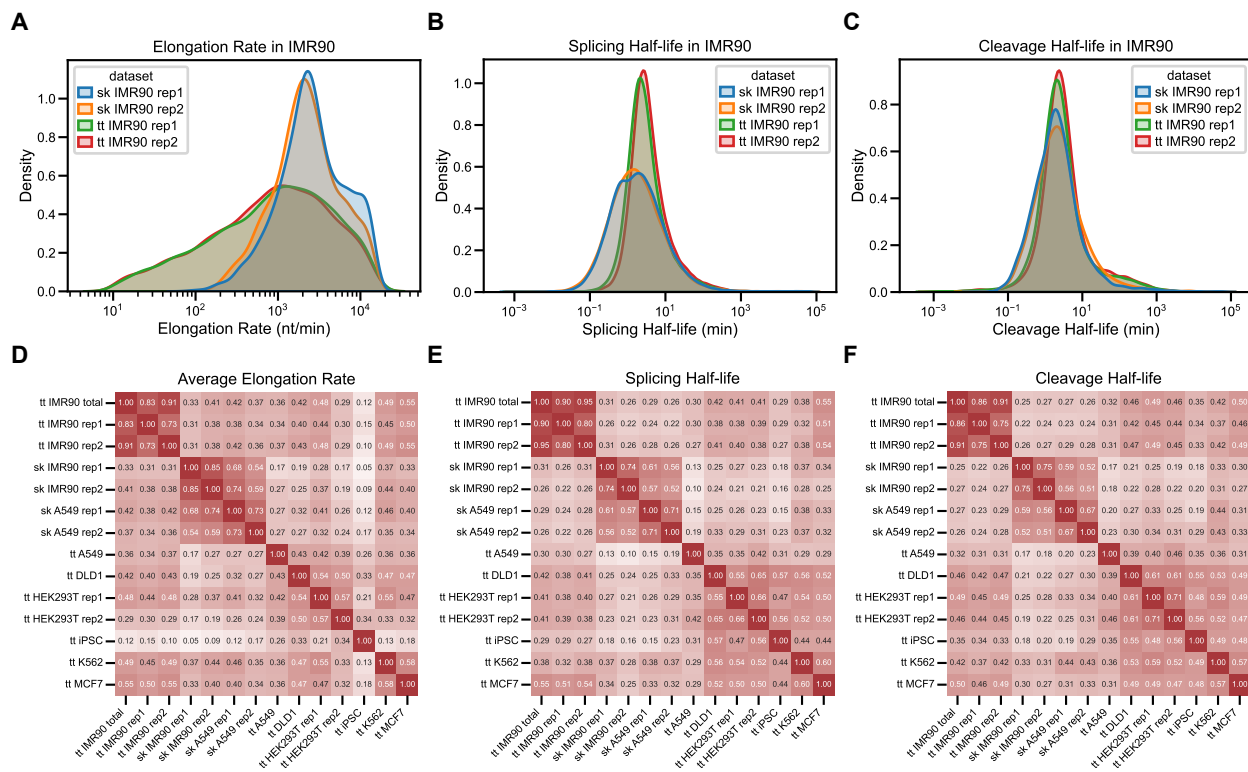


Figure 3. SkaTER reliably measures kinetics from both SkaTER-seq and TT-seq/PRO-seq: (A) Distribution of gene average elongation rates for 2 samples of SkaTER-seq and 2 samples of TT-SkaTER in IMR90 cells. Average elongation rates are 3.5 kb/min for SkaTER-seq and 2.2 kb/min for TT-SkaTER. (B) Correlation of elongation rates in 5 minute windows between SkaTER-seq replicate 1 and merged TT-SkaTER datasets. While data is highly correlated, we note elongation rates from TT-SkaTER are notably lower than SkaTER-seq rates. (C) Distribution of cleavage half lives with median values of 2.0 minutes and 2.5 minutes for SkaTER-seq and TT-SkaTER respectively. Heatmap of spearman correlation coefficients comparing gene average elongation rates (D), splicing rates (E), and cleavage rates (F). Rates are highly correlated between replicates, although there are generally widespread correlation between replicates.

Cleavage rates are regulated by RNA motifs

The PAS, canonically a AAUAAA motif, has a defined role at promoting efficient cleavage. In genes with multiple cleavage sites, the AAUAAA motif frequently occurs at the distal site, while weaker motifs occur at proximal sites (Tian & Manley, 2017). Recent studies in *Drosophila* indicate that cleavage rates are positively correlated with the number of nearby PAS motifs (Torres-Ulloa et al., 2024). To validate our cleavage rates, we compared cleavage rates between genes with and without a canonical PAS motif within a 100-nucleotide window upstream of the cleavage site. Across all our datasets, we find a significant increase in cleavage rates when a single PAS motif is present but no further increase with additional PAS (Figure 4A, Supplemental Figure 4A-B). This shows mammalian cleavage regulation is dependent on the presence of the PAS motif, not its frequency like in *Drosophila*.

Previous studies have suggested a positive interaction between cleavage and splicing of the terminal intron, mediated by interactions between cleavage and splicing factors (Dye & Proudfoot, 1999; Reimer et al., 2021). Therefore, we correlated splicing rates for final introns with cleavage rates of constitutive cleavage sites and found a strong positive association between cleavage and splicing. We also correlated cleavage rates with splicing rates of the first intron and found the correlations were weaker across all samples (Figure 4B). This suggests a coordination between final intron splicing and cleavage.

After observing the strong effect of PAS motif and splicing on cleavage rates, we wanted to determine if any RBP motifs have an association with cleavage rates. We collected a series of RBP motifs from RMAPs (Hwang et al., 2020), binned genes based on the presence or absence of the motif within the 100 nucleotide window upstream of the CPS, and performed a Student's T-test

comparing the cleavage rates for constitutive sites with the motif compared to without (Figure 4C). The strongest association was with the PAS motif, and the second-strongest association was with KHDRBS3, whose motif is highly similar to the canonical PAS motif and therefore was treated as a false positive. The third-strongest motif was SRSF9, which appeared to inhibit cleavage (Supplemental Figure 4C). Interestingly, we find SRSF3 to also be a cleavage inhibitor, matching previous reports (Schwich et al., 2021). SRSF9 is known to regulate alternative splicing and drive progression of multiple forms of cancer (Luo et al., 2024; Wang et al., 2022; Zuo et al., 2025). However, there is no documented association between SRSF9 and cleavage and polyadenylation regulation. Plotting SRSF9 motif enrichment around the CPS showed that the slowest 20% of genes had a significant enrichment of the motif around 100 nucleotides upstream of the CPS (Figure 4D). We repeated the analysis across all our datasets and found a significant increase in cleavage half-lives in the presence of the SRSF9 motif in most of the other samples (Figure 4E). This indicates SRSF9 binding motifs near the CPS can regulate cleavage rates.

The above results indicate that SRSF9 can act as a potential cleavage inhibitor. To test this hypothesis, we collected a SRSF9 CRISPR knock-down RNA-seq dataset from ENCODE (The ENCODE Project Consortium, 2012; Kagda et al., 2025; Luo et al., 2020). We then employed QAPA, a method for measuring PAS usage from total RNA-seq (Ha et al., 2018), to measure the relative usage in different APA sites in both wild type and knockdown conditions. QAPA employs Salmon (Patro et al., 2014; Patro et al., 2017) to assign fragments to different alternative polyadenylation isoforms. Using these values, we used a Cochran Armitage rank-sum test to calculate the significance of a change in 3' UTR length for genes in the SRSF9 KD samples as compared to wild type (Figure 4F) as can be seen in CTNS and PPP6R3 (Figure 4G-H). We saw

a notable shift towards shorter 3' UTR lengths, which we quantified using a one-sample Student's T-test. These findings indicate SRSF9 regulates APA outcomes by acting as a cleavage inhibitor.

