TRANSPARENT PROCESS

OPEN ACCESS

molecular systems biology

Check for updates

SLAM-Drop-seq reveals mRNA kinetic rates throughout the cell cycle

Haiyue Liu[1,†] (iD), Roberto Arsiè[1,†], Daniel Schwabe[2], Marcel Schilling[1,3] (iD), Igor Minia[1], Jonathan Alles[1], Anastasiya Boltengagen[1], Christine Kocks[1] (iD), Martin Falcke[2,4] (iD), Nir Friedman[5,6,7] (iD), Markus Landthaler[1,8,*] (iD) & Nikolaus Rajewsky[1,9,10,11,12,13,**] (iD)

Abstract

RNA abundance is tightly regulated in eukaryotic cells by modulating the kinetic rates of RNA production, processing, and degradation. To date, little is known about time-dependent kinetic rates during dynamic processes. Here, we present SLAM-Drop-seq, a method that combines RNA metabolic labeling and alkylation of modified nucleotides in methanol-fixed cells with droplet-based sequencing to detect newly synthesized and preexisting mRNAs in single cells. As a first application, we sequenced 7280 HEK293 cells and calculated gene-specific kinetic rates during the cell cycle using the novel package Eskrate. Of the 377 robust-cycling genes that we identified, only a minor fraction is regulated solely by either dynamic transcription or degradation (6 and 4%, respectively). By contrast, the vast majority (89%) exhibit dynamically regulated transcription and degradation rates during the cell cycle. Our study thus shows that temporally regulated mRNA degradation is fundamental for the correct expression of a majority of cycling genes. SLAM-Drop-seq, combined with Eskrate, is a powerful approach to understanding the underlying mRNA kinetics of single-cell gene expression dynamics in continuous biological processes.

Introduction
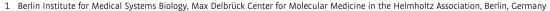
Progression through the cell cycle requires tight control of transcriptional (Dynlacht, 1997) and posttranscriptional (Blackinton & Keene, 2014) regulation to prevent deregulation that trigger cell death or lead to abnormal growth (Schafer, 1998). Many of the genes involved in coordinating this complex process are fine-tuned at specific stages during the cell cycle, and their dysregulation can lead to various diseases, most notably cancer (Hanahan & Weinberg, 2011; Rubin *et al*, 2020; Matthews *et al*, 2022). While the dynamics of gene expression have been extensively characterized in the cell cycle (Iyer *et al*, 1999; Cho *et al*, 2001; Whitfield *et al*, 2002; Liu *et al*, 2017), little is known how the underlying processes of mRNA transcription, splicing, and degradation influence gene expression and how their associated kinetic rates change depending on the cell cycle time.

Common approaches to studying RNA kinetic rates rely on 4-thiouridine (4sU) incorporation into nascent transcripts, followed by biochemical enrichment of labeled RNAs (Dölken *et al*, 2008; Miller *et al*, 2011; Rabani *et al*, 2011). To circumvent the main disadvantages associated with these approaches (e.g., high amount of starting material required, unspecific enrichment), SLAM-seq was developed (Herzog *et al*, 2017): instead of depending on enrichment steps, iodoacetamide (IAA) is used to alkylate the 4sU base in labeled RNAs. This alkylation results in diagnostic thymine-to-cytosine (T->C) transitions in a reverse-transcription dependent manner, which are detected and quantified with computational

1 Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany
2 Mathematical Cell Physiology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany
3 Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), University of Lübeck, Lübeck, Germany
4 Department of Physics, Humboldt University Berlin, Berlin, Germany
5 The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel
6 The Lautenberg Center for Immunology and Cancer Research, Institute of Medical Research Israel-Canada (IMRIC), Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel
7 The Center for Computational Medicine, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel
8 Institut für Biologie, Humboldt Universität zu Berlin, Berlin, Germany
9 Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
10 German Center for Cardiovascular Research (DZHK), Berlin, Germany
11 NeuroCure Cluster of Excellence, Berlin, Germany
12 German Cancer Consortium (DKTK), Berlin, Germany
13 National Center for Tumor Diseases (NCT), Berlin, Germany
*Corresponding author. Tel: +49 9406 3026; E-mail: markus.landthaler@mdc-berlin.de
**Corresponding author. Tel: +49 9406 2999; E-mail: rajewsky@mdc-berlin.de
†These authors contributed equally to this work

analysis after RNA sequencing. Although instructive, this and similar (Riml *et al*, 2017; Schofield *et al*, 2018) protocols assume RNA kinetic rates to be constant and average gene expressions for millions of cells, flattening the heterogeneity of single cells. In recent years, scRNA-seq has become the method of choice to analyze the complexity and heterogeneity of cell populations (Hashimshony *et al*, 2012; Macosko *et al*, 2015; Alles *et al*, 2017; Pepe-Mooney *et al*, 2019). Based on scRNA-data, gene-specific RNA synthesis and degradation rates in single cells have been calculated using precursor and mature mRNA counts (Manno *et al*, 2018), but the mRNA decay rate was assumed constant across the cell population. More recently, experimentally more complex single-cell methods have been established to investigate gene-specific mRNA kinetic rates based on SLAM-seq (scSLAM-seq (Erhard *et al*, 2019), NASC-seq (Hendriks *et al*, 2019), sci-fate (Cao *et al*, 2020), scNT-seq (Qiu *et al*, 2020)). These approaches were described in more detail in a recent review by Erhard *et al* (2022). While these methods all profile whole and newly synthesized transcriptomes, there are important differences: SMART-seq-based methods (i.e., scSLAM-seq, NASC-seq) normally capture deep coverage of transcripts in each cell, but assay only hundreds of cells due to labor and cost. Sci-fate relies on single-cell combinatorial indexing to obtain a high number of single cells. When many cells in heterogeneous populations need to be analyzed, droplet-based approaches such as scNT-seq are more advantageous, recovering thousands of cells per experiment with low cost and easy setup. Although all these approaches have been applied to investigate gene-specific mRNA kinetic rates in different systems, they only address discrete states of cell activation, differentiation, or infection. For biological processes that are continuous and dynamic, it is necessary to investigate the kinetic rates in a continuous time-resolved manner. For example, in the case of the cell cycle progression, it would be desirable to measure the RNA kinetic rates throughout the cell cycle as a function of time.

To overcome the limitations of current methods for analyzing time-dependent RNA kinetic rates, we developed SLAM-Drop-seq, an approach that employs 4sU metabolic RNA labeling, chemical conversion of labeled transcripts with IAA *in situ* inside fixed cells, followed by droplet-based scRNA-seq. The fixation step permeabilizes cell membranes and therefore allows the IAA-induced alkylation of 4sU residues to take place within intact cells. Compared to scNT-seq, which also applies droplet-based sequencing (Qiu *et al*, 2020), this is a novel step that reduces the hands-on time and complexity of the experiment. Using SLAM-Drop-seq, we quantified newly synthesized and preexisting mRNA transcripts in individual cells. Based on a time-dependent mRNA kinetic rate model, we developed the R package Eskrate to estimate the time-resolved mRNA kinetic rates. As a first application, we explored the dynamics of RNA kinetic rates along the cell cycle from a population of unsynchronized HEK293 cells. We assigned each cell to a specific time point of the cell cycle using Revelio (Schwabe *et al*, 2020), a recently developed approach to reconstructing the temporal continuity of the cell cycle from single-cell transcriptomic data of unperturbed cells. We profiled the time-dependent gene expression transcriptome-wide and identified 377 robustly cycling genes. The underlying rates of mRNA synthesis and degradation along the cell cycle were then calculated using Eskrate. We found that temporally regulated RNA degradation is considerably involved in cycling gene expression regulation in the cell cycle, and the gene expression of the majority of the analyzed cycling genes is the result of a close interplay between RNA production and decay.

Results

Experimental procedure and data

SLAM-Drop-seq combines 4-thiouridine (4sU) metabolic labeling and iodoacetamide (IAA) alkylation inside fixed cells with droplet-based scRNA-seq. We optimized the conditions to alkylate 4sU-labeled RNAs in methanol-fixed cell suspensions to preserve high-integrity RNAs (Appendix Figs S1A and B). For our SLAM-Drop-seq experiment, we incubated HEK293 cells with 4sU for different durations (0, 15, 30, 60 min of incubation, two biological replicates each). We alkylated fixed cells and sequenced their mRNA transcriptomes via a Drop-seq protocol. The accumulation of T->C transitions was detected in sequence reads marked with cellular and molecular barcodes by computational analysis to quantify newly transcribed and preexisting transcripts of each gene in each single cell (Fig 1A).

We obtained a total of 7,280 single cells that expressed at least 200 genes (Appendix Fig S2A). Although the sequencing depths were different between sequencing batches (batch 1 contained all of the 0, 30 and 60 min 4sU samples, while batch 2 contained the 15 min 4sU samples only; Appendix Fig S2B), we detected a median of more than 6,400 transcripts (UMIs) and 2,800 genes per cell for all labeling conditions (Appendix Fig S2C and D). Pseudo-bulk level gene expression (i.e., gene expression summed up across all single cells in each sample) correlation analysis showed high correlation coefficients among all SLAM-Drop-seq samples (Pearson $r > 0.99$; Fig 1B), indicating highly reproducible sequencing results and no apparent changes in gene expression in response to short 4sU incubations.

Newly synthesized and preexisting RNA molecules are quantified for both precursor and mature mRNAs in single cells

To distinguish newly synthesized transcripts from preexisting ones, the 4sU labeling status for each transcript had to be determined. The identification of 4sU labeled RNA molecules relies on the identification of T->C transitions. However, single-nucleotide polymorphisms (SNPs) and sequencing errors can confound the quantification of labeled RNAs. To remove SNPs, we mapped the sequenced reads to a HEK293 cell line-specific reference genome, which we created based on public HEK293 genomic sequencing data (see Materials and Methods). After this, we still observed 3.1% transcripts containing T->C transitions in the no-4sU labeled control (i.e., 0 min 4sU) samples (Appendix Fig S3A). We hypothesized that these transitions were mainly due to the unfiltered SNPs and sequencing errors. We further filtered SNP positions identified from the no-4sU control samples (see Materials and Methods). Then the true T->C transitions were distinguished from sequencing errors using a Bayesian model (see Materials and Methods). The resulting T->C conversions detected in the no-4sU labeled control samples dropped to 0.3% (Appendix Fig S3A).

The labeling status (labeled or unlabeled) of a read enabled us to distinguish between newly synthesized and preexisting transcripts.
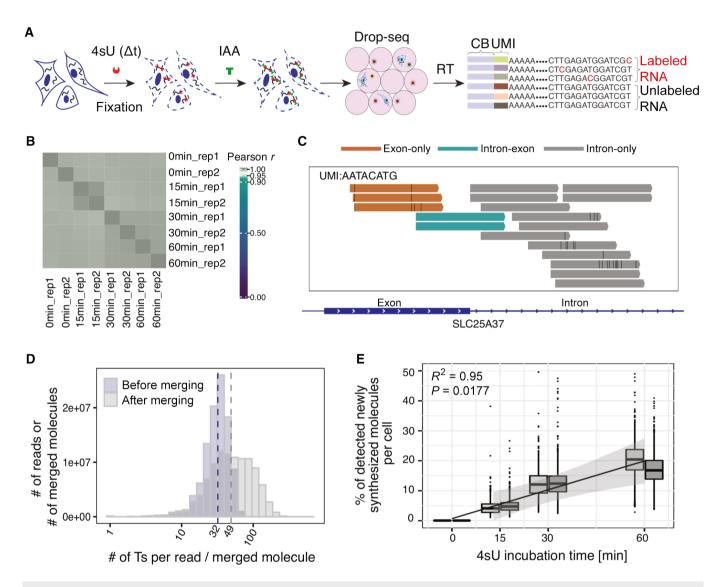
Figure 1. Newly synthesized and preexisting mRNAs in single cells are accurately detected with SLAM-Drop-seq by merging reads from the same UMI.

A   Schematic illustration of SLAM-Drop-seq experimental workflow. Briefly, cells were incubated with 4-thiouridine (4sU) and fixed. 4sU residues in newly synthesized RNAs were alkylated by iodoacetamide (IAA) *in situ* in fixed cells. This resulted in T->C transitions during the reverse transcription step, after encapsulation of single cells in oil droplets. scRNA-seq libraries were prepared after droplet lysis and the T->C conversions in barcoded reads were detected by computational analysis. CB = Cell barcode, UMI = Unique Molecular Identifier.

B   Pseudo-bulk gene expression levels are highly correlated between samples. Gene expression counts were summed across single cells in each sample and converted to counts per million (CPM). The log2(CPM + 1) values were used for the correlation analysis.

C   Different parts of the same transcript are captured and sequenced in SLAM-Drop-seq data. Example Genome browser track shows reads from one single UMI in one cell are mapped to both intronic and exonic regions of the *SLC25A37* gene.

D   The accuracy of identification of labeled transcripts increases after merging reads originated from the same transcript (UMI) as the observed number of thymidines (Ts) in each molecule increases. Data shown are from the 15 min 4sU labeled samples.

E   The molecules detected as newly synthesized increase proportionally to the 4sU labeling time. The regression was performed with the cross-replicate means per labeling time. The gray bands around the regression line display the 95% confidence intervals. The central band indicates the median value, while the lower and upper hinges of the boxes correspond to the 25th and 75th quantiles. The upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. There are two biological replicates for each 4sU incubation time, which are denoted by the side-by-side box plots. The number of cells in each boxplot is shown in Appendix Fig S2A.