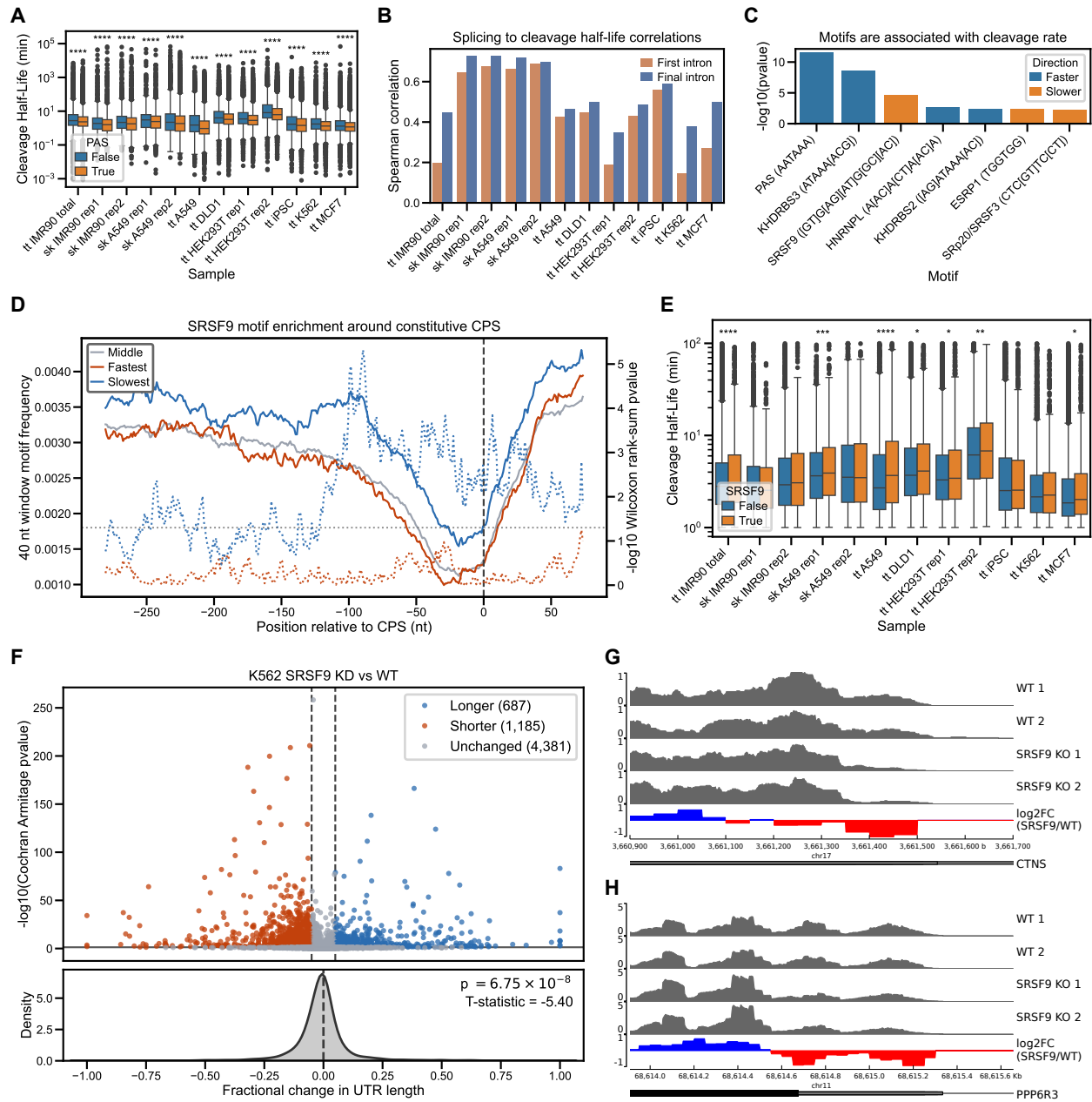


Figure 4. Cleavage rates are regulated by local RNA motifs: (A) Constitutive cleavage rates (show as half-life) across all cell lines, grouped by the presence or absence of the canonical PAS motif (AAUAAA) within a 100 nt window upstream of the cleavage site. Across all genes, sites with the canonical PAS have significantly shorter cleavage half-lives. (B) Bar plots of Spearman correlations between splicing rates of the final intron and cleavage rates indicate widespread positive correlation. This effect is stronger than the association with first intron splicing rates across all samples. (C). A screen of RBP motifs shows SRSF9's motif was significantly associated with slower cleavage rates. (D). Meta-plot of SRSF9 motif shows a significant enrichment for SRSF9 motifs among the slowest 20% of genes 100 nucleotides upstream of the CPS. Dotted lines indicate significance measured by negative $\log_{10}$ Wilcoxon rank-sum p-value. (E). Analysis across all samples of genes with cleavage rate half-lives of over 1 minute shows genes with the SRSF9 motif

have significantly slower cleavage rates. (F). Analysis of CPS usage in an SRSF9 knockdown shows significant shift towards shorter 3' UTR length, consistent with faster cleavage and SRSF9 activity as a cleavage inhibitor. (G-H). Sample genes (CTNS, PPP6R3) where SRSF9 Knockout is associated with a shortening of 3' UTR length

Splicing rates are spatially coordinated

Based on the previously identified interaction between cleavage rates and splicing rates, we investigated if introns that are close to each other in space are also correlated kinetically. We plotted splicing rates for the first three and last three constitutive introns. We found that across all cell lines, early introns spliced slower than the later introns (Figure 5A, Supplemental Figure 5A), matching previous reports (Bedi et al., 2021; Pai et al., 2017). We found that splicing rates between neighboring introns were highly correlated (Figure 5B). To separate local and gene-wide effects, we performed a permutation test by randomly rearranging the order of introns within a given gene and calculating the Spearman correlation. We found that while the permuted introns still had a strong correlation, the true correlation was much higher (Figure 5C). This observation holds across all our datasets (Supplemental Figure 5B). These results indicate that splicing is coordinated both locally between neighboring introns and at a gene-wide level, suggesting that both local and gene-wide factors play an important role in regulating the kinetics of splicing.

To validate our splicing rates, we collected splice site strength information using MaxENTScan (Yeo & Burge, 2004) for each intron. To account for the gene effect, we correlated the splicing rates of all the introns in a particular gene to their 3' splice site strengths (Figure 5D upper). This showed that most genes had a minor correlation between their intron splicing rates and 3' splice site strength estimated using a Student's T-test of the underlying Spearman correlation coefficient distribution (Figure 5D middle). To estimate the significance of this effect, we performed a permutation test by scrambling the association between splicing rates and splice site strength within each individual gene and recalculating Spearman correlations and T-statistics for the overall distributions. We find that the observed positive correlation between splicing rates

and 3' splice site strength is higher than expected by random chance (Figure 5D lower). Repeating this process across all our datasets showed widespread positive correlations between splice site strength and splicing rates, with 3' splice sites more significant than 5' (Supplemental Figure 5C). We also observed faster splicing with decreasing GC richness (Figure 5E) corresponding to previous reports in *Drosophila* (Pai et al., 2017). However, the effects of GC richness and intron length seem to vary between cell lines (Supplemental Figure 5C). These results indicate that, while intron specific factors have an impact on overall splicing rates, their impact is masked by much stronger gene level factors.

To test how multiple factors combine to predict splicing rates, we trained a generalized linear mixed model of splicing using variables including gene, topologically associated domain (TAD), relative position within the gene, neighboring splicing rates, intron length, GC content and splice site strengths. These factors alone were able to account for 48% of the variance in the samples. To test the significance of each factor, we retrained the model by removing one variable at a time and comparing the fits of the models using a likelihood-ratio test. We found that spatial factors such as neighboring splicing rates, relative position of the intron, gene and TAD had much higher statistical significance than intron specific factors such as GC content and splice site strengths (Figure 5F). This indicates spatial factors are more predictive of splicing rates than intron specific factors.

To explore potential mechanisms for this spatial coordination of splicing rates, we calculated the similarity of gene average splicing rates for genes within a TAD compared the rates with all other genes and calculated the probability of seeing this association using a Kolmogorov-Smirnov test. We then summed the negative $\log_{10}$ p-values to compute a genome wide statistic

for the association of gene average splicing rates and TADs. To determine the significance of this statistic, we employed a permutation test to randomly scramble the mapping of each gene to their respective TADs and recompute the statistic. This distribution showed that the true score was higher than expected by random chance (Figure 5G) across all cell lines with TAD data (Supplemental Figure 5D), suggesting TAD effects are partially responsible for gene level effects. We repeated this process for cleavage and elongation rates across our samples but found the effects to be inconsistent (Supplemental Figure 5E-F). This shows spatial coordination of transcription seems to be uniquely important for splicing rates.

Another potential mechanism for some of the spatial coordination of splicing rates is the effect of nuclear speckle proximity. Nuclear speckles are phase separated condensates rich in splicing factors and have been shown to impact splicing efficiency (Małszycki et al., 2026). Therefore, we correlated gene average splicing rates with their nuclear speckle proximity as measured by TSA-seq in K562 cells (Chen et al., 2018; Zhang et al., 2021) and found a significant positive correlation, with introns closer to speckles having higher splicing rates across only some of the cell lines (Figure 5H, Supplemental Figure 5G) with iPSCs having the greatest association and IMR90 and A549 having the least. We repeated this analysis for cleavage and gene average elongation rates (Supplemental Figure 5H-K) and found negative correlations between speckle proximity and elongation rates across all samples except iPSCs which had a strong positive correlation. This indicates nuclear speckles are associated with regulation of transcriptional kinetics in some cell lines but cannot explain the gene and TAD level coordination of splicing rates.

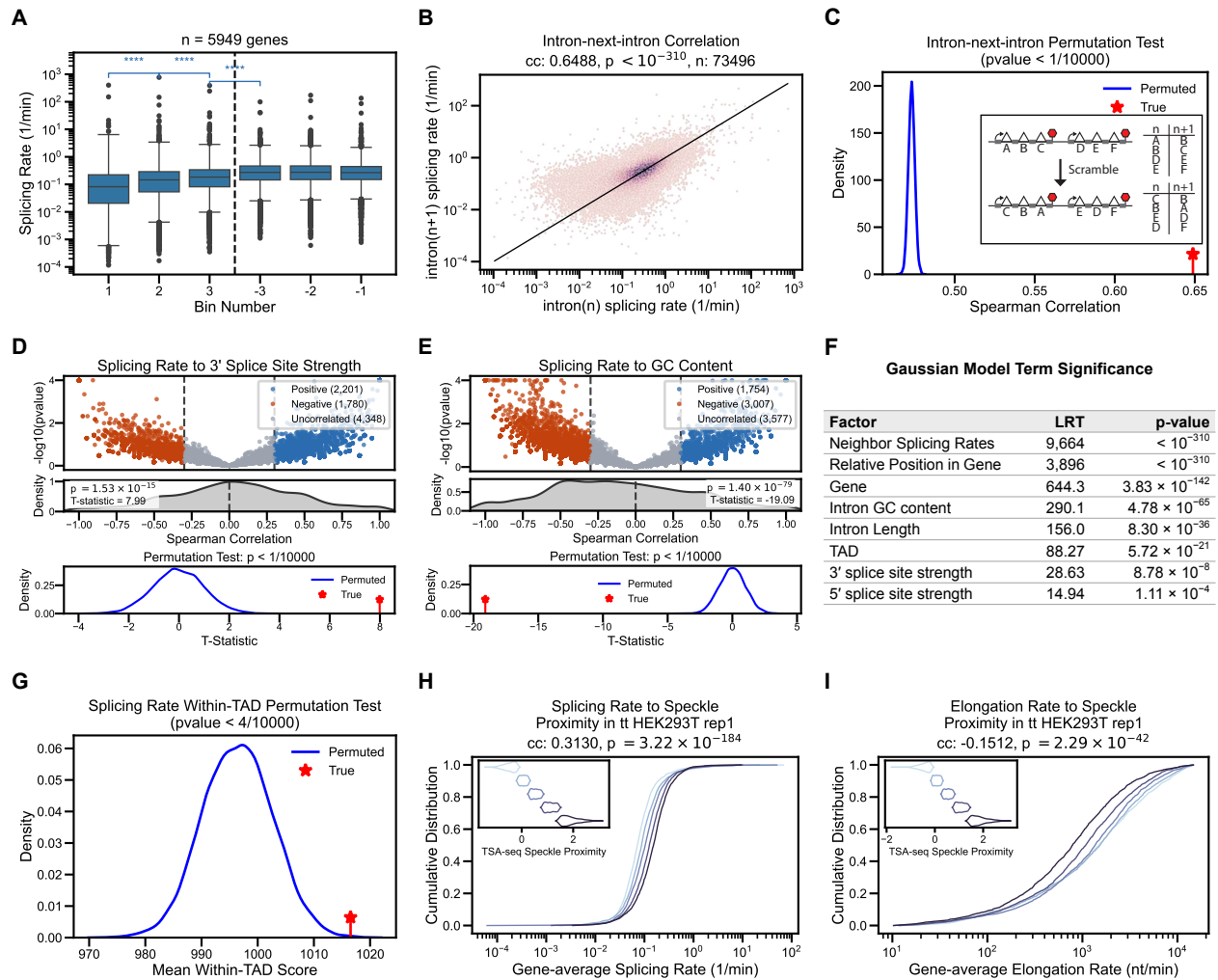


Figure 5. Splicing rates are strongly regulated by spatial factors: (A) Plot of splicing rates comparing the first three introns and last three introns genome wide in IMR90. Early introns have slower splicing rates than later introns. (B) Splicing rates for constitutive introns show a strong correlation with their next downstream constitutive intron within the same gene. (C) Scrambling the pairings of intron to their nearest downstream neighbor shows that the true correlation is much higher than would be expected by random chance. Additionally, introns within the same gene have highly correlated splicing rates, indicating the presence of a gene level effect on splicing rates. (D) (Upper) Scatter-plot of Spearman correlation vs negative $\log_{10}$ p-value comparing splicing rates with 3' splice site strength, with each gene represented by a single point. (Middle) Student's T-Test of the underlying Spearman correlations shows the population is shifted towards a positive correlation. (Lower) This was tested using a permutation test by randomly scrambling the mapping between splice site strengths and splicing rates. This found that the association was highly significant. (E) Similar analysis showing stronger negative association between splicing rate and intronic GC content. (F) Results of Gaussian model combining factors known to effect splicing rates. Likelihood-ratio tests (LRT) show the relative importance of each model factor. Rows are ordered by decreasing statistical significance, showing positional effects such as neighboring intron splicing rates, relative position within the gene, and gene identity are much more significant than intron intrinsic factors such as GC content, intron length, and splice site strengths. (G)

Association between gene average splicing rates and TAD calculated by summing the negative $\log_{10}$ p-value from a KS-test comparing rates within a given TAD to all other genes. Permuting the mapping of genes to TAD and computing a new significance score shows that gene average splicing rates are clustered by TAD. (H) Splicing rates of introns are associated with nuclear-speckle proximity in HEK293T cells, as measured by TSA-seq. (I) Meanwhile, elongation rates are negatively associated with nuclear-speckle proximity.

Elongation rates are determined by the local chromatin environment

Polymerase elongation rates are known to vary widely over the course of a gene, accelerating as it moves away from the promoter proximal pause (Danko et al., 2013; Jonkers et al., 2014). Plotting elongation rates against position from the TSS to the CPS of our SKaTER-seq data shows an acceleration moving away from the transcription start sites, with longer genes reaching higher elongation rates overall (Figure 6A). We see the same phenomenon in TT-SKaTER, where there is a clear positive association between gene length and gene average elongation rate across all of our investigated cell lines (Figure 6B, Supplemental Figure 6A).

Previous reports also indicated that elongation rates are slower in high GC regions (Mimoso & Adelman, 2023). After masking the promoter proximal region, we divided the gene body of all the genes into windows that are at a minimum 5 minutes long and contain at least 50 nucleotides to match the minimum window sizes measured by SKaTER-seq. We collected the first five windows per gene and correlated with local GC content. We see a very strong negative association between GC content and elongation rate across both methods and all cell lines (Figure 6C, Supplemental Figure 6B). These results show our elongation rates exhibit features previously observed in the literature.

Chromatin marks such as H3K4me3 have been documented to regulate elongation rates (Wang et al., 2023). Other factors such as H3K36me3, H4K20me1, and CTCF have been used to predict elongation rates in machine learning algorithms (Liu et al., 2025). To test the association between chromatin marks and elongation rates, we collected a set of histone and transcription factor ChIP-seq and ATAC-seq in IMR90 cells from ENCODE (The ENCODE Project Consortium, 2012; Kagda et al., 2025; Luo et al., 2020). We divided the transcriptome into

windows corresponding to at least five minutes of transcription and at least 50 nucleotides long (excluding the TSS). To account for potential biases introduced by including windows of different sizes, we gathered 1000 random samplings of each window size from the transcriptome and normalized the mark intensity. We then calculated the Spearman correlation of normalized mark intensity with elongation rates and found that none of the marks showed a particularly strong association with elongation rates (Figure 6D, Supplemental Figure 6C-D). To determine if there were any combinatorial effects of histone marks on elongation rates, we trained a random forest regression model on the data. The regression model identified H3K4me3, ATAC-seq chromatin accessibility, H3K27me3, and H3K56ac as the most important features for predicting elongation rates (Figure 6E). The model was able to accurately predict elongation rates in the test data with an R^2 of 0.7361 (Figure 6F). We repeated this process using our SKaTER-seq windowed data and found ATAC, H3K4me3, and H3K56ac were similarly highly predictive with R^2 of 0.844 and 0.837 for replicates 1 and 2 respectively (Supplemental Figure 6E-H) This was striking as these marks did not show strong correlations with elongation rates, indicating a combinatorial effect between these marks.

To understand this effect, we plotted the predicted elongation rates for H3K4me3 against H3K27me3 for the trained random forest regression model (Figure 6G). We see that most combinations of H3K4me3 and H3K27me3 intensity result in slow elongation rates. However, intermediate H3K4me3 intensity or low H3K27me3 with high H3K4me3 intensity are compatible with faster elongation rates. To validate the predictions, we overlaid a topological map of the windows from the test data with elongation rates over 3 kb/min in black and a second map of elongation rates under 50 nt/min in white. The histone mark intensity from these regions closely

matches the regions of high elongation rates predicted by the random forest regression model. Similarly, comparing H3K56ac intensity against ATAC-seq chromatin accessibility showed that specifically moderate H3K56ac intensity is compatible with higher elongation rates (Figure 6H, Supplemental Figure 6I-M). Taken together, this data indicates that just the right amount of particular combinations of chromatin marks, or a histone code, that is required for productive elongation, and moving out of these regions results in polymerase slowing.

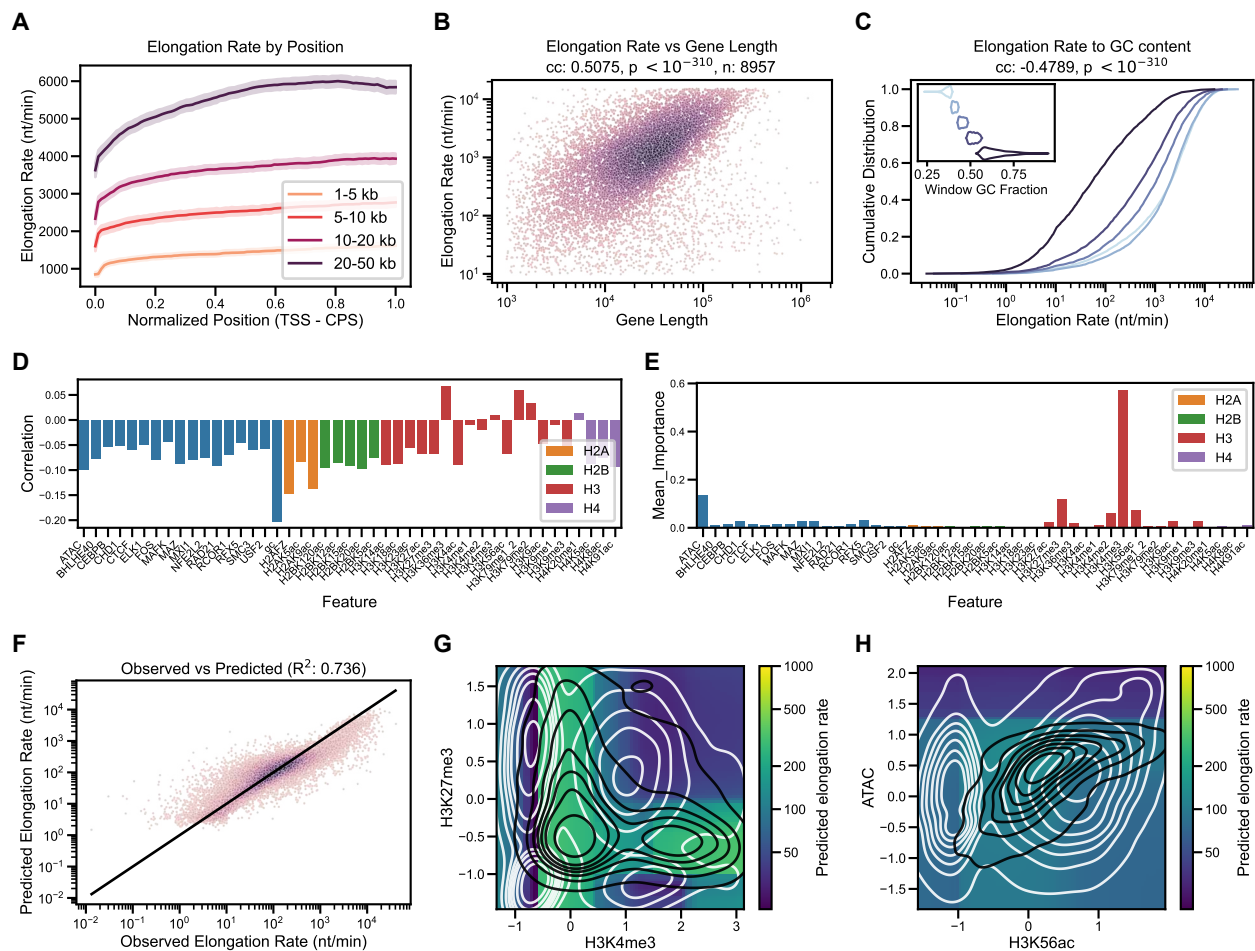


Figure 6. Elongation rates vary with genomic context and chromatin state.: (A). Elongation rates measured from SKaTER-seq are binned by gene length and plotted as a function of relative position along the gene body. This shows elongation rate increasing both with distance from the TSS and with increasing gene length. (B). Comparison between TT-SKaTER elongation rates and gene length reproduce the strong positive association. (C). TT-SKaTER elongation rates in multiple windows (each at least 5 minutes and 50 nt long) are plotted against local GC content CDF, showing a strong negative association between GC content and elongation rate. (D). ChIP signal for multiple marks including ATAC-seq, transcription factors, and histone marks was collected for multiple windows (excluding the first 500 nucleotides of each gene), normalized based on randomly selected windows of the same size within the transcriptome, and correlated with elongation rates. Most marks show minor negative associations. These marks were used to train a random forest regression model to predict elongation rates. (E). Permutation importance ranks chromatin features by how much randomizing each feature reduces random forest prediction accuracy. Higher importance values provide more unique predictive information to the model after accounting for the other marks. H3K4me3, H3K27me3, ATAC, and H3K56ac are the most important marks for elongation rate predictions. (F). The random forest model trained on the chromatin features is able to accurately predict elongation rates for held-out data. (G-H). A two-dimensional partial-dependence heatmap shows the random forest-predicted elongation rate across combinations of H3K4me3 and H3K27me3 values or H3K56ac and ATAC, highlighting how elongation rates can be a function of combinations of histone marks. Overlaid on this heatmap in

black is a KDE of the histone mark intensity of windows with fast elongation rates (over 3 kb/min) and a second KDE of slow elongation rates (under 50 nt/min) in white.