This classification is central to our analysis, but it is highly dependent on the number of 4sU nucleotides incorporated per molecule. In our experiments, the median number of thymines (Ts) per read (median length: 138 nucleotides) is around 32 (Appendix Fig S3B). Thus, it is possible that reads originated from newly synthesized transcripts are devoid of T->C conversions because it is known that

even with long labeling times, at most one in 40 uridines is expected to be substituted by 4sU (Herzog *et al*, 2017; Jürges *et al*, 2018). To minimize the chance of such false negatives, we took advantage of unique molecular identifiers (UMIs) introduced by the SLAM-Drop-seq protocol to computationally merge the reads that mapped to the same transcript. Since reads associated with the same UMI can

cover different parts of the same transcript due to random fragmentation and PCR amplification, merging reads from the same UMI recovers the full coverage of the sequencing data (Fig 1C). The *in silico* merged "molecules," as we refer to them, generally contained sequence information for longer parts of the fragments (median length: 190 nucleotides; Appendix Fig S3B) and higher number of Ts (median: 49; Fig 1D), which increased the expected number of T->C transitions in the newly synthesized RNA molecules. To identify the newly synthesized transcripts lacking detectable T->C conversions by chance (increased likelihood with a lower number of Ts), we calculated the posterior probability of a molecule to be newly synthesized based on its T->C conversions and the number of Ts observed using a Bayesian model. To obtain the prior parameter to the model, we performed an extra experiment in which HEK293 cells were incubated with 4sU for 24 h to label all newly synthesized transcripts. With the assumption of a constant 4sU incorporation rate, we observed the number of T->C conversions in molecules with a given number of Ts following the expected Poisson distribution (Appendix Fig S3C). We applied the Bayesian model to calculate the probability for a given molecule to be newly synthesized based on its number of observed Ts and T->C conversions (Appendix Fig S3D, also see Materials and Methods). After these quantifications, we observed a linear increase of the newly synthesized RNA fractions with increasing time of 4sU incubation (Fig 1E), further arguing for the reliability of SLAM-Drop-seq, regarding the accurate quantification of newly synthesized and preexisting RNAs.

To be able to infer gene-specific RNA kinetic rates of transcription, processing, and degradation, we not only distinguished between newly synthesized (i.e., labeled) and preexisting (i.e., unlabeled) transcripts but also classified each molecule as a precursor (i.e., unspliced) or mature (i.e., spliced) RNA. The transcripts were identified as unspliced or spliced based on the presence or absence of intron coverage (Appendix Fig S4A, see Materials and Methods). The fractions of unspliced molecules are independent of 4sU incubation times (Appendix Fig S4B), meaning that metabolic labeling does not perturb mRNA processing. Thus, SLAM-Drop-seq can distinguish and measure the products of mRNA synthesis and processing in parallel and in a reliable manner.

Reconstruction of the cell cycle and modeling of cell cycle time-dependent RNA kinetic rates

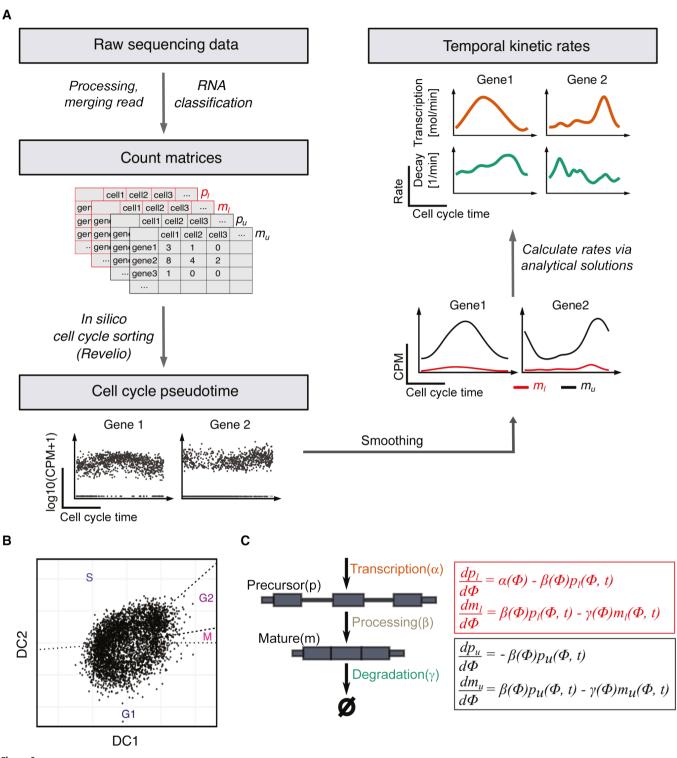From SLAM-Drop-seq data, we obtained gene expression count matrices for four types of mRNAs (i.e., "labeled precursor," "labeled mature," "unlabeled precursor," and "unlabeled mature"). For the calculation of RNA kinetic rates in single cells, we developed Eskrate, an R package, integrating the single-cell-level measurements of newly synthesized and preexisting precursor RNAs as well as newly synthesized and preexisting mature RNAs to infer gene-specific transcription, processing, and degradation rates as a function of time. The inputs to Eskrate are gene expression profiles of single cells of the aforementioned four types of RNA molecules measured by SLAM-Drop-seq and the biological times (e.g., cell cycle time) for every single cell (Fig 2A). We focus on the cell cycle as the first biological process to study the temporal dynamics of RNA kinetic rates, since gene expression is highly variable during the cell cycle and the underlying kinetic rates defining the expression of cycling genes are still not fully studied in unperturbed cells.

To analyze kinetic rates along the cell cycle, the individual cells first have to be classified according to cell cycle stages. To this end, we used Revelio, a principal component analysis (PCA)-based method (Schwabe *et al*, 2020), to reconstruct the cell cycle *in silico* from gene expression data of unsynchronized single cells. After filtering cells that showed more than 5% mitochondrial content, low expression of cycling genes, high expression of stress-related genes, or high contents of ribosomal protein-coding genes (see Materials and Methods), we analyzed the remaining cells using Revelio and obtained a two-dimensional representation placing each cell along a circular trajectory corresponding to their progression through the cell cycle (Figs 2B and EV1A and B). Analogously to Schwabe *et al* (2020) we utilized known cell cycle phase durations (Cheng & Solomon, 2008) to assign cell cycle phase boundaries and ordered the cells along the cell cycle process based on their angle in the two-dimensional space (Fig 2B). As expected, the mean captured RNA molecules per cell increased along the cell cycle progression (Fig EV1C). As an additional control, we compared the cell cycle time-dependent expression profiles recovered by Revelio between well-known cell cycle markers and housekeeping genes. While housekeeping genes such as HPRT1 show a constant expression profile along the cell cycle, the expression of well-known cell cycle marker genes is upregulated at specific cell cycle phases that recapitulate their known expression patterns (Whitfield *et al*, 2002; Fig EV1D).

Based on the obtained cell cycle time-resolved expression profiles of all four mRNA types, we implemented Eskrate by extending an RNA kinetic rate model commonly used for steady-state analyses (Zeisel *et al*, 2011; Manno *et al*, 2018) to a system of four ordinary differential equations (ODEs) describing RNA abundance changes over the cell cycle time for the described four types of RNAs

---

Figure 2.   SLAM-Drop-seq of unsynchronized HEK293 cells reveals temporally regulated RNA kinetic rates throughout the cell cycle.

A   The main steps of the data processing pipeline and new algorithm Eskrate (Estimate time-dependent RNA kinetic rates) for estimating time-dependent RNA kinetic rates from SLAM-Drop-seq data. Step 1. Raw sequencing data is processed and reads from the same UMI (transcript) are merged. Different RNA types are classified (i.e., labeled precursor p_l, labeled mature m_l, unlabeled precursors .25emp_u and unlabeled mature m_u) and gene expression counts are obtained for each RNA type. Step 2. After low quality cells are filtered, we use Revelio (Schwabe *et al*, 2020) to sort cells in cell cycle times. Single cell gene expression is transformed to gene expression over cell cycle time. Step 3. Time-dependent RNA kinetic rates are calculated using Eskrate via the simplified analytical solutions of the kinetic rate model (see Materials and Methods).

B   Based on gene expression, individual HEK293 cells were positioned along a circular trajectory representing the cell cycle using Revelio (Schwabe *et al*, 2020). DC = dynamic component. Shown are cells from the batch 1 experiments (0, 30 and 60 min 4sU samples).

C   The mathematical model of RNA kinetic rates. The model consists of two mRNA maturation states (precursor and mature) and includes three kinetic rates: transcription α, processing β and degradation γ. Since maturation status and labeling status are independent events, we obtain four RNA types from the SLAM-Drop-seq experiment: p_l, .5emm_l, p_u, and m_u. We model the dynamics of RNA abundances for each RNA type over the cell cycle time by ODEs involving the kinetic rates. RNA abundances are measured at given cell cycle time Φ and 4sU labeling time t.

A

B

C

$$\frac{dp_l}{d\Phi} = \alpha(\Phi) - \beta(\Phi)p_l(\Phi, t)$$

$$\frac{dm_l}{d\Phi} = \beta(\Phi)p_l(\Phi, t) - \gamma(\Phi)m_l(\Phi, t)$$

$$\frac{dp_u}{d\Phi} = -\beta(\Phi)p_u(\Phi, t)$$

$$\frac{dm_u}{d\Phi} = \beta(\Phi)p_u(\Phi, t) - \gamma(\Phi)m_u(\Phi, t)$$

Figure 2.

(Fig 2C). Importantly, we solved the time-dependent versions of these equations such that all kinetic rates are (continuous) functions of the cell cycle time. In this way, we obtained equations directly: linking the time-dependent abundances of the four RNA types to the time-dependent rates of transcription, splicing, and degradation (see Materials and Methods). Due to the fact that we captured a low amount of precursor mRNAs (Appendix Fig S4B), we simplified the

analytical solutions using approximations assuming that RNA processing (within minutes; Alpert *et al*, 2017) is much faster than RNA degradation (within hours; Murakawa *et al*, 2015; Schofield *et al*, 2018). We implemented Eskrate with simplified analytical solutions in which the time-dependent transcription and degradation rates were driven by time courses of labeled and unlabeled mature RNAs (see Materials and Methods).

While the derived analytic solutions enable us to calculate kinetic rates from the experimental data, one drawback is that the ground truth is unknown. Therefore, as a proof-of-concept, we generated synthetic data assuming that both transcription and degradation rates are cell cycle time-dependent (Appendix Fig S5A–C, also see Materials and Methods). We show that given a ground truth for kinetic rates, which generate certain expression profiles, the calculated kinetic rates accurately recapitulate the ground truths in all labeling cases using the full model (Appendix Fig S5D, see Materials and Methods). Using the simplified model, the kinetic rates are generally close to the ground truths as well. However, the rates calculated from longer 4sU labeling times are more precise (Appendix Fig S5D). Since we can only approximate the kinetic rates from the experimental data using the simplified model, in the following analyses, we show the results of kinetic rates in the cell cycle calculated from the 60 min 4sU labeled samples. Since it is theoretically possible that the same RNA levels result from different kinetic rate combinations, we additionally generated the same gene expression profile of mature RNA considering all possible kinetic regulation modes (i.e., constant transcription rate; constant degradation rate; and dynamic transcription and degradation rates). The calculated RNA kinetic rates recapitulate the simulated input transcription and degradation rates in all cases (Appendix Fig S6A). Therefore, both the full model and the simplified model are capable of distinguishing between gene expression patterns driven by changes in transcription compared to those driven by changes in degradation, even if both modes of regulation ultimately lead to the exact same time course of mature mRNA expression.

We calculate the transcription and degradation rates from the observed gene expression data (i.e., observation). Reversely, we can calculate the predicted gene expression (i.e., prediction) from the calculated kinetic rates (see Materials and Methods). The difference between the observation and the prediction can therefore be used as a measurement of the accuracy of the calculated rates. Using synthetic data, we show when the changes are small in transcription and degradation rates, the prediction changes linearly (or near linearly; Appendix Fig S6B and C). This analysis suggests that when errors are small in the calculated kinetic rates, the prediction should also contain small errors. In other words, the prediction should be close to the observation, if the calculated rates are close to the truth. Thus, we define genes as "well-predicted" for which an estimation of transcription and degradation rates is meaningful if the difference between prediction and observation is small (see Materials and Methods).