Discussion

Although many types of genome-wide RNA sequencing are now routinely performed, algorithms to accurately derive kinetic parameters from these data have been lacking. Here we developed a state model of transcription that we employed to measure genome-wide transcriptional elongation, splicing, and cleavage rates simultaneously from either DRB-released nascent RNA sequencing data (SKaTER-seq) or a combination of steady state TT-seq and PRO-seq data (TT-SKaTER). TT-SKaTER is a notable improvement over other TT-seq based approaches (Caizzi et al., 2021; Mimoso & Adelman, 2023; Stein et al., 2022; Žumer et al., 2021; Žumer et al., 2024) because it properly accounts for TT-seq coverage lost due to intron splicing and reports absolute elongation rates rather than normalized metrics like elongation index. SKaTER-seq is similarly able to measure elongation rates directly from a nascent RNA preparation, without requiring PRO-seq or another polymerase mapping method (Danko et al., 2013; Jonkers et al., 2014). Both SKaTER-seq and TT-SKaTER directly report splicing and cleavage rates without requiring the assumption of an elongation rate or multiple labeling time points used by current methods (Carrillo Oesterreich et al., 2016; Pai et al., 2017; Reimer et al., 2021; Torres-Ulloa et al., 2024; Wachutka et al., 2019). While SKaTER-seq splicing and cleavage rates can be confounded by the presence of contaminating mature mRNAs, this effect can be accounted for through the addition of a *Drosophila* spike-in and may be further reduced through future improvements in the nascent RNA preparations. While both methods are able to estimate global transcriptional kinetics, they each specialize in particular ways. SKaTER-seq is highly accurate in measuring regions with faster elongation rates, by using the wave-front associated with DRB release. TT-SKaTER in turn has higher resolution for splicing rates and regions with slower polymerase elongation rates. The

accuracy of both methods for kinetics they are not specialized in can be further improved with increased sequencing depth.

We found that cleavage is promoted by the strength of the PAS motif, not the number of motifs. This suggests a different regulatory method than in *Drosophila*, where PAS motif frequency is a dominant feature (Torres-Ulloa et al., 2024). We also found that cleavage rates and final intron splicing rates are strongly associated across cell lines. This matches a hypothesized association between final intron splicing and transcript cleavage (Berget, 1995; Dye & Proudfoot, 1999; Reimer et al., 2021). This association is further supported by our finding that SRSF9 can act as a cleavage inhibitor. This joins SRSF3 and SRSF7 (Schwich et al., 2021) in the evolving role of splicing factors in regulating cleavage and polyadenylation.

We observe several notable factors associated with splicing rates. First, we observe a dominant effect of spatial proximity on splicing rates, ranging from neighboring introns all the way up to broader TAD level associations. This TAD level effect is conserved for splicing rates across all of our samples, while cleavage and elongation rates do not show the same consistent effect. This indicates that splicing may be uniquely coordinated spatially. Once these larger effects are accounted for, we find that intron-specific factors, including splice site strengths, intron length, and GC content, are predictive of splicing rates. Some of these factors, such as GC content, have effects that vary by cell line, indicating that these factors are not direct regulators of kinetics, but instead acted upon by cell-wide factors. This is supported by observations that efficient splicing of GC rich introns requires nuclear speckle proximity (Małszycki et al., 2026). This hypothesis is further supported by the striking differences between the iPSC line and the other differentiated cell lines. iPSCs show the strongest association between splicing rates, GC content, and nuclear

speckle proximity. Further, iPSCs are the only sample to show a strong positive association between elongation rates and nuclear speckle proximity. This suggests that these effects are lost during differentiation.

We found that elongation rates are highly dependent on their local chromatin environment. The study of the interactions between chromatin and elongation rates is complicated by the associations between different chromatin marks. For example, most of the H3 lysine acetyl marks are present in the same regions of chromatin, and we also observe strong associations between GC content, H2AK9ac and H2AFZ. GC content is further associated with nuclear speckle proximity (Małszycki et al., 2026). To overcome this limitation, we employed a random forest regression model, which selects for the feature that contains the most information. We identified three chromatin marks that were of particular importance for elongation rates in both TT-SKaTER and SKaTER-seq datasets. Chromatin accessibility, characterized by ATAC-seq, is a predictable factor as the transcription machinery has fewer nucleosomes to disrupt. H3K4me3 similarly was recently documented to directly regulate transcriptional elongation rates (Wang et al., 2023). However, our results indicate that it is not the presence of H3K4me3 that matters, but rather the precise amount, with minor shifts in H3K4me3 levels associated with much slower or faster elongation rates. The third factor we identified is H3K56ac, a poorly studied H3 acetyl mark. Previous studies of H3K56ac have been hampered by the high cross-reactivity between the antibody used for H3K56ac and other acetyl marks (Pal et al., 2016). However, the presence of the other H3 acetyl marks in the same dataset indicate that our findings are unique to H3K56ac, not just acetylated chromatin. A role of H3K56ac in elongation has been documented in yeast (Värv et al., 2010), but not in humans. This effect is potentially mediated by the location of H3K56ac on the histone

core potentially weakening interactions with DNA (Kim et al., 2015) or its ability recruit chromatin remodelers (Duan et al., 2025). Like H3K4me3, we also observe a dose dependence of H3K56ac, where moderate amounts are associated with fast elongation, but high or low amounts indicate slower rates. These studies open up new avenues for studying the role of chromatin in regulating transcriptional elongation rates, and thereby alternative mRNA isoform formation.

Data Availability

All RNA-seq, SKaTER-seq, TT-seq, and PRO-seq data used in this study is available on GEO (www.ncbi.nlm.nih.gov/geo/). SKaTER-seq datasets for IMR90 and A549 are available at GSE146290. IMR90 TT-seq and PRO-seq datasets are available at GSE344536. A549 PRO-seq data was collected from GSE337411, A549 TT-seq from GSE287535 (Zhu et al., 2025), DLD1 RNA-seq, TT-seq, and PRO-seq from GSE180845 (Hu et al., 2021), HEK293T RNA-seq, TT-seq rep1 and PRO-seq rep1 from GSE218135 (Mimoso & Adelman, 2023), HEK293T TT-seq rep2 and PRO-seq rep2 from GSE267239 (Baluapuri et al., 2025), iPSC RNA-seq, PRO-seq, and TT-seq from GSE283681, GSE273680, and GSE300859 respectively (Li et al., 2025), K562 RNA-seq from GSE148433 (Caizzi et al., 2021), K562 PRO-seq from GSE154746 (Vihervaara et al., 2021), K562 TT-seq from GSE242380 (Žumer et al., 2024), MCF7 RNA-seq from GSE218069 (Watanabe et al., 2024), MCF7 PRO-seq from GSE298773 (Kim et al., 2026), and MCF7 TT-seq from GSE225508 (Gómez Acuña et al., 2024).

SKaTER-seq and TT-SKaTER pipelines are available at <https://github.com/TheRealGambleLab/skater-cli>. Software used for analysis and figure generation is available at <https://github.com/TheRealGambleLab/rates2chip> and <https://github.com/TheRealGambleLab/StochasticRNAModeler>.

Author Contributions

A.D.C., M.J.G, K.Y. and J.D.S developed state model and conceived the study. A.D.C. collected SKaTER-seq samples. J.D.S. collected TT-seq and PRO-seq samples, derived general formulas, constructed the SKaTER pipeline, and performed the analysis. Original draft was prepared by J.D.S. and M.J.G., A.D.C., and C.C.Q. reviewed and edited the text. M.J.G. and C.C.Q. supervised the work.

Conflicts of interests: Authors declare no competing interests.

Acknowledgements

We thank Dr. Kristy Stengel, Ariel Raskin, Dr. Scott Heibert, and Dr. Anna Gilbert for their help with PRO-seq. We thank the Einstein High Performance Computing Core for computer cluster administration. This work was supported by R01GM134379 and 5T32GM149364-04.

Methods

Cell Culture

IMR90 primary human fetal lung fibroblasts (ATCC CCL-186) were cultured in MEM supplemented with 10% fetal bovine serum (FBS) and 1x Penicillin-Streptomycin (pen-strep). A549 lung cancer cells (ATCC CCL-185) were cultured in RPMI supplemented with 10% FBS and 1x pen-strep. *Drosophila melanogaster* S2 cells (ATCC CRL-1963) were cultured at 28°C in Schneider's *Drosophila* Medium supplemented with 10% heat-inactivated FBS and 1x pen-strep.

SKaTER-seq

Cell treatment and nascent RNA isolation protocols were loosely adapted from previous studies and optimized for our method (Carrillo Oesterreich et al., 2016; Khodor et al., 2012; Mayer

et al., 2015; Nojima et al., 2015; Singh & Padgett, 2009; Wuarin & Schibler, 1994). Cells were grown to 90% confluence in nine 10 cm dishes (one dish per time point). Cells were treated with 100 μ M 5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside (DRB, Sigma D1916-50MG) for DMSO (Fisher Scientific BP231100) as a vehicle control for 3 hours. After 3 hours, cells were washed twice in 37°C 1x PBS and 37°C fresh media was added to each plate, except the 0-minute time point which was never washed out. Cells were immediately returned to the 37°C incubator. At each 5 minute time point, one plate of cells was removed from the incubator, quickly washed once in ice-cold 1x PBS, then 1 mL Cytoplasmic Lysis Buffer + *Drosophila* S2 cells (CL+S2 buffer, 25 mM Tris pH 7.9, 150 mM NaCl, 0.1 mM EDTA, 0.1% TritonX, 1 mM DTT, 1x protease inhibitor cocktail, 1.3×10^6 S2 cells/mL) was added directly to the plate. Cells were collected by scraping and kept on ice until the end of the time course.

After all samples were collected and incubated in CL+S2 buffer on ice for at least 10 min, the lysates were centrifuges at 3000 RPM for 5 min at 4°C. The supernatant was removed, and the pellet was washed in 1 mL CL buffer (without S2 cells) and re-centrifugated at the same speed. The pellet was then resuspended in 100 μ L Glycerol Resuspension Buffer (GR buffer, 20 mM Tris pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT), to which 1.1 mL Nuclear Lysis Buffer (NL buffer, 20 mM HEPES pH 7.6, 300 mM NaCl, 7.5 mM MgCl₂, 1% NP-40, 1 mM DTT, 1 M Urea) was subsequently added. The resulting lysate was incubated on ice (with regular vortexing) for 15 min, then centrifugated at 16,000 G for 10 min at 4°C. The resulting chromatin pellet was resuspended in TRIzol (Invitrogen 15596018) and RNA was extracted according to the manufacturer's protocol. To further deplete mRNA, the NEBNextPoly(A) mRNA Magnetic Isolation Module (NEB, E7490L) was used to capture and remove polyA+ mRNA. The

manufacturer's protocol was followed until the mRNA was bound to the beads. Following mRNA binding, the supernatant was removed from the beads and the remaining nascent RNA was ethanol precipitated and resuspended in ddH₂O.

Nascent RNA was quantified on a Denovix QFX fluorometer using the Qubit RNA HS assay according to the manufacturer's protocol. The quality of the RNA was assessed using an Agilent 2100 Expert Bioanalyzer instrument. Sequencing libraries were generated using the KAPA Stranded RNA-seq Kit with RiboErase (HMR) (Roche, 07962282001) according to the manufacturer's protocol. Final libraries were also subjected to bioanalyzer runs to ensure proper fragment length distribution. All 9 libraries were barcoded and sequenced together on two lanes of an Illumina HiSeq (paired-end 150 bp reads, stranded total RNA-seq).