RNA transcription and degradation rates are dynamically regulated in the cell cycle

As explained above, we analyzed the kinetic rates in the cell cycle using the simplified model from the 60 min 4sU samples. Since single-cell transcriptomic data are noisy, smoothed expression profiles along the cell cycle are used for the estimations (see Materials and Methods). To investigate the regulation of temporal RNA kinetic rates during the cell cycle, we only focused on genes whose expression oscillates along the cell cycle (i.e., cycling genes). The reliability of the smoothed profile for each gene is largely dependent on its dropout rate (Qiu, 2020). To avoid unreliable estimates from shallow gene expression data (Appendix Fig S7A), we down-sampled the gene expression matrix and defined thresholds of dropout rates

for cycling genes to consider the smoothed profiles reliable (Appendix Fig S7B–D, also see Materials and Methods). Applying these thresholds, we defined a core set of 399 cycling genes with confidence in their smoothed profiles.
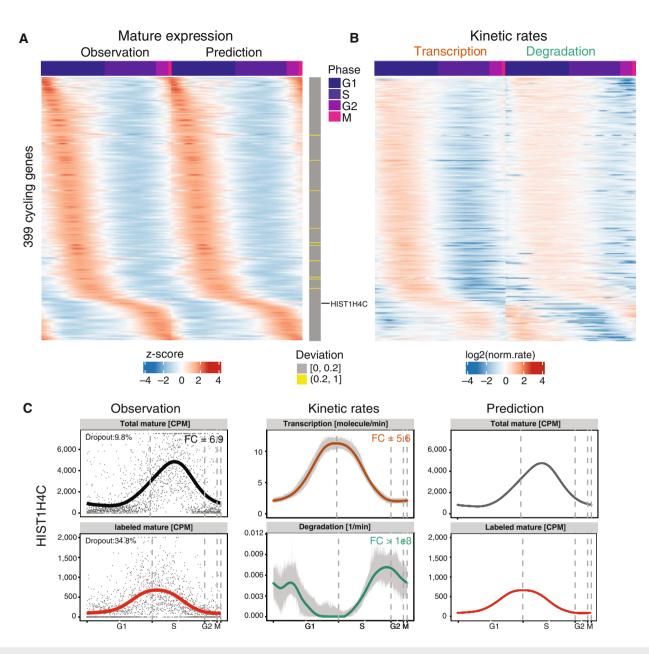
The expression profiles of these 399 cycling genes peak at specific stages of the cell cycle (Fig 3A). We identified 97% of them to be well-predicted by comparing gene-wise the predictions to the observations (Fig 3A). The corresponding calculated transcription and degradation rates for cycling genes presented dynamic patterns as well along the cell cycle (Fig 3B), indicating temporal regulation of both processes during cell cycle progression. Noticeably, the peak times of transcription and degradation are less ordered with respect to the peak times in gene expression (Fig 3A and B), indicating that distinct regulation patterns for transcription and degradation can yield similar mRNA expression profiles.

As an example, of a dynamically regulated gene, we show the detailed profiles for the observed gene expression and the calculated kinetic rates along the cell cycle for the HIST1H4C gene (Fig 3C). It is a well-known S-phase enriched gene and its observed gene expression peaked exactly during the S phase in our data. Its transcription rate peaked in the S phase, while its degradation rate showed a sharp increase preceding the M phase. These results are highly consistent with the known kinetic regulation of replication-dependent histone genes: in fact, it was previously reported that the upregulation of gene expression is mainly dependent on increased transcription, whereas the G2 phase associated decrease in detection is caused by decreased RNA stability (Harris et al, 1991).

To evaluate the robustness of our method and estimations, we compared the cell cycle time-dependent profiles of RNA expression, synthesis, and degradation rates from our data with those of another cell line (RPE1-FUCCI), obtained by using a different approach (scEU-seq, Battich et al, 2020). The comparison of shared genes between the two datasets revealed a similar overall pattern of peaking times along the cell cycle, but with a relatively consistent time delay (Fig EV2A–C). HEK293 kinetics profiles peaked at an earlier cell cycle time with respect to the one calculated in the scEU-seq paper. Interestingly, RPE1-FUCCI cells seem to lack any transcriptional activity at the beginning of the cell cycle. While various reasons could account for these discrepancies, the genetic background, cell of origin, cell source, and ways of cell cycle sorting are likely contributors.

We also tested the reproducibility of our data and the reliability of the developed algorithm Eskrate by calculating the synthesis and degradation rates using the raw sequencing data from the published sci-fate dataset (Cao et al, 2020). Specifically, we used the 4sU-labeled and unstimulated A549 cell line sample to estimate the RNA kinetic rates. When compared to the 140 shared genes with HEK293 results, the two datasets exhibited a high degree of similarity (Fig EV2D–F). This finding not only confirms the reliability of the Eskrate algorithm by demonstrating its potential to analyze other datasets but also highlights the dynamic regulation of RNA kinetic rates in cell cycle variable genes across different human cell lines.

Furthermore, we averaged the calculated degradation rates and compared them to independent studies on RNA half-lives in HEK293 and K562 cells calculated from constant degradation estimates (Murakawa et al, 2015; Schofield et al, 2018). We found that our estimates correlated with published data (Spearman correlation coefficients were 0.60 and 0.46, respectively; Appendix Fig S8A and B).
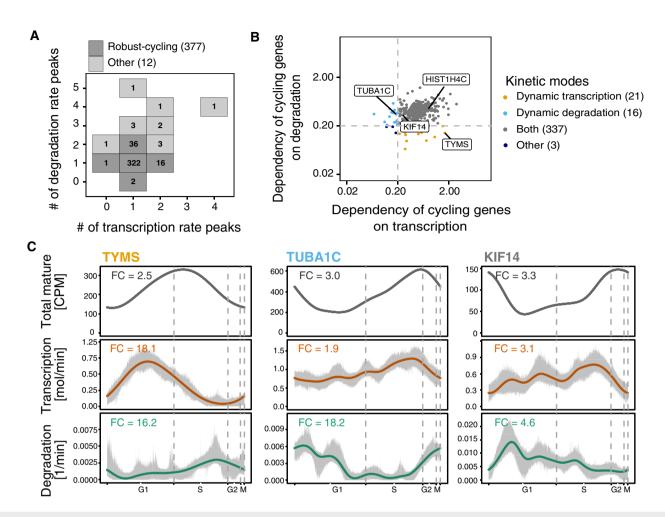
Figure 3. Gene expression and kinetic rates for cycling genes are highly dynamic along the cell cycle.

A   Gene expression of mature mRNAs for 399 cycling genes as quantified from sequencing data (i.e., observation) and as predicted (i.e., prediction) from calculated transcription and degradation rates along the cell cycle. The gray vertical bars on the right show the mean relative deviations between observation and prediction (see Materials and Methods). 389 out of those genes were identified as well-predicted (deviation ≤ 0.2). Genes on rows were ordered by the peak time of the observation. Expression values were centered and scaled across cells for each gene.

B   The calculated transcription and degradation rates for the genes shown in (A) along the cell cycle. The rates were normalized to their mean per gene and log2 transformed (transcription rates [molecules/min], degradation rates [1/min]).

C   Profiles of the observed gene expression, the calculated kinetic rates and the corresponding predicted expression of *HIST1H4C* gene. Left panel: Observed gene expression levels (i.e., points) and their smoothed profiles (i.e., lines). CPM, counts per million. Total mature RNA is the sum of labeled and unlabeled mature RNAs. Middle panel: Kinetic rates calculated from the smoothed gene expression profiles are shown by the solid lines. The gray bands show the 90% confidence intervals (i.e., the 5–95% quantiles) of the rates calculated from bootstrapping (resampling of cells 100 times with replacement). Right panel: The predicted expression profiles of labeled and total mature RNAs that derived from the calculated rates.

Cycling genes exhibit different modes of kinetic regulation

We wondered how frequent different modes of kinetic regulation are in human cycling genes and which kinetic parameters

(transcription and/or degradation rate) drive the dynamic changes of gene expression. To this end, we filtered the well-predicted cycling genes to exclude those that exhibit overly "wiggly" transcription or degradation profiles and obtained a set of 377 genes,

Figure 4. Cycling gene expression can result from various kinetic modes that often involve dynamically regulated RNA decay.

A Identification of the robust-cycling genes for kinetic rate regulation investigation. 377 out of the 389 well-predicted cycling genes were defined as robust-cycling genes by excluding those showing multiple transcription or degradation peaks.

B Cycling genes exhibit diverse kinetic modes. The kinetic modes were defined based on the dependency of dynamic transcription and dynamic degradation. The dependency on transcription (or degradation) is represented by the mean relative deviation between prediction from constant transcription (or degradation) and prediction from dynamic model (see Materials and Methods). 'Both' means genes are regulated by both dynamic transcription and dynamic degradation. Genes that were not identified into dynamic transcription, dynamic degradation or both were assigned into the "other" group. The labeled genes represent example cases in each mode.

C Profiles of the predicted expression, transcription and degradation rates along the cell cycle for example genes, which are labeled in (B) (dynamic transcription: TYMS; dynamic degradation: TUBA1C; dynamic transcription and dynamic degradation: KIF14). Gray areas around rates' profiles indicate the 90% confidence intervals calculated from bootstrapping (resampling of cells 100 times with replacement).

which we called "robust-cycling genes" (Fig 4A). We classified these genes according to their dependency on transcription and degradation dynamics by determining the predicted gene expression changes when kinetic rates are set constant. We hypothesized that forcing a constant transcription rate should lead to large deviations of the resulting expression profile from the observed one for genes temporally regulated by dynamic transcription. Analogously, for genes with their temporal expression patterns dominantly regulated by dynamic turnover, a constant degradation rate should cause large deviation in its predicted expression. Using this approach, we assigned the robust-cycling genes to the following three major regulatory classes: (i) dynamic transcription, (ii) dynamic degradation; (iii) dynamic transcription and degradation (Fig 4B; Dataset EV1).

We found that the majority of the robust-cycling genes (337 out of 377) depend on dynamic changes in both rates, implicating

posttranscriptional regulation as an important regulatory mechanism for mRNA expression and oscillating gene expression during the cell cycle in HEK293 cells. Surprisingly, we only found 21 genes that achieve their cycling mRNA levels by dynamic transcription alone and 16 by dynamic changes in degradation rates (Fig 4B). As expected, when looking at the kinetic profiles without fixed parameter, genes classified as "dynamic transcription regulated" showed relatively constant degradation rate profiles throughout the cell cycle (e.g., TYMS in Fig 4C), while genes classified as "dynamic degradation regulated" exhibited roughly constant transcription rate profiles (e.g., TUBA1C in Fig 4C).

Our results show that expression of the majority of the cycling mRNAs is regulated at the level of transcription as well as degradation (Fig 4B). Examples for this mode of regulation are HIST1H4C (Fig 3C) and KIF14 mRNAs. KIF14 encodes a member

of the kinesin-3 superfamily of microtubule motor proteins, which is involved in chromosome segregation, mitotic spindle formation, and cytokinesis during the cell cycle (Carleton *et al*, 2006). KIF14 mRNA levels went up before G2 (Fig 4C), mostly due to increased transcription (Fig 4C). Toward the end of mitosis, the KIF14 transcript numbers quickly diminished due to a concomitant decrease in transcription and increase in degradation (Fig 4C).

To investigate how gene expression changes are defined by changes of transcription and degradation rates, we compared the peak times of both parameters for the 337 robust-cycling genes. The lag times between the two rates come from a broad distribution (Appendix Fig S9A), indicating highly variable and gene-specific coordination of mRNA kinetic rates. For example, mRNA expression profiles of SYNE2 and CLSPN genes peak during the G1 phase (Appendix Fig S9B). While their transcription rates are similar, their degradation rates have very different peak patterns. Overall, our findings demonstrate that our approach enables a description of transcription and degradation rates during the cell cycle with high temporal resolution.

Discussion

In this study, we introduced SLAM-Drop-seq to distinguish newly synthesized and preexisting polyadenylated RNAs in single cells. We developed the R package Eskrate that enabled us to study temporal changes of RNA kinetic rates during the cell cycle based on SLAM-Drop-seq data. We inferred RNA kinetic rates in terms of cell cycle time from unsynchronized HEK293 cells and identified 377 robust-cycling genes with confident rate estimates. We found the majority of cycling genes to be temporally regulated by changes in not only synthesis but also degradation rates. The temporal interplay between transcription and decay to achieve dynamic gene expression along the cell cycle revealed different modes of mRNA kinetic rates.

Previous studies on mRNA expression regulation throughout the cell cycle largely focused on transcriptional regulation (Cho *et al*, 2001; Liu *et al*, 2017). The underlying assumption of transcriptional control as a main regulatory mechanism is the basis of many methods to estimate RNA kinetic rates (Skinner *et al*, 2016). Here, we show that only a small fraction of cycling genes seems to be temporally regulated by transcription alone. Therefore, at least for cell cycle regulation in HEK293 cells, our data suggest that the generalization provided by our approach (time-dependent rates with biophysical units) is not only of theoretical interest but also key in revealing the underlying biological processes: we found the majority of cycling genes to rely on some level of time-dependent degradation in addition to their transcriptional regulation to achieve final expression patterns. In addition, we also identified 16 robust-cycling genes with near-constant transcription over the entire cell cycle that solely relied on dynamic degradation rates to modulate their expression (Fig 4B). A recent study showed that alternative polyadenylation site usage is highly regulated in cell-cycle-related genes at the single cell level, suggesting that posttranscriptional regulation is important for the regulation of gene expression during cell cycle progression (Wang *et al*, 2022). The switch of polyA sites of several cell cycle genes could potentially influence their turnover rate. Our study thus underscores the diverse temporal regulation types of transcription and

degradation for defining gene expression in cell cycle progression. While the temporal changes of transcription and degradation rates have been previously revealed during the cell cycle (Eser *et al*, 2014; Battich *et al*, 2020), we infer these rates with a greater time resolution.

While this study was conducted, several approaches to measuring RNA half lives in single cells were published (Erhard *et al*, 2019; Hendriks *et al*, 2019; Battich *et al*, 2020; Cao *et al*, 2020; Qiu *et al*, 2020, 2022). NASC-seq and scSLAM-seq combined metabolic RNA labeling with the Smart-seq2 approach (Picelli *et al*, 2013). While these methods have the advantage of full-length coverage of RNA transcripts compared to Drop-Seq (3′ end sequencing), the cell throughput is an order of magnitude smaller, capturing only a few hundred single cells (Erhard *et al*, 2019; Hendriks *et al*, 2019). scEU-seq performed pull-down to enrich for metabolically labeled RNAs, which is laborious, low-throughput, and prone to introduce biases. Sci-fate gained high sequencing throughput by using combinatorial indexing strategies (Cao *et al*, 2020). scNT-seq adapted RNA metabolic labeling to droplet-based sequencing approach, which effectively increases the throughput at a low cost (Qiu *et al*, 2020). Regarding the estimation of RNA kinetic rates, most of these studies (Erhard *et al*, 2019; Hendriks *et al*, 2019; Cao *et al*, 2020; Qiu *et al*, 2020) fit the degradation rates under a steady-state assumption. Battich *et al* (2020) estimated the synthesis and degradation rates in the cell cycle with lower pseudotime resolution by pooling cells at close cell cycle times (Battich *et al*, 2020). One advanced computation tool was proposed by Qiu *et al* (2022), by which dynamical models and machine learning were used to predict dynamic cell transitions (Qiu *et al*, 2022). Compared to these studies, SLAM-Drop-seq obtains a high number of single cells with less labor at a lower cost. The 4sU labeled nucleotides are chemically converted *in situ* in fixed cells, which makes the method applicable for primary cells and complex tissues. Moreover, the kinetic model and algorithm we implemented take all kinetic rates as a function of time, meaning we are able to predict the absolute number of transcripts that are produced during a certain (cell cycle) time.

We are aware that the Drop-seq-based sequencing method has a 3′ bias, and this can lead to potential underestimation of unspliced RNA molecules (i.e., precursors). Nonetheless, due to the simplification of the mRNA kinetic model, the precursor mRNAs were not used for the kinetic rates calculation. Thus the 3′ bias of the sequencing should not influence the kinetic rates shown in this manuscript. However, full-length scRNA-seq methods might be better options to apply the full kinetic rate model if the number of single cells could scale to similar levels to Drop-seq. Alternatively, when the accuracy of base calling for long-read sequencing will match that of short read sequencing, their incorporation into the workflow will be a more reliable alternative.

Due to the low number of precursor RNA captured in our dataset (Appendix Fig S4B), we decided to adjust our model to increase the robustness of the transcription and degradation rates at the cost of not being able to investigate mRNA processing (i.e., splicing) rates for our application of SLAM-Drop-seq to cycling HEK293 cells. However, with deeper sequencing data, our mathematical framework should be capable of correctly estimating time-dependent splicing rates as well.

Materials and Methods

Reagents and Tools table

Reagent/Resource	Reference/ Source	Identifier/Catalog number
Experimental models		
HEK293 Flp-In T-Rex cells	Thermo Fisher Scientific	Cat # R78007
Oligonucleotides		
Read1CustSeqB	Macosko *et al* (2015)	GCCTGTCCGCGGAAGCAGTGGTATCA ACGCAGAGTAC
TSO	Macosko *et al* (2015)	AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG
SMART PCR oligo	Picelli *et al* (2014)	AAGCAGTGGTATCAACGCAGAGT
New-P5-SMART PCR hybrid oligo	Macosko *et al* (2015)	AATGATACGGCGACCACCGAGATCTAC ACGCCTGTCCGCGGAAGCAGTGGTATC AACGCAGAGT*A*C
Chemicals, enzymes and other reagents		
Dulbecco's modified Eagle medium	Gibco	Cat # 41965039
AMPure XP beads	Beckman Coulter	Cat # A63881
Bovine serum albumin (BSA)	Sigma-Aldrich	Cat # A8806
Dithiothreitol (DTT)	Sigma-Aldrich	Cat # D9779
DPBS	Gibco	Cat # 14190250
ECL detection reagent	GE Healthcare	Cat # RPN2209
Exonuclease I	New England Biolabs	Cat # M0293
Fetal bovine serum	Gibco	Cat # 10270106
Ficoll PM 400	Sigma-Aldrich	Cat # F4375
ʟ-Glutamine	Gibco	Cat # 25030081
Iodoacetamide	Sigma-Aldrich	Cat # I6125-5G
Maxima H- RT enzyme	Thermo Fisher	Cat # EP0753
Methanol z. A. (min. 99.8%)	Th. Geyer	Cat # 11646935
MTSEA-XX-biotin	Biotium	Cat # 90066
Nylon membrane	Amersham Hybond-N+	Cat # RPN203B
N-Lauroylsarcosine (Sarkosyl)	Sigma-Aldrich	Cat # L7414
Phase Lock Gel Heavy tubes	QuantaBio	Cat # 2302830
Phenol/Chloroform/Isoamylalkohol	Carl Roth	Cat # A156.1
QX200 Droplet Generation Oil	Bio-Rad	Cat #1864006
SDS 20%	Carl Roth	Cat # 1057.1
Streptavidin-HRP	Pierce	Cat # 21130
Superase•In RNAse Inhibitor	Thermo Fisher	Cat # AM2694
4-Thio uridine	ChemGem	Cat # RP-2304
Trizol	Thermo Fisher	Cat # 15-596-018
TrypLE express enzyme	Gibco	Cat # 12605036
Software		
bcl2fastq v2.20.0		
STAR v2.6.0a	Dobin *et al* (2013)	
bcftools v1.9	Li (2011)	
Drop-seq tools v2.2.0	Macosko *et al* (2015)	
Revelio	Schwabe *et al* (2020)	
GATK toolkit	McKenna *et al* (2010)	
Seurat	Macosko *et al* (2015)	

Reagents and Tools table (continued)

Reagent/Resource	Reference/ Source	Identifier/Catalog number
DropletUtils	Griffiths *et al* (2018) and Lun *et al* (2019)	
FastQC v0.11.5		
Samtools v1.6		
Bedtools v2.30.0		
Pysam v 0.15.4		
VarScan v2.3.9		
sam2tsv.jar		
Python 2.7.18 & python 3.9.0		
R 'mgcv' packages	Wood (2017)	
R 3.6.0		
Other		
Nadia instrument	Dolomite Bio	
TC20 cell counter	Bio-Rad	
Bio-Dot apparatus	Bio-Rad	
Direct-zol RNA Miniprep	Zymo Research	
QuBit 2.0 Fluorometer	Thermo Fisher	
UV Stratalinker Ultraviolet Crosslinker	Stratagene	
Imager AI680	GE Amersham	
2100 Bioanalyzer Instrument	Agilent	
Nextera XT DNA Library Preparation Kit	Illumina	
Nextseq 500/550 HO v2 kit	Illumina	
Illumina Nextseq 500	Illumina	
Illumina NovaSeq 6000	Illumina	

Methods and Protocols

Cell culture

Human HEK293 Flp-In T-Rex cells were cultured at 37°C with 5% CO_2 and were grown in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and 2 mM L-Glutamine. TrypLE Express Enzyme was used to detach cells and split them 2–3 times per week. The cell line was tested for mycoplasma contamination before experiments.

SLAM-Drop-seq protocol

- Split cells the day before the experiment in order to reach ~ 60/70% confluence the day after (exponential growth phase).
- Incubate cells with 300 μM 4sU for the desired amount of time.
- Wash cells without detaching from the plate with warm DPBS and dissociate them with TrypLE Express Enzyme for 1–2 min.
- Using cold DPBS, first wash cells in cold DPBS and then resuspend them at a final concentration of 1–2 million cells/ml (or lower).
  CRITICAL STEPS: be as quick as possible in this and previous steps, since transcription happening during this time will introduce unwanted signal to the results.

- Fix cells by gently adding drop by drop four volumes of cold methanol, with mild shaking of the dissociated single cells.
- Leave samples in fixation buffer (80% methanol, 20% DPBS) for at least 20 min at −20°C.
- Add IAA to final concentration of 10 mM to alkylate 4sU residues in cell suspensions. Same volume of fixation buffer was added to control samples.
- Alkylation reaction is performed overnight (16–18 h) in the dark with mild rotation at room temperature.
- Using the fixation buffer, wash cell samples once (300 *g*, 2′, 4°C) and resuspend them. Storage at −80°C is possible.
  SUGGESTED STEP: split the samples to control if the alkylation was efficient. Refer to the next section "Biotinylation blocking assay" to perform it.
- Rehydrate samples in rehydration buffer (DPBS, 0.01% BSA, 1:100 Superase · In RNAse Inhibitor).
- Quench the remaining IAA by adding 100 mM DTT (from freshly prepared 1 M DTT, dissolved in rehydration buffer) for 5 min at room temperature.
- Wash samples once in rehydration buffer and resuspend them in 400 μl DPBS - BSA 0.01% + 1 U/μl Superase · In RNAse Inhibitor.
- Pass cell suspensions through a 40 μm cell strainer and count single cells (we used a TC20 automated cell counter).
- Proceed immediately to the chosen approach for single-cell RNA library preparation.

N.B.: when working with 4sU and IAA reduce to the minimum light exposure, because both molecules are unstable and light-sensitive.

Biotinylation blocking assay protocol

- After the overnight alkylation step, spin down a fraction of the samples (at least 50,000 cells) and dissolve it in TRIzol to check for alkylation efficiency.
- Extract the RNA following the TRIzol Reagent manual or using Direct-zol purification columns.
- Biotinylate the purified RNA (same amount between samples) with MTSEA-XX-Biotin in biotinylation buffer (20 mM Tris–HCl pH 7.5, 1 mM EDTA, in ddH2O, 100 μl final volume) for more than 1 h in the dark at room temperature.
- Remove excess of MTSEA-XX-Biotin with Phenol/Chloroform (for RNA extraction) by using Phase Lock Gel Heavy tubes.
- Blot the samples to a nylon membrane ($\sim$ 100 μl) using a dot-blot apparatus.
- Cross-link the RNA to the membrane with 2,400 μJ UV_{254nm}.
- Block the membrane in blocking solution (DPBS, 10% SDS, 1 mM EDTA) for 20 min at room temperature.
- Incubate the membrane with 1:10,000 dilution of 1 mg/ml streptavidin-HRP in blocking solution for 10 min at room temperature.
- Wash the membrane six times with a blocking solution containing decreasing concentration of SDS (10, 1, 0.1%, applied twice each) for 10 min.
- Develop the biotin signal with ECL detection reagent and detect the chemiluminescence with an appropriate imager system.

Single-cell library preparation and sequencing

Single cells were encapsulated using a Nadia Instrument and following the protocol from the manufacturer (version 1.8). 75,000 single cells in 250 μl rehydration buffer were encapsulated using 250 μl lysis buffer (6% Ficoll PM-400, 0.2% Sarkosyl, 20 mM EDTA, 50 mM DTT, 200 mM Tris pH 7.5) and 3 ml QX200 Droplet Generation Oil. After encapsulation, the transcripts captured by barcoded beads were reverse transcribed (Maxima H- RT enzyme) and treated with exonuclease I following the instruction manual. The beads were than counted, and PCR amplification was run with SMART PCR primers (4,000 beads/sample, 6 reactions, 14 cycles). The PCR reactions were purified two times with 0.6 volumes of AMPure XP beads, quantified with QuBit, and analyzed on a Bioanalyzer DNA HS chip. 1,000 pg DNA libraries were used for the fragmentation and amplification steps (11 cycles) using the Nextera XT v2 DNA sample preparation kit. The libraries were double-purified with 0.6 volumes of AMPure XP Beads, quantified, and pooled. The 15 min 4sU-labeled samples were sequenced on a Illumina Nextseq500 sequencer (library concentration 1.8 pM; Nextseq 500/550 High Output v2 kit (150 cycles) in paired-end mode; read 1 = 20/21 nt using the custom primer Read1CustSeqB (Macosko *et al*, 2015), read 2 = 133/132 nt). The 0, 30 and 60 min 4sU-labeled samples were sequenced on a Illumina NovaSeq 6000 device (SP configuration, pair-end; read 1 = 20 nt using the custom primer Read1CustSeqB (Macosko *et al*, 2015), index 1 (i7) = 8 nt, read 2 = 150 nt).

HEK293 cell line–specific reference genome creation

To increase mapping sensitivity and most importantly avoid mistaking single-nucleotide polymorphisms (SNPs) from the minor allele in HEK293 cells as T->C conversions, a cell type-–specific reference genome was created by amending the hg38 reference genome using SNPs detected from public bulk genome DNA (gDNA) sequencing data (SRA accession number: SRR2123657). To identify the SNPs in the HEK293 gDNA data, we mapped raw reads to the hg38 reference genome using STAR v2.6.0a (Dobin *et al*, 2013). SNPs were called using bcftools v1.9 (Li, 2011) "mpileup" and "call" functions. Then we filtered SNPs by the quality score using bcftools "view" function with parameter "%QUAL>20." The generated SNP vcf file was then sorted to list chromosomes in lexicographic order. Very importantly, indels were filtered out to avoid changes in the length of the reference. Finally, the hg38 reference genome was corrected from the identified SNPs in HEK293 gDNA data using the "FastaAlternateReference-Maker" function from the GATK toolkit (McKenna *et al*, 2010).