TT-seq

TT-seq was adapted from previously published protocol (Schwalb et al., 2016). Cells were grown to 70% confluence in two 15 cm plates. One dish was treated with 500 μ M 4-thiouridine (4sU, Sigma T4509-25MG) for 10 minutes. After labeling, media was immediately aspirated and cells were collected by scraping into 2 mL TRIzol (Invitrogen 15596018). The second unlabeled plate was similarly collected by scraping into TRIzol. Enrichment, library preparation and sequencing were performed at the Harvard Nascent Transcriptomics Core using a *Drosophila* spike-in following the published protocol (Schwalb et al., 2016).

Alignment

SKaTER-seq, TT-seq, and RNA-seq libraries were aligned to hg38 (or a combined hg38 dm6 genome where applicable) using STAR v2.7.11 (Dobin et al., 2013) with --

outFilterMultimapNmax 1. PCR duplicates were removed using PicardTools MarkDuplicates v2.26.10 using default parameters (*Picard Toolkit*, 2019).

PRO-seq

PRO-seq was performed following the rapid PRO-seq protocol (Cingaram et al., 2025) using 25 μ M Biotin-11-CTP and 25 μ M UTC, ATP, and GTP. Sequencing was performed by VANTAGE Sequencing Core. Sequencing adaptors were trimmed using Cutadapt 3.5 (Martin, 2011), and alignment to hg38 was performed using BWA v0.7.17 (Li & Durbin, 2009).

Kinetic Measurements

Gene Identification

The first step of the SKaTER-seq and TT-SKaTER analysis pipeline is creating an annotation of the genome. Starting with a RefSeq annotation, we collect the set of all TSS and PAS corresponding to a given gene. Then, we *de novo* identify a set of introns from RNA-seq data between the upstream most TSS and the downstream most CPS. Introns are filtered to include only those longer than 50 nucleotides and with junctional coverage of at least 4 reads. Additionally, we discard any introns whose junctional coverage is under 5% of the maximum junctional coverage for the gene. Genes were retained for SKaTER-seq when coverage averaged at least 1 reads per kb in at least one selected time point; TT-SKaTER required mean PRO-seq and TT-seq coverage of at least 1 read per kb.

The current pipeline is limited to genes with a single TSS. If a gene has two or more annotated TSS, we calculate the distance between the sites. We consider TSSs within 150 bp to be sufficiently close to be modeled using the upstream TSS. If the start sites are over 150 bp

apart, we use the coverage from the first 200 nucleotides PRO-seq in TT-SKaTER or the 100 to 300 nucleotides 5-minute sample in SKaTER-seq to determine usage. If any TSS has over 10% of the maximum coverage, the gene is marked as having multiple TSS and is not solved by the algorithm. Otherwise, the TSS with the maximum coverage is treated as the sole TSS and the others are discarded. Similar to TSS, CPSs within 150 nucleotides are grouped to the downstream most site, due to lack of resolution. However, CPSs with more than 150 nucleotide spacing are kept. TT-SKaTER is unable to model alternative terminal exons. Therefore, genes with an intronic 3' splice site downstream of a cleavage site are discarded from TT-SKaTER (but not SKaTER-seq). Additionally, when modeling coverage, we include a region downstream of the final PAS up to either the next downstream TSS or 5 kb, whichever is shorter.

We divide each gene into events rather than simultaneously modeling every intron in a gene. An event is defined by a group of introns with overlapping genomic intervals, whose processing cannot be treated as independent, together with all CPSs in the gene. For each group, we generate a state model containing the group's 5' and 3' splice sites, the junctions connecting them, and all CPSs. Introns that conflict with a CPS are included in every group. Constitutive intervals are modeled using this shared, cleavage-associated set of introns and CPSs. Thus, each interval is fit with the smallest state model capable of predicting its coverage.

Junctional observations are assigned to events according to the mapped 3' end of the read pair. The event for a junction contains the measured intron, introns that conflict with the measured intron and any introns intervening between the 3' splice site. In SKaTER-seq events with introns intervening between the 3' splice site and the final nucleotide of the fragment were excluded. In TT-SKaTER, these events were included due to the simpler solutions to the state model.

State Model Solutions

For each event, we construct a directed state graph of possible lifecycles of an individual mRNA molecule. Each state records the processing status of every splice junction and CPS in the event. A splice junction becomes eligible when Pol II transcribes its 3' splice site, and a CPS becomes eligible after transcription of the CPS. States can additionally transition to a new state through transcription of the next critical point or by processing an eligible splicing or cleavage site. Predicted coverage at a genomic position is the sum of occupancies of all nascent states that retain that position, together with the compatible mature-RNA states. Predicted junctional splicing fractions are calculated by summing the occupancies of the states that report the junction as spliced or unspliced, respectively. If an intron or set of introns intervene between the 3' splice site, and the distance between the 3' splice site and final fragment position is greater than 200 nucleotides, we assume that all intervening introns must be spliced or otherwise ineligible for splicing.

The number of mRNA molecules N in a state m can be described by the following formula.

$$(2) \quad N_m = \begin{cases} \sum_{y \in \mathbb{Y}} P_y F_{v_y}(t, \tau_y) & \Lambda_m = 0 \\ \sum_{y \in \mathbb{Y}} P_y f_{v_y}(t, \tau_y) & \Lambda_m > 0 \end{cases}$$

Where $\mathbb{Y}$ is the set of all paths v_y that end in the state m , τ_y is the time when the critical point to enter (or exit) the first state in v_y is transcribed, P_y is the probability of a given mRNA being in the first state in path v_y at time τ_y . $f_{v_y}(t, \tau_y)$ and $F_{v_y}(t, \tau_y)$ are defined as:

$$(3) \quad f_{[1, \dots, m]}(t, \tau) = \left(\prod_{i=1}^{m-1} (\lambda_i) \right) \left(\frac{G \lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=0}^m \left(\frac{C_{[0, \dots, i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i(t-\tau)} \right) \right)$$

$$(4) \quad F_{\emptyset}(t, \tau) = G\lambda_n(t - \tau) - \frac{C_{[0]}}{\Lambda_0} (1 - e^{-\Lambda_0(t-\tau)})$$

$$(5) \quad F_{[1, \dots, m]}(t, \tau) = \left(\prod_{i=1}^m (\lambda_i) \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} (t - \tau) - \sum_{i=0}^m \left(\frac{C_{[0, \dots, i]}}{\Lambda_i \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} (1 - e^{-\Lambda_i(t-\tau)}) \right) \right)$$

Where Λ_0 is the DRB release rate, G is the number of transcription sites, λ_n is the initiation rate at unblocked transcription sites, and Λ_i is the sum of all splicing rates leaving state i . λ_{v_i} defined locally as the splicing/cleavage rate for moving between states v_i and v_{i+1} or, in other words, the rate of splicing used to leave state i in the particular path, and C is defined as follows:

$$(6) \quad C_{[0, \dots, m]} = \frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=0}^{m-1} \left(\frac{C_{[0, \dots, i]}}{\prod_{j=i}^m ((\Lambda_j - \Lambda_i))} \right)$$

.

The probability P_m of entering state m by elongation after time δ can be defined as the sum of all paths $x \in \mathbb{X}$ able to reach state m multiplied by the probability of initially entering the state of origin x_0 by elongation P'_{x_0}

$$(7) \quad P_n = \sum_{x \in \mathbb{X}} (P'_{x_0} * p_x)$$

where p_x is defines as follows:

$$(8) \quad p_{[a]} = e^{-\Lambda_a \delta}$$

$$(9) \quad p_{[a,b]} = \frac{\lambda_a}{\Lambda_a - \Lambda_b} (p_{[b]} - p_{[a]})$$

$$(10) \quad p_{[a,b,\dots,n]} = \frac{\lambda_a}{\Lambda_a - \Lambda_b} (p_{[b,\dots,n]} - q_{a:[b,\dots,n]})$$

$$(11) \quad q_{a:[b,c]} = \frac{\lambda_b}{\Lambda_a - \Lambda_c} (p_{[c]} - p_{[a]})$$

$$(12) \quad q_{a:[b,c,\dots,n]} = \frac{\lambda_b}{\Lambda_a - \Lambda_c} (p_{[c,\dots,n]} - q_{a:[c,\dots,n]})$$

In steady state, this set of formulas simplifies. In particular, we only are concerned with the set of possible paths $y' \in \mathbb{Y}'$ containing entrances into states that can reach state m . The formula then simplifies to formula simplifies as follows.

$$(13) \quad N_m = \begin{cases} \sum_{y' \in \mathbb{Y}'} P_{y'} F'_{v_{y'}}(\delta_{y'}) & \Lambda_m = 0 \\ \sum_{y' \in \mathbb{Y}'} P_{y'} f'_{v_{y'}}(\delta_{y'}) & \Lambda_m > 0 \end{cases}$$

Where $f'_{v_{y'}}(\delta_{y'})$ and $F'_{v_{y'}}(\delta_{y'})$ are defined as:

$$(14) \quad f'_{[1,\dots,m]}(\delta) = \left(\prod_{i=1}^{m-1} \lambda_i \right) \left(\frac{G \lambda_n}{\prod_{i=1}^m \Lambda_i} - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} e^{-\Lambda_i \delta} \right)$$

$$(15) \quad f'_{[1,\dots,m]}(\delta \rightarrow \infty) = \left(\prod_{i=1}^{m-1} \lambda_i \right) \frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i}$$

$$(16) \quad F'_{[1,\dots,m]}(\delta) = \left(\prod_{i=1}^m \lambda_i \right) \left(\frac{G\lambda_n}{\prod_{i=1}^m \Lambda_i} \delta - \sum_{i=1}^m \frac{C'_{[1,\dots,i]}}{\Lambda_i \prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} (1 - e^{-\Lambda_i \delta}) \right)$$

Where Λ_0 is the DRB release rate, G is the number of transcription sites, λ_n is the initiation rate at unblocked transcription sites, and Λ_i is the sum of all splicing rates leaving state i . λ_{v_i} defined locally as the splicing/cleavage rate for moving between states v_i and v_{i+1} or, in other words, the rate of splicing used to leave state i in the particular path, and $C_{v_{y'}}$ is defined as follows:

$$(17) \quad C'_{[1,\dots,m]} = \frac{G\lambda_n}{\prod_{i=1}^m (\Lambda_i)} - \sum_{i=1}^{m-1} \left(\frac{C'_{[1,\dots,i]}}{\prod_{j=i+1}^m (\Lambda_j - \Lambda_i)} \right)$$

These formulas are all derived in the supplemental text.

For DRB released time points, SKaTER-seq integrates from τ to t , and it integrates from τ to ∞ in steady state. TT-SKaTER meanwhile uses the steady state solutions but integrates over an interval from τ to $\tau + l$ where l is the labeling time used in TT-seq. We developed an automated pipeline to handle the spool-out of these formulas

Time to position calculation

Both approaches require the time $\tau(n)$ at which Pol II reaches nucleotide n , but estimate it differently. In SKaTER-seq, elongation rates are estimated for each interval between consecutive

DRB-release samples. We calculate $\tau(n)$ by accumulating distance divided by the fitted rate across these time intervals.