Data processing, mapping, splice status tagging, and mismatches tagging

Raw sequencing data were demultiplexed using bcl2fastq v2.20.0. The sequencing quality was checked using FastQC v0.11.5. Since we applied Drop-seq for single-cell RNA sequencing, we followed the Drop-seq Core Computational Protocol (https://github.com/broadinstitute/Drop-seq) to process the data. Drop-seq tool v2.2.0 (Macosko *et al*, 2015) was used to tag cell barcodes and molecular barcodes, to trim the 3′ poly(A) tail and potential 5' SMART adapter sequences, and to filter out barcodes with low quality bases. The reads were then aligned to the HEK293 reference genome that we described above using STAR v2.6.0a (Dobin *et al*, 2013) with optional parameters "-outFilterMismatchNmax 10 and –outSAMat-tributes AS NH nM NM MD" in order to write the MD tag, which carried the encoding mismatched and deleted reference bases into the aligned output files. Typically, around 80% of the reads mapped uniquely. Non-uniquely mapped reads were not used for downstream analysis.

Gene annotation tags were added to the aligned reads by using the Drop-seq tools v2.2.0 with filtered genome annotation file (GENCODE annotation release 29), which is described below. As we wanted to estimate kinetic rates for well-annotated genes, we filtered the genome annotation file to keep well-supported transcripts (transcript support level: 1, 2 and NA) for protein coding genes. For our interest of exploring other gene types, we also added "lincRNA" and "miRNA" from the genome annotation to our filtered gtf file. The Drop-seq tools was further exploited to identify and correct potential barcode errors. The number of cells (cell barcodes associated with single-cell transcriptomes) was determined by extracting the number of reads per cell, then plotting the cumulative distribution of reads against the cell barcodes ordered by descending number of reads and selecting the inflection point ('knee') of the distribution.

The gene annotation tags that were added by Drop-seq Tools contained genes on both forward and reverse stands within the genomic location that the read mapped to. To annotate the specific gene name that the read mapped to, we further tagged the aligned reads with gene names stating the unique gene the read mapped to by considering the strands information and the genes that located at the same strand that the read mapped to.

In order to quantify mature and precursor RNAs for the kinetic rates modeling, we annotated the splice status for each uniquely aligned read with a tag stating the specific genomic location that the read mapped to (i.e., intron-only, intron-exon, exon-only, exon-exon, etc.) by intersecting the aligned read with gene annotation bed files containing intron and exon coordination information using bedtools "intersect" function. The bed files were generated from the gene annotation gtf file and split into intron, exon, and ambiguous bed files. To avoid ambiguous tagging where both intron and exon present at the read mapped locus, we annotated the ambiguous loci from the bed files and filtered them from the intron and exon bed files. The quantification of spliced and unspliced transcripts was described in later sections.

To count the T->C conversions, we need to find out T->C mismatches between the sequenced reads and the reference genome. We tagged each aligned read with a specific mismatch tag indicating the number of mismatches for all mismatch types using a python script that was adapted from "conversiontag" function in NASC-seq analysis pipeline (Hendriks *et al*, 2019). This script utilized pysam (https://github.com/pysam-developers/pysam) to call single base mismatches by using the MD tags in conjunction with the CIGAR tags, which carried the mismatches, insertions, and deletions in the reads in the aligned files without requiring access to the entire original reference. Since the reads in SLAM-Drop-seq were not-stranded, both sense and antisense transcripts were sequenced. As the reads were from the first strand of cDNA, T->C mismatches seen in reads mapped to the sense strand and A->G mismatches seen in reads mapped to the antisense strand were counted as T->C conversions. T->C conversions for each gene in each cell were counted by summing up the unique T->C positions in all reads that uniquely mapped to the gene in each single cell, which was described in the later section.

SNPs filtering

To reduce mistakes in T->C conversion quantification caused by SNPs, we mapped sequencing reads to the HEK293 specific reference genome as described in the mapping step above. To further get rid of SNPs that possibly presented in different cell passages, we checked the observed mismatches in no-4sU labeled samples and identified mismatches in positions with high frequency (mismatch rate > 0.5 in position with reads depth $\geq$ 2) across all cells as SNPs. We then ignored these defined SNP positions when we count T->C conversions in all SLAM-Drop-seq samples.

Reads merging and transcript splice status identification

In SLAM-Drop-seq, unique molecular identifiers (UMIs) were used to identify different transcripts. The UMI tag was added to the first stand cDNA, which was PCR amplified and fragmented during library preparation. Since the fragmentation was random, different parts of each transcript with the same UMI were sequenced. To check the T->C conversion in each transcript (UMI), we merged all reads with the same UMI in each cell by collapsing the overlapping positions. Meanwhile, we checked each position in the merged sequence where at least one mismatch was observed and marked the read coverage and mismatch number at that position.

To identify the splice status of the sequenced transcript at the time that the 4sU labeling experiments ended, we checked the splice status tags of all reads that with the same UMI and identified the splice status for each transcript with the following logics: (i) If all reads from the transcript were mapped to exonic regions, the transcript was identified as "spliced"; (ii) If there was at least one read from the transcript were mapped to intronic regions, the transcript was identified as "unspliced"; (iii) Transcripts were defined as "ambiguous" if they fit neither of the above criteria, and they were not used for the kinetic estimation.

Newly synthesized and preexisting transcript quantification

To quantify the newly synthesized and preexisting transcripts, we need to identify labeling status from the T->C conversions observed in each transcript. As described above, reads that mapped to different genomic locations from the same transcript were merged. Thus, the number of T->C conversions for each transcript could be counted by summing up all the unique positions with a seen T->C conversion over reads that have the same UMI. This required the correct identification of T->C conversions in each position. However, the sequencing errors in high-throughput sequencing could be confounded as T->C conversions. To distinguish the real T->C conversions from sequencing errors, we applied a Bayesian statistics method modeled as the mixture of two binomial distributions as described in the following formula:

$$p_c(k_l; n_l, \varepsilon_s, \rho) = \rho \cdot Binom(k_l; n_l, 1-\varepsilon_s) + (1-\rho) \cdot Binom(n_l-k_l; n_l, \varepsilon_s) \quad (1)$$

where p_c was the posterior probability of T->C conversion at locus l; k_l was the number of reads supporting T->C conversions at locus l; n_l was the total read coverage at locus l; ε_s was the sequencing error rate in high-throughput sequencing, which was approximated as 0.1% (Pfeiffer *et al*, 2018); ρ was the T->C conversion rate in all reads (prior probability), which was approximated by the ratio of observed number of confident T->C conversions (> 50% Ts were converted to Cs at each locus) to the observed number of Ts over all sequenced molecules for each sample. For each sequenced molecule (UMI), the total number of T->C conversion was calculated by summing up the posterior probabilities of T->C conversions over all positions within the molecule.

Due to inefficient 4sU incorporation, not all newly synthesized RNAs can be labeled with 4sU (Herzog *et al*, 2017). To recover the newly synthesized RNAs, which were not labeled, we applied a Bayesian statistics method with a mixture of Poisson and Binomial distributions. The method was described by the following function and terms:

$$p_{new}(k_m; n_m, \rho_l, \varepsilon_c, \theta) = \theta \cdot Poisson(k_m; n_m \cdot \rho_l) + (1-\theta) \cdot Binom(k_m; n_m, \varepsilon_c) \quad (2)$$

where p_{new} was the posterior probability of molecule to be newly synthesized given the T->C conversions we observed in molecule m; k_m was the number of T->C conversions in molecule m; n_m was the total number of Ts in molecule m; ε_c was the conversion error rate (false-positive rate of T->C conversions), which was calculated from the no 4sU labeled cells ($\varepsilon_c = 0.00018$); θ was the newly synthesized molecule fraction (prior probability), which was calculated for each gene in each sample by dividing the number of molecules with T->C conversions by the total number molecules observed; ρ_l was the T->C incorporation rate constant in newly

synthesized molecules, which was approximated by the mean T->C conversion rate (# of T->C divided by the # of Ts) over all precursor molecules in cells that were labeled with 4sU for 24 h. Under constant 4sU incorporation rate assumption, the number of T->C conversions in a newly synthesized molecule with a certain number of Ts should be a Poisson event. We did observe that the distribution of the fraction of T->C conversions recapitulated the density of Poisson distribution with the parameter $\rho_l * \#Ts$ (Appendix Fig S3C). In the above equation, the Poisson distribution modeled the T->C conversion observed in the newly synthesized RNA molecules and the Binomial distribution modeled the T->C conversion in the preexisting RNA molecules. For each gene, the number of newly synthesized molecules was calculated by summing up the posterior probabilities (p_{new}) over all molecules.

Gene expression normalization, smoothing, and correlation analysis

Gene expression was normalized by dividing the raw counts to the total molecules captured in each cell and then multiplied with the scaling factor 10^6 transforming the unit of our reads from absolute counts to counts per million (CPM). To get rid of the unwanted variabilities in single cells for kinetic rates estimation, we smoothed the gene expression profile for each gene along the cell cycle using penalized spline from the Generalized Additive Models (GAM; Wood, 2017) using R 'mgcv' packages (parameters used in the model were: bc = 'cc', k = 20, gamma = 1.4).

To check the gene expression correlation between samples, the normalized gene expression (CPM) was averaged over all cells for each gene and log2 transformed with pseudocount 1. Pair-wise gene expression correlations between different samples were calculated using the common set of genes across samples.

Cell filtering

We discarded cells that contained more than 5% mitochondrial RNA contents and clustered cells using Seurat UMAP algorithm (Macosko et al, 2015). Cells from batch 1 (0, 30, and 60 min 4sU samples) and batch 2 samples (15 min 4sU samples) were pooled and clustered separately. To focus the analysis on cells that were cycling, we measured the averaged expression of well-known cell cycle marker genes and stress-related genes for each cell cluster. The clusters that showed relatively low expression of cycling genes or relatively high expression of stress-related genes were filtered out. We also removed cell clusters that showed relatively higher contents of ribosomal protein coding genes.

In silico cell cycle sorting

The sequenced HEK293 cells were unsynchronized. Thus, it is reasonable to assume that the single cells occupy one cell cycle uniformly. We sorted the filtered cells to a continuous cell cycle progression using the Revelio algorithm (Schwabe et al, 2020). It first identified the highly variable genes using similar method that defined in Seurat package (Macosko et al, 2015). By default, the variable genes are used for PCA. Here, we used the intersection of variable genes and the default variable genes that provided by Revelio package instead since we found it improved the cell cycle sorting. This algorithm then transformed high-dimensional scRNA-seq data of immortalized cell lines into a two-dimensional circular trajectory with approximately uniform cell density in phase space. Due to the

simplicity of the cell cycle signal, we approximated the cell cycle progression by the angular component from this two-dimensional trajectory. This provided us with an ordering of the cells that follows the cell cycle progression. We then based our analysis on the time courses of individual genes that stemmed from this order of cells. With the help of estimated HEK293 cell cycle phase durations (Cheng & Solomon, 2008), we also estimated rough cell cycle phase boundaries and durations.

Time-dependent RNA kinetic rate model

We develop a theoretical framework for estimating RNA kinetic rate based on a typical experimental setup for metabolic labeling. We first introduce a general model and its solutions for the dynamics of different mRNA types that includes the three rate parameters: transcription rate $\alpha(\Phi)$ in $\frac{molecules}{h}$, splicing rate $\beta(\Phi)$ in $\frac{1}{h}$ and degradation rate $\gamma(\Phi)$ in $\frac{1}{h}$, all depending on the cell cycle phase Φ. In contrast to RNA velocity (Manno et al, 2018), all three rates have units associated to them enabling absolute quantification rather than relative quantification. We also allow all three parameters to vary in time, enabling us to model the cell cycle as an example for a biological process.

We have previously introduced the distinction between precursor mRNA p and mature mRNA m and the fact that they can both be quantified by scRNA-seq. The additional classification into labeled and unlabeled mRNA results in a total of four different types of mRNA (since both classifications are independent): unlabeled precursors p_u, unlabeled matures m_u, labeled precursors p_l, and labeled matures m_l. We note that the amounts of unlabeled precursors p_u and labeled precursors p_l sum up to yield the total amount of precursors p. The same is true for the mature mRNAs m. Therefore, the quantities p_u and p_l (as well as m_u and m_l) signify a splitting up of the quantity p (and m) into two summands. This yields four linearly independent observables to infer three kinetic rates such that the problem is no longer under-determined. We now proceed by modeling the concentration levels of the four different mRNA types in a gene-specific manner across the cell population.

In Schwabe et al, 2020, it was observed that the cell cycle of immortalized cell lines is embedded as an approximately circular shape in two dimensions within the normalized gene expression space. For the modeling task at hand, we therefore simplify the cell cycle to a motion on a simple cycle in phase space. All cycles that are topologically equivalent to a circle allow for the definition of a phase variable to uniquely parametrize a position on the cycle. The position of a cell along the cell cycle is then uniquely defined by its cell cycle phase Φ and attains values between 0 and 2π.

The concentration levels of the four mRNA types p_u, m_u, p_l, m_l depend on the position the cell attains within the cell cycle and the duration for which 4sU has been incorporated into the system. We hence choose to model these as functions of two variables: the cell cycle phase Φ with $\Phi \in [0, 2\pi]$ and the labeling time t with $t \in \mathbb{R}_0^+$ and such that $t = 0$ denotes the initiation of labeling. This yields the functions

p_u—unlabeled precursors (preexisting precursors)
m_u—unlabeled matures (preexisting matures)
p_l—labeled precursors (newly synthesized precursors)
m_l—labeled matures (newly synthesized matures)

These four variables are density fields on the phase range $[0, 2\pi]$ changing in time and the two parameters Φ and t are independent. The densities given for $t = 0$ can be considered the initial densities. We obtain the trajectory of an individual cell in phase space if we incorporate the time dependence of its position like

$$\tilde{\Phi}(t) = \tilde{\Phi}_0 + \omega \cdot t \tag{3}$$

Here, $\tilde{\Phi}_0$ is the cell cycle phase the cell was in when labeling was initiated at $t = 0$, and ω is the average velocity with which the cells move along the cell cycle. This is due to the fact that labeling time and cell cycle phase increase simultaneously in a linear way once a particular cell is chosen. For reasons of simplicity, we assume a typical cell cycle to take $T = 19.33$ h (Cheng & Solomon, 2008) and set ω constant by defining $\omega = \frac{2\pi}{T}$. Taking the example of unlabeled precursors p_u, we can then describe the amount of unlabeled precursor mRNA molecules contained within an individual cell throughout the labeling experiment by moving along a curve $c^{p_u}(t)$ such that $c^{p_u}(t) := p_u(\tilde{\Phi}(t), t) = p_u(\tilde{\Phi}_0 + \omega \cdot t, t)$ for $t > 0$.

The parametrization $(\tilde{\Phi}(t), t)$ turns out to correspond to the characteristics of the partial differential equations (PDEs) of interest. This enables us to translate a system of PDEs to a system of ordinary differential equations (ODEs) in the process of finding solutions.

The total amounts of precursor RNA $p(\Phi)$ and the total amounts of mature RNA $m(\Phi)$ are defined by

$$p(\Phi) := p_u(\Phi, t) + p_l(\Phi, t), \tag{4}$$

$$m(\Phi) := m_u(\Phi, t) + m_l(\Phi, t). \tag{5}$$

Furthermore, at the initiation of labeling $(t = 0)$, there is no newly synthesized RNA yet, thus $p_l(\Phi, 0) = m_l(\Phi, 0) = 0$ for all Φ. We conclude $p(\Phi) = p_u(\Phi, 0)$ and $m(\Phi) = m_u(\Phi, 0)$ for all Φ.

The functions p_u, m_u, p_l, m_l describe distributions of molecules over the entire cell cycle. Without loss of generality and to simplify explanations, we concentrate our description on the unlabeled precursors p_u. Its distribution of molecules over the cell cycle is variable with respect to time t. Let t_1 and t_2 be two distinct time points with $0 \le t_1 < t_2$. Two aspects govern the changes of the distribution from time t_1 to time t_2. On the one hand, all cells progress through the cycle. This would mean that the distribution of p_u gets rotated along the cycle since the cells transport the density of p_u with them. This part of the dynamics is termed advection and is given by $-\omega \frac{\partial p_u(\Phi, t)}{\partial \Phi}$, which stems from the basic transport equation.

On the other hand, during the time $\Delta t := t_2 - t_1$, transcriptional changes occur and act on the distribution of p_u. In the case of unlabeled precursors, there is only splicing of already existing p_u molecules occurring since no new unlabeled precursors are created for $t \ge 0$. The transcriptional changes during Δt are hence given by $-\beta(\Phi) \cdot p_u(\Phi, t)$. This allows us to write down a PDE for the dynamics of p_u. It has the structure of a general balance equation for concentration changes, which is

(*local concentration changes*) = (*change due to flux gradients*)
+ (change due to local processes). (6)

We obtain

$$\frac{\partial p_u(\Phi, t)}{\partial t} = -\omega \frac{\partial p_u(\Phi, t)}{\partial \Phi} + (-\beta(\Phi) p_u(\Phi, t)) \tag{7}$$

The initial condition (initial distribution) at time $t = 0$ is given by $p(\Phi)$, since $p_u(\Phi, 0) = p(\Phi)$.

For the other mRNA types, we can argue very similarly and have to only adjust the transcriptional changes. The unlabeled matures m_u gain the number of unlabeled precursors that is spliced during Δt but lose what is degraded of m_u. The labeled precursors p_l are similar to the unlabeled precursors p_u apart from the fact that transcription is a source for newly produced p_l. The dynamics for the labeled matures m_l is analog to the dynamics of m_u.

We can derive a solution for the PDEs with their initial conditions with the help of the method of characteristics (Evans, 2010). We again demonstrate this only for p_u. We parametrize the variables Φ and t with a new variable τ and get

$$\frac{dp_u(\tilde{\Phi}(\tau), \tilde{t}(\tau))}{d\tau} = \frac{\partial p_u(\tilde{\Phi}(\tau), \tilde{t}(\tau))}{\partial \Phi} \cdot \frac{d\tilde{\Phi}(\tau)}{d\tau} + \frac{\partial p_u(\tilde{\Phi}(\tau), \tilde{t}(\tau))}{\partial t} \cdot \frac{d\tilde{t}(\tau)}{d\tau} \tag{8}$$

via the chain rule. We now choose $\tilde{t}(\tau)$ such that

$$\frac{d\tilde{t}(\tau)}{d\tau} = 1 \Rightarrow \tilde{t} = \tau + c_{\tilde{t}} \Rightarrow \text{choose } c_{\tilde{t}} = 0 \Rightarrow \tilde{t}(\tau) = \tau \text{ and rewrite } t = \tau.$$

We conclude that

$$\frac{dp_u(\tilde{\Phi}(t), t)}{dt} = \frac{\partial p_u(\tilde{\Phi}(t), t)}{\partial \Phi} \cdot \frac{d\tilde{\Phi}(t)}{dt} + \frac{\partial p_u(\tilde{\Phi}(t), t)}{\partial t}. \tag{9}$$

Now, we choose $\tilde{\Phi}(t)$ such that

$$\frac{d\tilde{\Phi}(t)}{dt} = \omega \Rightarrow \tilde{\Phi}(t) = \omega t + c_{\tilde{\Phi}} \Rightarrow \text{choose } c_{\tilde{\Phi}} = \tilde{\Phi}_0 \Rightarrow \tilde{\Phi}(t) = \omega t + \tilde{\Phi}_0.$$

This implies

$$\frac{dp_u(\tilde{\Phi}(t), t)}{dt} = \omega \cdot \frac{\partial p_u(\tilde{\Phi}(t), t)}{\partial \Phi} + \frac{\partial p_u(\tilde{\Phi}(t), t)}{\partial t}. \tag{10}$$

If we compare Equation (10) to the initial PDE (7), we can conclude:

$$\frac{dp_u(\tilde{\Phi}(t), t)}{dt} = -\beta(\tilde{\Phi}(t)) p_u(\tilde{\Phi}(t), t). \tag{11}$$

which is an ODE for p_u along a parametrization $(\tilde{\Phi}(t), t)$, the solution of which describes the trajectory of a cell in phase space. The initial condition becomes

$$p_u(\tilde{\Phi}(0), 0) = p(\tilde{\Phi}_0). \tag{12}$$

We note that the parametrization for which the ODE in Equation (11) is satisfied coincides with the parametrization required for

the trajectory of an individual cell (see Equation (3)). That means that the characteristics of the PDE are precisely the individual trajectories of cells.

We also observe that the time derivative along this characteristic is equal to the transcriptional changes from PDE in Equation (7). This enables us without further derivations to write down the ODEs for all other mRNA types along their characteristic curves. The corresponding ODEs yield

$$
\begin{cases}
\dfrac{dp_u(\tilde{\Phi}(t),t)}{dt} = \quad\quad\quad\quad -\beta(\tilde{\Phi}(t))\,p_u(\tilde{\Phi}(t),t), \\[2mm]
\dfrac{dm_u(\tilde{\Phi}(t),t)}{dt} = \beta(\tilde{\Phi}(t))\,p_u(\tilde{\Phi}(t),t) - \gamma(\tilde{\Phi}(t))\,m_u(\tilde{\Phi}(t),t), \\[2mm]
\dfrac{dp_l(\tilde{\Phi}(t),t)}{dt} = \quad \alpha(\tilde{\Phi}(t)) \quad\quad -\beta(\tilde{\Phi}(t))\,p_l(\tilde{\Phi}(t),t), \\[2mm]
\dfrac{dm_l(\tilde{\Phi}(t),t)}{dt} = \beta(\tilde{\Phi}(t))\,p_l(\tilde{\Phi}(t),t) \quad -\gamma(\tilde{\Phi}(t))\,m_l(\tilde{\Phi}(t),t).
\end{cases}
\tag{13}
$$

with initial conditions

$$
\begin{cases}
p_u(\tilde{\Phi}(0),0) = p(\tilde{\Phi}_0), \\[1mm]
m_u(\tilde{\Phi}(0),0) = m(\tilde{\Phi}_0), \\[1mm]
p_l(\tilde{\Phi}(0),0) = 0, \\[1mm]
m_l(\tilde{\Phi}(0),0) = 0.
\end{cases}
\tag{14}
$$

It can be shown that the following solutions indeed solve the system (13) as well as their corresponding PDE counterparts:

$$
\begin{cases}
p_u(\Phi,t) = p(\Phi - \omega t) \cdot e^{-\frac{1}{\omega}\int_{\Phi-\omega t}^{\Phi}\beta(\varphi)d\varphi}, \\[2mm]
m_u(\Phi,t) = p(\Phi-\omega t)\cdot\dfrac{1}{\omega}\int_{\Phi-\omega t}^{\Phi}\left[\beta(\varphi)\cdot e^{-\frac{1}{\omega}\int_{\varphi}^{\Phi}\gamma(\overline{\varphi})d\overline{\varphi}-\frac{1}{\omega}\int_{\Phi-\omega t}^{\varphi}\beta(\overline{\varphi})d\overline{\varphi}}\right]d\varphi \\[2mm]
\quad\quad\quad + m(\Phi-\omega t)\cdot e^{-\frac{1}{\omega}\int_{\Phi-\omega t}^{\Phi}\gamma(\varphi)d\varphi}, \\[2mm]
p_l(\Phi,t) = \dfrac{1}{\omega}\int_{\Phi-\omega t}^{\Phi}\left[\alpha(\varphi)\cdot e^{-\frac{1}{\omega}\int_{\varphi}^{\Phi}\beta(\overline{\varphi})d\overline{\varphi}}\right]d\varphi, \\[2mm]
m_l(\Phi,t) = \dfrac{1}{\omega^2}\int_{\Phi-\omega t}^{\Phi}\left[\beta(\varphi)\cdot e^{-\frac{1}{\omega}\int_{\varphi}^{\Phi}\gamma(\overline{\varphi})d\overline{\varphi}}\cdot\int_{\Phi-\omega t}^{\varphi}\left[\alpha(\overline{\varphi})\cdot e^{-\frac{1}{\omega}\int_{\overline{\varphi}}^{\varphi}\beta(\overline{\overline{\varphi}})d\overline{\overline{\varphi}}}\right]d\overline{\varphi}\right]d\varphi.
\end{cases}
\tag{15}
$$

With respect to the total precursor RNA $p(\Phi)$ and the total mature RNA $m(\Phi)$, we obtain the system

$$
\begin{cases}
\dfrac{dp(\Phi)}{d\Phi} = \quad \dfrac{1}{\omega}\alpha(\Phi) \quad -\dfrac{1}{\omega}\beta(\Phi)p(\Phi), \\[2mm]
\dfrac{dm(\Phi)}{d\Phi} = \dfrac{1}{\omega}\beta(\Phi)p(\Phi) - \dfrac{1}{\omega}\gamma(\Phi)m(\Phi),
\end{cases}
\tag{16}
$$

with initial conditions $p(\Phi_0) =: p_{\Phi_0}$ and $m(\Phi_0) =: m_{\Phi_0}$. It can be shown that a solution for system (16) is given by

$$
\begin{cases}
p(\Phi) = p_{\Phi_0}\cdot e^{-\frac{1}{\omega}\int_{\Phi_0}^{\Phi}\beta(\varphi)d\varphi} \\[2mm]
\quad\quad + \dfrac{1}{\omega}\int_{\Phi_0}^{\Phi}\left[\alpha(\varphi)\cdot e^{-\frac{1}{\omega}\int_{\varphi}^{\Phi}\beta(\overline{\varphi})d\overline{\varphi}}\right]d\varphi, \\[2mm]
m(\Phi) = m_{\Phi_0}\cdot e^{-\frac{1}{\omega}\int_{\Phi_0}^{\Phi}\gamma(\varphi)d\varphi} \\[2mm]
\quad\quad + \dfrac{1}{\omega}\int_{\Phi_0}^{\Phi}\left[\beta(\varphi)\cdot p(\varphi)\cdot e^{-\frac{1}{\omega}\int_{\varphi}^{\Phi}\gamma(\overline{\varphi})d\overline{\varphi}}\right]d\varphi.
\end{cases}
\tag{17}
$$

Solving the inverse problem

The solution formulas from the previous section state how the concentrations of the four mRNA types p_u, m_u, p_l, and m_l can be calculated given dynamic rate parameters $\alpha(\Phi)$, $\beta(\Phi)$, and $\gamma(\Phi)$. However, in reality it is rather of interest to estimate the mRNA kinetic rates that caused the observable profiles for the different mRNA types. This is called an inverse problem.

Typical solutions to inverse problems concerning our model and structure of the previous solutions in system (15) would entail double integrals, which we cannot hope to solve analytically and which would be extremely unstable numerically.