In TT-SKaTER, we use PRO-seq density as a local measure of polymerase dwell time. We assume that the total number of PRO-seq reads separating two positions in a given gene is proportional to the amount of time it takes to transcribe that region multiplied by a gene specific polymerase dwell time $\lambda_{p,j}$ defined as the time per PRO-seq read.

$$(18) \quad \lambda_{p,j} = \begin{cases} \frac{\text{CPS}_0}{\bar{v} \sum_{i=0}^{\text{CPS}_0-1} R(i)} & j=0 \\ \frac{\lambda_{p,j-1}}{1 - \Lambda_{C,j-1} \lambda_{p,j-1} \sum_{i=\text{CPS}_{j-1}}^{\text{CPS}_j-1} R(i)} & j>0 \end{cases}$$

where $R(i)$ is the number of PRO-seq reads at position i , CPS_0 is the position of the first CPS, and $\bar{v}$ is the gene average elongation rate, $\Lambda_{C,j} = \sum_{a=0}^j \lambda_{c,a}$ is the cumulative cleavage rate for cleavage sites up to and including CPS_j . Before the first CPS, the amount of time taken to reach a particular position $\tau(n)$ can be defined as:

$$(19) \quad \tau(n) = \sum_{i \leq n} R(i) \lambda_{p,0}$$

Past the CPS, this coverage will be an underestimate due to the cleavage and termination of some portion of polymerases. Assuming that termination is fast relative to cleavage, for $\text{CPS}_j \leq i < \text{CPS}_{j+1}$ we calculate the elapsed time $\Delta t(i)$ from CPS_j to nucleotide i as

$$(20) \quad \Delta t(i) = -\frac{1}{\Lambda_{C,j}} \ln \left(1 - \Lambda_{C,j} \lambda_{p,j} \sum_{k=\text{CPS}_j}^i R(k) \right)$$

These formulas have an additional constraint of

$$(21) \quad \Lambda_{C,j} \lambda_{p,j} \sum_{k=\text{CPS}_j}^{\text{CPS}_{j+1}-1} R(k) < 1$$

These formulas are derived in the supplemental text.

Spike-in Normalization

Depending on the time since DRB wash out a sample was collected, we expect SKaTER-seq samples to have different levels of nascent RNA. These expected global changes make normalization to total human read counts an inappropriate strategy. Our cross-species spike-in of the same number of untreated (steady state) *Drosophila* S2 cells to every sample allows us to normalize across time points. Reads that map to the *Drosophila* genome are used to calculate a scaling factor used to normalize the human data. We define this scaling factor β as:

$$(22) \quad \beta_{\text{time point}} = \frac{\text{Total drosophila reads}_{\text{time point}}}{\text{Total drosophila reads}_{\text{steady state}}}$$

mRNA contamination

While the poly(A) depletion described above removes the bulk of mRNA contamination, residual mRNA in the nascent RNA purification would affect the estimations of splicing and cleavage kinetics if not taken into consideration computationally. Assuming mRNA contamination from *Drosophila* RNA is proportional to the mRNA contamination from human samples to which it was spiked in, we can use the spiked-in *Drosophila* RNA to obtain the following contamination factors.

$$(23) \quad c_t = \frac{\gamma_t - \alpha}{\max(\gamma_t) - \alpha} c$$

where α is 0.31 (the minimum *Drosophila* exon intron ratio), γ_t is the observed exon/intron ratio in sequenced *Drosophila* RNAs from time t , c_t is the time point specific contamination metric, and c is an optimized gene specific contamination level. Values for γ_t were calculated using CollectRnaSeqMetrics from PicardTools (*Picard Toolkit*, 2019). Contamination is added to predicted mRNA and junctional coverage as fully mature mRNAs corresponding to predicted ratios of alternative splicing outcomes once fully spliced.

Differential evolution optimization algorithm

A differential evolution optimization algorithm (Storn & Price, 1996) was used along with the above model of nascent RNA transcription and co-transcriptional processing to determine the set of rates that best fit the SKaTER-seq or TT-SKaTER data. We start with a population of 100 potential solutions each containing numeric values for each parameter (rate) necessary to predict coverage of an individual gene. Rates making up each solution in the first population are chosen randomly between the set upper and lower bounds for a rate from a uniform distribution on the log scale. For each solution, we model the coverage based on the rates and compute the negative log likelihood (NLL) of the coverage and junctional scores.

Coverage is scored using the Poisson distribution where c_i is the observed coverage at position i and μ_i its model-predicted mean.

$$(24) \quad \text{NLL}_{\text{coverage}} = \sum_i (\mu_i - c_i \ln(\mu_i) + \ln(c_i!))$$

Junctional coverage is scored using a binomial distribution where s_i is the observed splicing status and q_i its predicted probability of observing a spliced read.

$$(25) \quad \text{NLL}_{\text{junction}} = - \sum_i \left(\ln \left(\binom{1}{s_i} \right) + s_i \ln(q_i) + (1 - s_i) \ln(1 - q_i) \right)$$

The differential evolution was performed with a mutation factor selected uniformly between 0.5 and 1.0 for each generation and a fixed crossover rate of 0.8. Optimization was run for a maximum of 50,000 generations for both SKaTER-seq and TT-SKaTER, with minimum generation counts of 1,000 and 100, respectively. If the difference between the highest and lowest negative log likelihood dropped below 10^{-4} after the minimum number of generations, then the population was considered converged, and the optimization stopped.

Rate filtering

The differential evolution algorithm was run a minimum of 3 and a maximum of 9 times for each gene. These independent runs use different random initial populations. Runs were ranked by their final negative log likelihood, and additional runs were launched until three solutions were within 0.1% of the best score or the maximum of nine runs was reached. Genes that failed to converge to a shared best solution were excluded. Individual parameter estimates were flagged for filtering when they were not reproduced within 1% (on the log scale) between the two best runs, did not converge within 1% across the final optimizer population, or reached a search-boundary value. Rates whose required model parameters did not meet these criteria were not used in downstream analysis.

Stochastic modeler

Pipeline

We constructed a pipeline to create stochastically model TT-seq coverage for a particular gene with specified critical points, splicing connections, and kinetic parameters. For each simulated mRNA molecule, we randomly sampled a final transcription position, a delay representing time from transcribing the final position until sample collection (between zero and labeling time), and a fragment length for paired end reads (based on the distribution of constitutively intronic coverage from a TT-seq dataset). Splice-site and cleavage times were then sampled from exponential distributions beginning when the corresponding 3' splice site or CPS was transcribed. Molecules with a final transcription position in a spliced intron or past a cleaved CPS, were discarded. The remaining molecules were converted to paired-end reads using the empirical fragment-length distribution, from which final-nucleotide coverage and junctional splicing status were calculated using the same rules as the TT-seq loading pipeline. PRO-seq positions could be similarly randomly drawn from random positions within the gene, with coverage removed past the CPS to model termination.

State model testing

All TT-SKaTER simulations used a 10-minute labeling period. Except where PRO-seq depth was varied, PRO-seq coverage was set to a uniform profile. Simulations used independent random seeds for read generation and optimization. Accuracy heatmaps and the colors in Figure 2F classify an estimate as accurate when it is within 10% of its input value.

For Figure 2F, we simulated a single intron with its 5' and 3' splice sites at positions 100 and 300–10100, respectively, in 200-nt increments. The CPS was 1000 nt downstream of the 3' splice site and 5000 nt of sequence past the CPS was included. Input splicing and cleavage rates were 0.2 and 0.3 min⁻¹, respectively. Elongation rates were 1, 2, or 4 kb/min. Figure 2F displays the subset with TT-seq depth below 20 fragments/nt, with each position, rate, and depth combination simulated 30 times.

Figure 2G and Supplemental Figure 2A use the same single-intron dataset, with Figure 2G showing elongation-rate accuracy and Supplemental Figure 2A showing splicing-rate accuracy. The 5' was located at 100 bp from the TSS with the intron length varying from 200 to 10KB with 200 nt increments. The CPS was 1000 nt downstream of the 3' splice site, and 2000 nt of sequence downstream of the CPS was modeled. Input splicing and cleavage rates were 0.2 and 0.3 min⁻¹, respectively, and input elongation rates were 1, 2, or 4 kb/min. Each position, rate, and TT-seq depth combination was simulated 30 times.

Supplemental Figure 2B uses a fixed 2000 nt long intron with splice sites at 100 and 2100 nucleotides, a single CPS at position 3100, and the modeled end at position 8100. Splicing and cleavage rates were 0.2 and 0.3 min⁻¹, elongation rates were 1, 2, or 4 kb/min, and both TT-seq and PRO-seq depths were varied independently. The same 2000 nt long intron architecture and rates was used for Supplemental Figure 2C-F, while the modeled end ranged from 3300 to 13100 in 200-nt increments. These datasets used 30 replicates per condition.

For Figure 2H, the constitutive single-intron gene had splice sites at positions 100 and 2100, a CPS at 3100, and an endpoint at 8100. Alternative 5' splice-site genes used 5' splice sites at 100 and 200 with a shared 3' splice site at 2200. Alternative 3' splice-site genes used a shared

5' splice site at 100 and 3' splice sites at 2100 and 2200. Cassette-exon genes used positions 100, 1100, 1200, and 2200. All alternative-event genes had a CPS at 3200 and an endpoint at 8200. Each splice rate was varied independently over 0.03, 0.2, and 0.9 min⁻¹, while cleavage and elongation were fixed at 0.3 min⁻¹ and 2 kb/min. These simulations were performed 100 times per architecture, kinetic-parameter combination, and depth.

Across all panels, TT-seq depth was evaluated at 500, 200, 100, 50, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001, 0.0005, 0.0002, and 0.0001 fragments/nt.

QAPA analysis

3' UTR libraries were constructed based on the hg38 genome using QAPA (Ha et al., 2018) with PAS locations from the PolyASite database (Herrmann et al., 2019). Transcript quantification was performed on the raw fastq files using Salmon (Patro et al., 2017). Counts for each individual PAS were summed across replicates. Fractional UTR lengths for each gene with multiple PAS were calculated by normalizing the UTR length between zero for the shortest and one for the longest then calculating the count weighted average UTR length. Fractional changes in UTR length were calculated by subtracting the normalized UTR length for the perturbation from the control. A Cochran-Armitage trend test was used to test for systematic shifts in PAS usage between conditions across PAS isoforms ordered by 3' UTR length.

Stranded BigWigs of sequencing data for SRSF9 knockdown and controls were downloaded from encode. Replicates were averaged using BigWigCompare, and log2 fold changes were computed using BigWigCompare with a pseudo-count of 0.1 (Ramírez et al., 2016). Final sample tracks were created using pyGenomeTracks (Lopez-Delisle et al., 2021; Ramírez et al., 2018).

Permutation Testing

Neighboring intron

Reliably fit constitutive introns were collected from each gene and sorted based on the distance of their 5' splice site from the TSS. Each intron was paired with its nearest downstream neighbor if possible, with alternative or not reliably fit introns ignored. A Spearman correlation was then calculated between the two sets. To test for the effect of local intron order on rates, we randomly permuted the intron rates independently within every gene and recalculated the pooled correlation. We repeated this for 10000 permutations. We then compared the observed and permuted correlations to calculate a significance.

Intron properties

For GC content, intron length, and 5' and 3' splice-site strength, we calculated a Spearman correlation between the property and splicing rate separately within each gene, excluding genes with fewer than two variable introns. We summarized the set of within-gene correlations with a one-sample T-statistic with a population mean of zero, corresponding to an uncorrelated Spearman correlation. We then performed a permutation test with 10000 permutations by randomly shuffling the order of the property values (GC content, intron length, 5' and 3' splice site strengths) within each gene, recalculating the Spearman correlations and population level T-statistic. We then compared the observed and permuted T-statistics to calculate a significance.

TAD effects

Genes were assigned to TADs only when the genes were fully contained within a TAD. For gene average splicing analysis, rates were $\log_{10}$ -transformed and gene-average rates were

used. For elongation rates, gene average elongation rates were used, and only genes with a single cleavage site were considered when testing cleavage rates. For each TAD, we compared its rate distribution within each individual TAD with the pooled distribution from all other TADs using the Kolmogorov-Smirnov statistic and summed these statistics across TADs. We generated a null distribution from 10000 random reallocations of the rate values among TADs, retaining the number of genes assigned per TAD. We then compared the observed summed statistic with this distribution.

Random Forest Regression

Normalization

We downloaded ENCODE BigWig signal tracks for histone marks, transcription factors, and ATAC-seq in IMR90. We used ENCODE fold-change-over-control tracks, so fold change was provided by the ENCODE processing pipeline rather than recomputed from raw reads. For every analysis window, we extracted the exact mean BigWig value across the genomic interval for each track.

For TT-SKaTER, we used genes with reliable gene-average elongation-rate estimates and PRO-seq coverage. We excluded the first 500 nt after the TSS to avoid effects from the promoter and promoter proximal pause, and the rest of the gene was partitioned into consecutive windows. A window was closed at the first position at which it was at least 50 nt long, represented at least 5 minutes of predicted transcription time, and contained at least one PRO-seq read. Its elongation rate was calculated as window length divided by the predicted transit time, using the PRO-seq derived time-to-position relationship. For SKaTER-seq, we used the fitted elongation-rate windows

after excluding the first 500 nt at the TSS. We extracted coverage from the same genomic windows for all features.

Chromatin-mark signal was normalized separately for each mark to account for window length. For every observed window size, we sampled 1,000 random windows of the same size from the genomic spans represented by the analyzed genes and calculated the mean and standard deviation of each track. We transformed the signal to the log scale using $\log_2(x + p)$, where p was the first percentile of positive values for that track in the random-window distribution. Each observed signal was converted to a z-score using the corresponding null distribution. Windows larger than the sampled range used the largest available null distribution.

Analysis

We fit random forest regressors using the Scikit-Learn module to $\log_{10}$ -transformed elongation rates using z-scores of chromatin-mark intensities as predictors. The Random Forest Regression model used 200 trees, a minimum split size of 10, and a minimum leaf size of 5 with 10% of the data set aside for test. Testing data was used to predict elongation rates and compared to the observed rates to calculate an R^2 . Scikit-learn permutation importance function was run with 100 independent permutations of each feature. Two-feature partial-dependence plots were calculated by averaging model predictions over the training data while varying the selected pair of mark intensities across its observed grid. Contour plots of observed fast- and slow-elongation windows were overlaid for comparison.

Bibliography

- Baluapuri, A., Zhao, N. C., Marina, R. J., Huang, K.-L., Kuzkina, A., Amodeo, M. E., Stein, C. B., Ahn, L. Y., Farr, J. S., Schaffer, A. E., Khurana, V., Wagner, E. J., & Adelman, K. (2025). Integrator loss leads to dsRNA formation that triggers the integrated stress response. *Cell*, 188(12), 3184–3201.e21. <https://doi.org/10.1016/j.cell.2025.03.025>
- Bedi, K., Magnuson, B., Narayanan, I. V., Paulsen, M. T., Wilson, T. E., & Ljungman, M. (2021). Cotranscriptional splicing efficiencies differ within genes and between cell types. *RNA*, 27(7), 829–840. <https://doi.org/10.1261/rna.078662.120>
- Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*, 15(3), 163–175. <https://doi.org/10.1038/nrg3662>
- Berget, S. M. (1995). Exon Recognition in Vertebrate Splicing. *Journal of Biological Chemistry*, 270(6), 2411–2414. <https://doi.org/10.1074/jbc.270.6.2411>
- Cáceres, J. F., & Kornblihtt, A. R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends in Genetics*, 18(4), 186–193. [https://doi.org/10.1016/S0168-9525\(01\)02626-9](https://doi.org/10.1016/S0168-9525(01)02626-9)
- Caizzi, L., Monteiro-Martins, S., Schwalb, B., Lysakovskaia, K., Schmitzova, J., Sawicka, A., Chen, Y., Lidschreiber, M., & Cramer, P. (2021). Efficient RNA polymerase II pause release requires U2 snRNP function. *Molecular Cell*, 81(9), 1920–1934.e9. <https://doi.org/10.1016/j.molcel.2021.02.016>
- Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., & Neugebauer, K. M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell*, 165(2), 372–381. <https://doi.org/10.1016/j.cell.2016.02.045>
- Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E. K., Adam, S. A., Goldman, R., Van Steensel, B., Ma, J., & Belmont, A. S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *Journal of Cell Biology*, 217(11), 4025–4048. <https://doi.org/10.1083/jcb.201807108>
- Cingaram, P. K. R., Beckedorff, F., Yue, J., Liu, F., Gomes Dos Santos, H., Arigela, H., & Shiekhhattar, R. (2025). Enhancing transcriptome mapping with rapid PRO-seq profiling of nascent RNA. *Molecular Cell*, 85(15), 2988–3001.e4. <https://doi.org/10.1016/j.molcel.2025.06.029>
- Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., & Bentley, D. L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation

- Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. *Molecular Cell*, 76(6), 896–908.e4. <https://doi.org/10.1016/j.molcel.2019.09.031>
- Danko, C. G., Hah, N., Luo, X., Martins, A. L., Core, L., Lis, J. T., Siepel, A., & Kraus, W. L. (2013). Signaling Pathways Differentially Affect RNA Polymerase II Initiation, Pausing, and Elongation Rate in Cells. *Molecular Cell*, 50(2), 212–222. <https://doi.org/10.1016/j.molcel.2013.02.015>
- Darzacq, X., Shav-Tal, Y., De Turris, V., Brody, Y., Shenoy, S. M., Phair, R. D., & Singer, R. H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nature Structural & Molecular Biology*, 14(9), 796–806. <https://doi.org/10.1038/nsmb1280>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Duan, S., Nodelman, I. M., Zhou, H., Tsukiyama, T., Bowman, G. D., & Zhang, Z. (2025). H3K56 acetylation regulates chromatin maturation following DNA replication. *Nature Communications*, 16(1), 134. <https://doi.org/10.1038/s41467-024-55144-7>
- Dye, M. J., & Proudfoot, N. J. (1999). Terminal Exon Definition Occurs Cotranscriptionally and Promotes Termination of RNA Polymerase II. *Molecular Cell*, 3(3), 371–378. [https://doi.org/10.1016/S1097-2765\(00\)80464-5](https://doi.org/10.1016/S1097-2765(00)80464-5)
- The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. <https://doi.org/10.1038/nature11247>
- Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I., & Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biology*, 15(5), R69. <https://doi.org/10.1186/gb-2014-15-5-r69>
- Gómez Acuña, L. I., Flyamer, I., Boyle, S., Friman, E. T., & Bickmore, W. A. (2024). Transcription decouples estrogen-dependent changes in enhancer-promoter contact frequencies and spatial proximity. *PLOS Genetics*, 20(5), e1011277. <https://doi.org/10.1371/journal.pgen.1011277>
- Gressel, S., Schwalb, B., Decker, T. M., Qin, W., Leonhardt, H., Eick, D., & Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife*, 6, e29736. <https://doi.org/10.7554/eLife.29736>
- Ha, K. C. H., Blencowe, B. J., & Morris, Q. (2018). QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. *Genome Biology*, 19(1), 45. <https://doi.org/10.1186/s13059-018-1414-4>
- Herrmann, C. J., Schmidt, R., Kanitz, A., Artimo, P., Gruber, A. J., & Zavolan, M. (2019). PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3' end sequencing. *Nucleic Acids Research*, gkz918. <https://doi.org/10.1093/nar/gkz918>

- Hu, S., Peng, L., Xu, C., Wang, Z., Song, A., & Chen, F. X. (2021). SPT5 stabilizes RNA polymerase II, orchestrates transcription cycles, and maintains the enhancer landscape. *Molecular Cell*, 81(21), 4425–4439.e6. <https://doi.org/10.1016/j.molcel.2021.08.029>
- Hwang, J. Y., Jung, S., Kook, T. L., Rouchka, E. C., Bok, J., & Park, J. W. (2020). rMAPS2: an update of the RNA map analysis and plotting server for alternative splicing regulation. *Nucleic Acids Research*, 48(W1), W300–W306. <https://doi.org/10.1093/nar/gkaa237>
- Jonkers, I., Kwak, H., & Lis, J. T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife*, 3, e02407. <https://doi.org/10.7554/eLife.02407>
- Kagda, M. S., Lam, B., Litton, C., Small, C., Sloan, C. A., Spragins, E., Tanaka, F., Whaling, I., Gabdank, I., Youngworth, I., Strattan, J. S., Hilton, J., Jou, J., Au, J., Lee, J.-W., Andreeva, K., Graham, K., Lin, K., Simison, M., ... Hitz, B. C. (2025). Data navigation on the ENCODE portal. *Nature Communications*, 16(1), 9592. <https://doi.org/10.1038/s41467-025-64343-9>
- Khodor, Y. L., Menet, J. S., Tolan, M., & Rosbash, M. (2012). Cotranscriptional splicing efficiency differs dramatically between *Drosophila* and mouse. *RNA*, 18(12), 2174–2186. <https://doi.org/10.1261/rna.034090.112>
- Kim, H. B., Nandu, T., Kim, Y. J., Camacho, C. V., & Kraus, W. L. (2026). Dose-dependent hormone actions at estrogen receptor enhancers specify distinct molecular and biological outcomes. *Science Advances*, 12(23), eaec2424. <https://doi.org/10.1126/sciadv.aec2424>
- Kim, J., Lee, J., & Lee, T.-H. (2015). Lysine Acetylation Facilitates Spontaneous DNA Dynamics in the Nucleosome. *The Journal of Physical Chemistry B*, 119(48), 15001–15005. <https://doi.org/10.1021/acs.jpcb.5b09734>
- Kwak, H., Fuda, N. J., Core, L. J., & Lis, J. T. (2013). Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science*, 339(6122), 950–953. <https://doi.org/10.1126/science.1229386>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, Z., Burgos-Bravo, F., Xu, K., Li, C., Kwan, K. Y., Tong, A. B., Shan, Z., Wang, H., Takaku, M., Li, J., Shi, Z., Lyumkis, D., Bustamante, C., & Fei, J. (2025). Phase-separated NDF–FACT condensates facilitate transcription elongation on chromatin. *Nature Cell Biology*, 27(11), 1938–1951. <https://doi.org/10.1038/s41556-025-01778-8>
- Liu, L., Zhao, Y., Hassett, R., Toneyan, S., Koo, P. K., & Siepel, A. (2025). Probabilistic and machine-learning methods for predicting local rates of transcription elongation from nascent