We notice that the system of ODEs (13) contains the rate parameters $\alpha(\Phi)$, $\beta(\Phi)$, and $\gamma(\Phi)$ directly without involving integrals. This suggests a potential strategy for obtaining the rates directly from the dynamics. However, the ODEs describe the trajectory of a single cell throughout the labeling experiment. That means the observable data would have to consist of multiple measurements of the same cell at different labeling times. However, scRNA-seq is not capable of such data.

Rather, we obtain measurements of multiple cells, which roughly cover the range of the cell cycle phase $[0, 2\pi]$ but only for one single labeling time t^*. Hence, the observable data p_u, m_u, p_l, and m_l are functions of the cell cycle phase Φ for fixed labeling times t^*. The numerical derivatives one can calculate from such observations coincide with the partial derivatives of p_u, m_u, p_l, and m_l with respect to cell cycle phase Φ. What remains to be done is to investigate these partial derivatives to find out if they can be utilized to infer rate parameters.

We have calculated these partial derivatives. They are given by

$$
\begin{cases}
\dfrac{\partial p_u(\Phi,t)}{\partial\Phi} = \dfrac{1}{\omega}\alpha(\Phi-\omega t)\dfrac{p_u(\Phi,t)}{p(\Phi-\omega t)} - \dfrac{1}{\omega}\beta(\Phi)p_u(\Phi,t), \\[2mm]
\dfrac{\partial m_u(\Phi,t)}{\partial\Phi} = \dfrac{1}{\omega^2}\alpha(\Phi-\omega t)\cdot\int_{\Phi-\omega t}^{\Phi}\left[\beta(\varphi)e^{-\frac{1}{\omega}\int_{\varphi}^{\Phi}\gamma(\overline{\varphi})d\overline{\varphi}-\frac{1}{\omega}\int_{\Phi-\omega t}^{\varphi}\beta(\overline{\varphi})d\overline{\varphi}}\right]d\varphi \\[2mm]
\quad\quad + \dfrac{1}{\omega}\beta(\Phi)p_u(\Phi,t) - \dfrac{1}{\omega}\gamma(\Phi)m_u(\Phi,t), \\[2mm]
\dfrac{\partial p_l(\Phi,t)}{\partial\Phi} = \dfrac{1}{\omega}\alpha(\Phi) - \dfrac{1}{\omega}\alpha(\Phi-\omega t)\dfrac{p(\Phi)}{p(\Phi-\omega t)} + \dfrac{1}{\omega}\alpha(\Phi-\omega t)\dfrac{p_l(\Phi,t)}{p(\Phi-\omega t)} \\[2mm]
\quad\quad - \dfrac{1}{\omega}\beta(\Phi)p_l(\Phi,t), \\[2mm]
\dfrac{\partial m_l(\Phi,t)}{\partial\Phi} = -\dfrac{1}{\omega^2}\alpha(\Phi-\omega t)\cdot\int_{\Phi-\omega t}^{\Phi}\left[\beta(\varphi)e^{-\frac{1}{\omega}\int_{\varphi}^{\Phi}\gamma(\overline{\varphi})d\overline{\varphi}-\frac{1}{\omega}\int_{\Phi-\omega t}^{\varphi}\beta(\overline{\varphi})d\overline{\varphi}}\right]d\varphi \\[2mm]
\quad\quad + \dfrac{1}{\omega}\beta(\Phi)p_l(\Phi,t) - \dfrac{1}{\omega}\gamma(\Phi)m_l(\Phi,t).
\end{cases}
\tag{18}
$$

Additionally, the concentrations of the total precursor mRNAs p and mature mRNAs m are given by the sum of unlabeled and labeled amounts. The ODEs they obey are:

$$
\begin{cases}
\dfrac{dp(\Phi)}{d\Phi} = \quad \dfrac{1}{\omega}\alpha(\Phi) \quad -\dfrac{1}{\omega}\beta(\Phi)p(\Phi), \\[2mm]
\dfrac{dm(\Phi)}{d\Phi} = \dfrac{1}{\omega}\beta(\Phi)p(\Phi) - \dfrac{1}{\omega}\gamma(\Phi)m(\Phi).
\end{cases}
\tag{19}
$$

Due to the fact that $p_u + p_l = p$ and $m_u + m_l = m$, we still only obtain four linearly independent equations.

Nevertheless, the six equations in (18) and (19) serve as a basis for inferring the rate parameters. The presence of an

integral in the dynamics for m_u and m_l makes these equations unsuitable for obtaining unique solutions of $\alpha(\Phi)$, $\beta(\Phi)$, and $\gamma(\Phi)$. This leaves dynamics of m as the only one involving the dynamics of mature mRNA molecules. Out of the remaining three equations involving precursors, we choose two to form the following system from which we intend to obtain the dynamic rates $\alpha(\Phi)$, $\beta(\Phi)$ and $\gamma(\Phi)$:

$$
\begin{cases}
\dfrac{\partial p_u(\Phi,t)}{\partial \Phi} = \dfrac{1}{\omega}\alpha(\Phi-\omega t)\dfrac{p_u(\Phi,t)}{p(\Phi-\omega t)} - \dfrac{1}{\omega}\beta(\Phi)p_u(\Phi,t), \\
\dfrac{dp(\Phi)}{d\Phi} = \dfrac{1}{\omega}\alpha(\Phi) - \dfrac{1}{\omega}\beta(\Phi)p(\Phi), \\
\dfrac{dm(\Phi)}{d\Phi} = \dfrac{1}{\omega}\beta(\Phi)p(\Phi) - \dfrac{1}{\omega}\gamma(\Phi)m(\Phi).
\end{cases}
\tag{20}
$$

The quantities $p(\Phi)$, $m(\Phi)$, and $p_u(\Phi,t^*)$ for fixed labeling time t^* are all measured within the labeling experiment. Since they cover the entire range of $\Phi\in[0,2\pi]$, we also know $p(\Phi-\omega t^*)$. The derivatives with respect to cell cycle phase can be numerically inferred in theory. Hence, we obtain a linear system (20) consisting of three equations with the three unknowns $\alpha(\Phi)$, $\beta(\Phi)$, and $\gamma(\Phi)$. This system has a unique solution, which is the analytical solution to the inverse problem.

In practice, current noise levels in scRNA-seq experiments cannot be neglected. Our solution to the inverse problem can for now only be seen as a theoretical solution since noise levels in fact would have a huge impact on numerical derivation, making this approach currently infeasible. Due to the presence of the term $\alpha(\Phi-\omega t)$, the above system should be solved by discretization. We would discretize over the entire interval $[0,2\pi]$ in order to solve the coupled system for all discretized time points at once. However, the equations in system (20) suggest a very weak coupling between the discretized variables, resulting in a very sparse parameter matrix. Such weak coupling of the parameter matrix is not capable of solving the issue of high noise terms in numerical derivation. Therefore, we suggest to reconsider this direct approach again once noise levels in scRNA-seq have been substantially reduced making numerical derivation viable. For now, we choose a different approach for estimating the parameters of the mRNA kinetic rates.

General rate estimation for the full model

Instead of solving the inverse problem directly, we now consider the solutions from system (15) in order to infer the rate parameters $\alpha(\Phi)$, $\beta(\Phi)$, and $\gamma(\Phi)$. These analytic solutions involve integrals over the rate parameters, which we want to now remove from the equations. We consider that in case we have access to data with relatively short labeling times $0 < t \ll T = \frac{2\pi}{\omega}$ (much shorter than the duration of the entire cell cycle), the cell cycle phase-dependent changes of $\alpha(\Phi)$, $\beta(\Phi)$, and $\gamma(\Phi)$ in the interval $[\Phi-\omega t, \Phi]$ can be neglected. Then with the help of the mean value theorem, we conclude that approximations of the form

$$
\int_{\Phi-\omega t}^{\Phi} f(y)dy \approx \omega t \cdot f(\Phi)
\tag{21}
$$

hold true for $\alpha(\Phi)$, $\beta(\Phi)$, $\gamma(\Phi)$. The solutions for p_u, m_u, p_l, and m_l from equations (15) can then be approximated and written without integrals:

$$
\begin{cases}
\tilde{p}_u(\Phi,t) = p(\Phi-\omega t)\cdot e^{-t\beta(\Phi)}, \\
\tilde{m}_u(\Phi,t) = \dfrac{\beta(\Phi)p(\Phi-\omega t)}{\gamma(\Phi)-\beta(\Phi)}\cdot e^{-t\beta(\Phi)} + \left(m(\Phi-\omega t) - \dfrac{\beta(\Phi)p(\Phi-\omega t)}{\gamma(\Phi)-\beta(\Phi)}\right)\cdot e^{-t\gamma(\Phi)}, \\
\tilde{p}_l(\Phi,t) = \dfrac{\alpha(\Phi)}{\beta(\Phi)}\cdot\left(1 - e^{-t\beta(\Phi)}\right), \\
\tilde{m}_l(\Phi,t) = \dfrac{\alpha(\Phi)}{\gamma(\Phi)} - \dfrac{\alpha(\Phi)}{\gamma(\Phi)-\beta(\Phi)}\cdot e^{-t\beta(\Phi)} + \dfrac{\alpha(\Phi)\beta(\Phi)}{\gamma(\Phi)(\gamma(\Phi)-\beta(\Phi))}\cdot e^{-t\gamma(\Phi)}.
\end{cases}
\tag{22}
$$

During the experiment, we consider the short-fixed labeling time $t^* = 0.25\text{h} \ll 19.33\text{h}$. We measure $p_u(\Phi,t^*)$, $m_u(\Phi,t^*)$, $p_l(\Phi,t^*)$, and $m_l(\Phi,t^*)$ for all $\Phi\in[0,2\pi]$, and thus implicitly $p(\Phi-\omega t^*)$ and $m(\Phi-\omega t^*)$. Hence, we can solve the approximated solutions for $\alpha(\Phi)$ and $\beta(\Phi)$ directly:

$$
\begin{cases}
\beta(\Phi) = -\dfrac{1}{t}\log\left(\dfrac{\tilde{p}_u(\Phi,t^*)}{p(\Phi-\omega t^*)}\right), \\
\alpha(\Phi) = \dfrac{p(\Phi-\omega t^*)\cdot\beta(\Phi)\cdot\tilde{p}_l(\Phi,t^*)}{p(\Phi-\omega t^*)-\tilde{p}_u(\Phi,t^*)}.
\end{cases}
\tag{23}
$$

A general solution for $\gamma(\Phi)$ involves solving the quadratic equation

$$
\begin{aligned}
0 = \gamma^2(\Phi) \\
-\beta(\phi)\cdot\left(\dfrac{\tilde{p}_l(\Phi,t^*)}{\tilde{m}_l(\Phi,t^*)} + \dfrac{p(\Phi-\omega t^*)}{m(\Phi-\omega t^*)} + 1\right)\cdot\gamma(\Phi) \\
+\beta^2(\Phi)\cdot\dfrac{\tilde{p}_l(\Phi,t^*)}{\tilde{m}_l(\Phi,t^*)}\cdot\dfrac{p(\Phi-\omega t^*)}{m(\Phi-\omega t^*)}\cdot\left(\dfrac{m(\Phi-\omega t^*)-\tilde{m}_u(\Phi,t^*)}{p(\Phi-\omega t^*)-\tilde{p}_u(\Phi,t^*)}\right).
\end{aligned}
\tag{24}
$$

Its solution is a basic application of the quadratic solution formula:

$$
\begin{aligned}
\gamma_{1,2}(\Phi) = \dfrac{\beta(\Phi)}{2}\cdot\left(\dfrac{\tilde{p}_l(\Phi,t^*)}{\tilde{m}_l(\Phi,t^*)} + \dfrac{p(\Phi-\omega t^*)}{m(\Phi-\omega t^*)} + 1\right) \\
\pm\dfrac{\beta(\Phi)}{2}\ \sqrt{\left(\dfrac{\tilde{p}_l(\Phi,t^*)}{\tilde{m}_l(\Phi,t^*)} + \dfrac{p(\Phi-\omega t^*)}{m(\Phi-\omega t^*)} + 1\right)^2 - 4\dfrac{\tilde{p}_l(\Phi,t^*)}{\tilde{m}_l(\Phi,t^*)}\dfrac{p(\Phi-\omega t^*)}{m(\Phi-\omega t^*)}\left(\dfrac{m(\Phi-\omega t^*)-\tilde{m}_u(\Phi,t^*)}{p(\Phi-\omega t^*)-\tilde{p}_u(\Phi,t^*)} + 1\right)}
\end{aligned}
\tag{25}
$$

We have to again consider the issue of noise in scRNA-seq data when applying these analytic solutions to currently available experimental data. We observe that the solutions for $\alpha(\Phi)$ and $\gamma(\Phi)$ in Equations (23), (25) depend on the splicing rate $\beta(\Phi)$. The solution for $\beta(\Phi)$ itself consists of the logarithm of the fraction between unlabeled precursors p_u at the time of measurement Φ and the total precursors p at the time of labeling initiation $(\Phi-\omega t)$. While the mature gene counts in RNA-seq data are known to be subjected to large amounts of both technical and biological noise, the detection rate for precursors is even an order of magnitude lower. We therefore expect $\beta(\Phi)$ to be extremely sensitive given the formula in

Equation (23). Due to the dependencies, any noise will be propagated to $\alpha(\Phi)$ and $\gamma(\Phi)$.