- RNA sequencing data. *Nucleic Acids Research*, 53(4), gkaf092. <https://doi.org/10.1093/nar/gkaf092>
- Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F., & Manke, T. (2021). pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics*, 37(3), 422–423. <https://doi.org/10.1093/bioinformatics/btaa692>
- Luo, C., He, J., Yang, Y., Wu, K., Fu, X., Cheng, J., Ming, Y., Liu, W., & Peng, Y. (2024). SRSF9 promotes cell proliferation and migration of glioblastoma through enhancing CDK1 expression. *Journal of Cancer Research and Clinical Oncology*, 150(6), 292. <https://doi.org/10.1007/s00432-024-05797-0>
- Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., Baymuradov, U. K., Graham, K., Litton, C., Miyasato, S. R., Strattan, J. S., Jolanki, O., Lee, J.-W., Tanaka, F. Y., Adenekan, P., ... Cherry, J. M. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Research*, 48(D1), D882–D889. <https://doi.org/10.1093/nar/gkz1062>
- Malszycki, M., Martina, L., Ilık, İ. A., Salgado Figueroa, D., Dasgupta, N., Çoşar, M. F., Kim, K.-T., Carraco, G., Fauler, B., Meierhofer, D., Mielke, T., Imai, H., Alev, C., Ay, F., & Aktaş, T. (2026). Nuclear speckles enable processing of RNA from GC-rich isochores. *Cell*, S009286742600098X. <https://doi.org/10.1016/j.cell.2026.01.011>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet.journal*, 17(1), 10. <https://doi.org/10.14806/ej.17.1.200>
- Mayer, A., Iulio, J. di, Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J. A., & Churchman, L. S. (2015). Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. *Cell*, 161(3), 541–554. <https://doi.org/10.1016/j.cell.2015.03.010>
- Mimoso, C. A., & Adelman, K. (2023). U1 snRNP increases RNA Pol II elongation rate to enable synthesis of long genes. *Molecular Cell*, 83(8), 1264–1279.e10. <https://doi.org/10.1016/j.molcel.2023.03.002>
- Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., Carmo-Fonseca, M., & Proudfoot, N. J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*, 161(3), 526–540. <https://doi.org/10.1016/j.cell.2015.03.027>
- Pai, A. A., Henriques, T., McCue, K., Burkholder, A., Adelman, K., & Burge, C. B. (2017). The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. *Elife*, 6, e32537. <https://doi.org/10.7554/eLife.32537>

- Pal, S., Graves, H., Ohsawa, R., Huang, T.-hsiang, Wang, P., Harmacek, L., & Tyler, J. (2016). The Commercial Antibodies Widely Used to Measure H3 K56 Acetylation Are Non-Specific in Human and Drosophila Cells. *PLOS ONE*, 11(5), e0155409. <https://doi.org/10.1371/journal.pone.0155409>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417–419. <https://doi.org/10.1038/nmeth.4197>
- Patro, R., Mount, S. M., & Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature Biotechnology*, 32(5), 462–464. <https://doi.org/10.1038/nbt.2862>
- Picard toolkit*. (2019,). Broad Institute.
- Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., Habermann, B., Akhtar, A., & Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications*, 9(1), 189. <https://doi.org/10.1038/s41467-017-02525-w>
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1), W160–W165. <https://doi.org/10.1093/nar/gkw257>
- Reimer, K. A., Mimoso, C. A., Adelman, K., & Neugebauer, K. M. (2021). Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Molecular Cell*, 81(5), 998–1012.e7. <https://doi.org/10.1016/j.molcel.2020.12.018>
- Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., & Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science*, 352(6290), 1225–1228. <https://doi.org/10.1126/science.aad9841>
- Schwich, O. D., Blümel, N., Keller, M., Wegener, M., Setty, S. T., Brunstein, M. E., Poser, I., Mozos, I. R. D. L., Suess, B., Münch, C., McNicoll, F., Zarnack, K., & Müller-McNicoll, M. (2021). SRSF3 and SRSF7 modulate 3'UTR length through suppression or activation of proximal polyadenylation sites and regulation of CFIm levels. *Genome Biology*, 22(1), 82. <https://doi.org/10.1186/s13059-021-02298-y>
- Singh, J., & Padgett, R. A. (2009). Rates of in situ transcription and splicing in large human genes. *Nature Structural & Molecular Biology*, 16(11), 1128–1133. <https://doi.org/10.1038/nsmb.1666>
- Stein, C. B., Field, A. R., Mimoso, C. A., Zhao, C., Huang, K.-L., Wagner, E. J., & Adelman, K. (2022). Integrator endonuclease drives promoter-proximal termination at all RNA polymerase

- II-transcribed loci. *Molecular Cell*, 82(22), 4232–4245.e11. <https://doi.org/10.1016/j.molcel.2022.10.004>
- Storn, R., & Price, K. (1996). Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, 341–359. <https://doi.org/https://doi.org/10.1023/A:1008202821328>
- Tian, B., & Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology*, 18(1), 18–30. <https://doi.org/10.1038/nrm.2016.116>
- Torres-Ulloa, L., Calvo-Roitberg, E., & Pai, A. A. (2024). Genome-wide kinetic profiling of pre-mRNA 3' end cleavage. *RNA*, 30(3), 256–270. <https://doi.org/10.1261/rna.079783.123>
- Värv, S., Kristjuhan, K., Peil, K., Lööke, M., Mahlakõiv, T., Paapsi, K., & Kristjuhan, A. (2010). Acetylation of H3 K56 Is Required for RNA Polymerase II Transcript Elongation through Heterochromatin in Yeast. *Molecular and Cellular Biology*, 30(6), 1467–1477. <https://doi.org/10.1128/MCB.01151-09>
- Veloso, A., Kirkconnell, K. S., Magnuson, B., Biewen, B., Paulsen, M. T., Wilson, T. E., & Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Research*, 24(6), 896–905. <https://doi.org/10.1101/gr.171405.113>
- Vihervaara, A., Mahat, D. B., Himanen, S. V., Blom, M. A., Lis, J. T., & Sistonen, L. (2021). Stress-induced transcriptional memory accelerates promoter-proximal pause release and decelerates termination over mitotic divisions. *Molecular Cell*, 81(8), 1715–1731.e6. <https://doi.org/10.1016/j.molcel.2021.03.007>
- Wachutka, L., Caizzi, L., Gagneur, J., & Cramer, P. (2019). Global donor and acceptor splicing site kinetics in human cells. *Elife*, 8, e45056. <https://doi.org/10.7554/eLife.45056>
- Wang, H., Fan, Z., Shliaha, P. V., Miele, M., Hendrickson, R. C., Jiang, X., & Helin, K. (2023). H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. *Nature*, 615(7951), 339–348. <https://doi.org/10.1038/s41586-023-05780-8>
- Wang, X., Lu, X., Wang, P., Chen, Q., Xiong, L., Tang, M., Hong, C., Lin, X., Shi, K., Liang, L., & Lin, J. (2022). SRSF9 promotes colorectal cancer progression via stabilizing DSN1 mRNA in an m6A-related manner. *Journal of Translational Medicine*, 20(1), 198. <https://doi.org/10.1186/s12967-022-03399-3>
- Watanabe, K., Yamamoto, T., Fujita, T., Hino, S., Hino, Y., Yamazaki, K., Ohashi, Y., Sakuraba, S., Kono, H., Nakao, M., Ochiai, K., Dan, S., & Saitoh, N. (2024). Metabolically inducing defects in DNA repair sensitizes *BRCA* –wild-type cancer cells to replication stress. *Science Signaling*, 17(862), eadl6445. <https://doi.org/10.1126/scisignal.adl6445>

- Wuarin, J., & Schibler, U. (1994). Physical Isolation of Nascent RNA Chains Transcribed by RNA Polymerase II: Evidence for Cotranscriptional Splicing. *Molecular and Cellular Biology*, 14(11), 7219–7225. <https://doi.org/10.1128/mcb.14.11.7219-7225.1994>
- Yeo, G., & Burge, C. B. (2004). Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology*, 11(2–3), 377–394. <https://doi.org/10.1089/1066527041410418>
- Zhang, L., Zhang, Y., Chen, Y., Gholamalamdari, O., Wang, Y., Ma, J., & Belmont, A. S. (2021). TSA-seq reveals a largely conserved genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. *Genome Research*, 31(2), 251–264. <https://doi.org/10.1101/gr.266239.120>
- Zhu, J., Bao, L., Xu, S., & Ji, X. (2025). FeaSion decodes the regulatory landscape and functional diversity of RNA polymerase II CTD phosphorylation. *Science Advances*, 11(48), eadz2345. <https://doi.org/10.1126/sciadv.adz2345>
- Žumer, K., Maier, K. C., Farnung, L., Jaeger, M. G., Rus, P., Winter, G., & Cramer, P. (2021). Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. *Molecular Cell*, 81(15), 3096–3109.e8. <https://doi.org/10.1016/j.molcel.2021.05.028>
- Žumer, K., Ochmann, M., Aljahani, A., Zheenbekova, A., Devadas, A., Maier, K. C., Rus, P., Neef, U., Oudelaar, A. M., & Cramer, P. (2024). FACT maintains chromatin architecture and thereby stimulates RNA polymerase II pausing during transcription in vivo. *Molecular Cell*, 84(11), 2053–2069.e9. <https://doi.org/10.1016/j.molcel.2024.05.003>
- Zuo, X., Xu, J., Yang, D., Yang, C., Liu, X., Chen, N., Wang, H., Luo, X., Luo, Q., Wang, Y., Wei, Q., Yang, Y., Zhao, H., Xu, J., Wang, W., Liu, T., & Yi, P. (2025). SRSF9 Forms Phase-Separated Condensates to Promote Ovarian Cancer Progression by Inducing RNA Alternative Splicing That Is Inhibited by m6A Modification. *Cancer Research*, 85(20), 3894–3909. <https://doi.org/10.1158/0008-5472.CAN-24-4233>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement.pdf](#)