Steady-state estimations of the splicing rate β also do not yield reliable results. In preliminary data from bulk RNA-seq involving metabolic labeling, we found a huge dynamic range for the β-estimates, spanning many orders of magnitude. Since there is no ground truth data, it is difficult to estimate the reliability of such data but due to the low data coverage, we are careful to trust such estimates. While our model might be useful for future experimental data with reduced noise levels, we now incorporate additional simplifications into the model in order to actually estimate a transcription rate $\alpha(\Phi)$ and a degradation rate $\gamma(\Phi)$ from currently available experimental data.

Since we introduce simplifications to our model in the next section, we from now on refer to the approximated rate parameters in Equations (23), (25) presented in this section as the estimates of the full model. The approximated rate parameters from the next section are then be referred to as the estimates of the simplified model.

Rate estimation for a simplified model

The splicing of transcripts (typically minutes) acts on a different time scale than the degradation (typically hours; Alpert *et al*, 2017). We therefore assume that $\gamma(\Phi) \ll \beta(\Phi)$ and can also conclude $\frac{e^{-t\beta(\Phi)}}{\beta(\Phi)} \ll \frac{e^{-t\gamma(\Phi)}}{\gamma(\Phi)}$. The approximation $e^{-t\beta(\Phi)} \ll e^{-t\gamma(\Phi)}$ introduces a larger error than the previous two simplifications but we find it reasonable in our setting as it severely simplifies the solution for $\tilde{m}_u(\Phi, t)$. Incorporating these simplifications, we obtain:

$$
\begin{cases}
\check{p}_u(\Phi, t) = p(\Phi - \omega t) \cdot e^{-t\beta(\Phi)}, \\
\check{m}_u(\Phi, t) = (m(\Phi - \omega t) + p(\Phi - \omega t)) \cdot e^{-t\gamma(\Phi)}, \\
\check{p}_l(\Phi, t) = \frac{\alpha(\Phi)}{\beta(\Phi)} \cdot \left(1 - e^{-t\beta(\Phi)}\right), \\
\check{m}_l(\Phi, t) = \frac{\alpha(\Phi)}{\gamma(\Phi)} \cdot \left(1 - e^{-t\gamma(\Phi)}\right).
\end{cases}
\tag{26}
$$

We note that the solutions concerning the mature RNAs are very similar to the solutions of the precursors, with the splicing rate $\beta(\Phi)$ and the degradation rate $\gamma(\Phi)$ interchanged. This is due to the fact that with our basic simplification that splicing is much faster than degradation, we are close to removing precursors from our system of equations and considering the system

$$
\begin{cases}
\dfrac{dm_u(\tilde{\Phi}(t), t)}{dt} = -\gamma(\tilde{\Phi}(t)) \, m_u(\tilde{\Phi}(t), t), \\
\dfrac{dm_l(\tilde{\Phi}(t), t)}{dt} = \alpha(\tilde{\Phi}(t)) - \gamma(\tilde{\Phi}(t)) \, m_l(\tilde{\Phi}(t), t).
\end{cases}
\tag{27}
$$

Since we find in our experimental data that the fraction of reads associated with precursors is on average less than 2% (Appendix Fig S4B), we do drop the precursors from the calculation of $\tilde{m}_u(\Phi, t)$ and obtain the final approximations

$$
\begin{cases}
\overline{m}_u(\Phi, t) = m(\Phi - \omega t) \cdot e^{-t\gamma(\Phi)}, \\
\overline{m}_l(\Phi, t) = \frac{\alpha(\Phi)}{\gamma(\Phi)} \cdot \left(1 - e^{-t\gamma(\Phi)}\right).
\end{cases}
\tag{28}
$$

We remark that starting from the simplified model (27) without precursors altogether leads to the exact same approximated

solutions now mentioned and is a convenient verification for the validity of our complex solution (15) and the subsequent simplification steps. The approximations (28) can be solved for $\alpha(\Phi)$ and $\gamma(\Phi)$:

$$
\begin{cases}
\gamma(\Phi) = -\dfrac{1}{t} \log\left(\dfrac{\overline{m}_u(\Phi, t)}{m(\Phi - \omega t)}\right), \\
\alpha(\Phi) = -\dfrac{1}{t} \cdot \dfrac{m(\Phi - \omega t) \cdot \overline{m}_l(\Phi, t)}{m(\Phi - \omega t) - \overline{m}_u(\Phi, t)} \cdot \log\left(\dfrac{\overline{m}_u(\Phi, t)}{m(\Phi - \omega t)}\right).
\end{cases}
\tag{29}
$$

We have so far always done approximations on the side of the RNA concentration levels and utilized the true parameters $\alpha(\Phi)$ and $\gamma(\Phi)$. However, we do not know $\overline{m}_u$ or $\overline{m}_l$ but we have observations for m_u and m_l and want to take advantage of these values. Hence, we utilize the structure of (29) and define approximations $\widehat{\alpha}(\Phi)$, $\widehat{\gamma}(\Phi)$:

$$
\begin{cases}
\widehat{\gamma}(\Phi) = -\dfrac{1}{t} \log\left(\dfrac{m_u(\Phi, t)}{m(\Phi - \omega t)}\right), \\
\widehat{\alpha}(\Phi) = -\dfrac{1}{t} \cdot \dfrac{m(\Phi - \omega t) \cdot m_l(\Phi, t)}{m(\Phi - \omega t) - m_u(\Phi, t)} \cdot \log\left(\dfrac{m_u(\Phi, t)}{m(\Phi - \omega t)}\right).
\end{cases}
\tag{30}
$$

The approximations in (30) are the formulas implemented to obtain the results from the main text.

In order to get predictions from the calculated parameters, we utilize the exact solution formula to the simplified model (27) where precursors are removed. This leads us to the following prediction formulas:

$$
\begin{cases}
\widehat{m}_u(\Phi, t) = m(\Phi - \omega t) \cdot e^{-\frac{1}{\omega}\int_{\Phi - \omega t}^{\Phi} \widehat{\gamma}(\varphi) d\varphi}, \\
\widehat{m}_l(\Phi, t) = \dfrac{1}{\omega} \int_{\Phi - \omega t}^{\Phi} \left[\widehat{\alpha}(\varphi) \cdot e^{-\frac{1}{\omega}\int_{\varphi}^{\Phi} \widehat{\gamma}(\overline{\varphi}) d\overline{\varphi}}\right] d\varphi.
\end{cases}
\tag{31}
$$

We will later compare the differences between the observation m_u and the prediction $\widehat{m}_u$, as well as between m_l and $\widehat{m}_l$ and between their sums $m = m_u + m_l$ and $\widehat{m} := \widehat{m}_u + \widehat{m}_l$.

Simulating synthetic data

In experimental data the task is typically to calculate rate parameters from observed gene expression time courses. While the previously derived formulas enable us to calculate estimates for these rates, one drawback is that the ground truth is unknown. We therefore, as a proof-of-concept, show that given a ground truth for $\alpha(\Phi)$, $\beta(\Phi)$, and $\gamma(\Phi)$, which generate certain p_u, m_u, p_l, m_l, p, and m with the help of the formulas (15), we get reasonable rate parameters with our formulas for the full model from Equations (23), (25) and for the simplified model from Equation (30).

One issue in the formulas (17) is that we still need to choose initial values $p(\Phi_0)$ and $m(\Phi_0)$. This could for example be done via the steady-state solutions $p(0) = \frac{\alpha(0)}{\beta(0)}$ and $m(0) = \frac{\alpha(0)}{\gamma(0)}$. However, after generating a time course over one cell cycle, we often observe that $p(0) \neq p(2\pi)$ and $m(0) \neq m(2\pi)$. Therefore, we continuously generate time courses for $p(\Phi)$ and $m(\Phi)$ over multiple periods (multiple cell cycles). We notice that the simulation converges after very few (fewer than 10) cycles such that $|p(\Phi) - p(\Phi + T)| < \varepsilon_p$ and $|m(\Phi) - m(\Phi + T)| < \varepsilon_m$. Hence, after our convergence criteria are met, we obtain the corresponding time courses $p(\Phi)$ and $m(\Phi)$ to a given input $\alpha(\Phi)$, $\beta(\Phi)$, $\gamma(\Phi)$. During the simulation of p_u, m_u, we

then choose initial values $p(\Phi - \omega t)$ and $m(\Phi - \omega t)$ (see system (15)) from these converged time courses making them close to the true initial values and non-arbitrary.

As a proof-of-concept for our model, we choose arbitrary time courses for $\alpha(\Phi)$, $\beta(\Phi)$, $\gamma(\Phi)$. With the help of the solution formulas from system (15), we then generate time courses for the different mRNA types given the dynamic parameters. Afterward, we plug these time courses into the formulas (23), (25) and then compare the resulting estimate for the full model to the ground truth. One example resulting from this strategy is depicted in Appendix Fig S5. We observe that the estimates for the rate parameters from our approximations (23), (25) of the full model recapitulate the ground truth very well.

We now choose a specific shape (such as a simple sine curve) that the time course $m(\Phi)$ should attain. We can then get different sets of parameters $\alpha(\Phi)$ and $\gamma(\Phi)$ that generate this time course. First, we set $\beta(\Phi) = \beta$ constant for the remainder of this section and assume that the parameter is known. Then, we consider the following three cases:

- Case I where γ is a constant and known parameter,
- Case II where α is a constant and known parameter,
- Case III where γ is a non-constant but known parameter.

Case I: We know from (16)

$$m'(\Phi) = \frac{1}{\omega}\beta p(\Phi) - \frac{1}{\omega}\gamma m(\Phi) \Leftrightarrow p(\Phi) = \frac{\gamma}{\beta}m(\Phi) + \omega\frac{1}{\beta}m'(\Phi). \tag{32}$$

We also have in (16)

$$\begin{aligned} p'(\Phi) &= \frac{1}{\omega}\alpha(\Phi) - \frac{1}{\omega}\beta p(\Phi) \\ &\Leftrightarrow \alpha(\Phi) = \omega p'(\Phi) + \beta p(\Phi) \\ &\Rightarrow \alpha(\Phi) = \gamma m(\Phi) + \omega\left(1 + \frac{\gamma}{\beta}\right)m'(\Phi) + \omega^2\frac{1}{\beta}m''(\Phi). \end{aligned} \tag{33}$$

Hence, choosing this alpha will generate the desired time course $m(\Phi)$.

Case II: When α and β are constant, we can find a simple solution for $p(\Phi)$ from (16):

$$p'(\Phi) = \frac{1}{\omega}\alpha - \frac{1}{\omega}\gamma p(\Phi) \Rightarrow p(\Phi) = \frac{\alpha}{\beta} + \left(p_0 - \frac{\alpha}{\beta}\right)e^{-\frac{1}{\omega}\beta\Phi}. \tag{34}$$

Coupled with the ODE for the total mature counts in (16), we obtain:

$$\begin{aligned} m'(\Phi) &= \frac{1}{\omega}\beta p(\Phi) - \frac{1}{\omega}\gamma(\Phi)m(\Phi) \\ &\Leftrightarrow \gamma(\Phi) = \beta\frac{p(\Phi)}{m(\Phi)} - \omega\frac{m'(\Phi)}{m(\Phi)} \\ &\Rightarrow \gamma(\Phi) = \frac{\alpha}{m} + \left(\beta\frac{p_0}{m} - \frac{\alpha}{m}\right)e^{-\frac{1}{\omega}\beta\Phi} - \omega\frac{m'(\Phi)}{m(\Phi)}. \end{aligned} \tag{35}$$

For an appropriately chosen p_0, the calculated $\gamma(\Phi)$ will then generate the desired $m(\Phi)$.

Case III: We can extend case I and instead of choosing constant γ, we can choose any differentiable $\gamma(\Phi)$. Then from (16)

$$\begin{aligned} m'(\Phi) &= \frac{1}{\omega}\beta p(\Phi) - \frac{1}{\omega}\gamma(\Phi)m(\Phi) \Leftrightarrow p(\Phi) \\ &= \frac{1}{\beta}\gamma(\Phi)m(\Phi) + \omega\frac{1}{\beta}m'(\Phi). \end{aligned} \tag{36}$$

Analogously to previous considerations, we obtain

$$\begin{aligned} \alpha(\Phi) &= \left(\gamma(\Phi) + \omega\frac{1}{\beta}\gamma'(\Phi)\right)m(\Phi) + \omega\left(1 + \frac{1}{\beta}\gamma(\Phi)\right)m'(\Phi) \\ &+ \omega^2\frac{1}{\beta}m''(\Phi). \end{aligned} \tag{37}$$

We now conclude that if we utilize the same $m(\Phi)$ in all three of these cases, we will have produced the same expression profile $m(\Phi)$, which is generated by three different types of regulation. We can show that our calculated parameter solutions from the previous sections can recapitulate these differences when we only provide the input p_u, m_u, p_l, m_l, p, and m (Appendix Fig S6A).

SLAM-Drop-seq data: Peak identification for gene expression and kinetic rates

To investigate the variability of gene expression and kinetic rates over the cell cycle, we implemented the peaking calling method to identify peak patterns of the profiles over cell cycle time. Since we smoothed the profiles by penalized splines, we can investigate their derivative in order to calculate the locations of local extrema. We defined the global fold change (*fc*) and the global difference (*diff*) of a profile as:

$$fc(global) = \frac{\max(global)}{\min(global)} \tag{38}$$

$$diff(global) = \max(global) - \min(global) \tag{39}$$

If *fc* (*global*) of the profile was larger or equal to 1.5, it was considered as a peaking gene or peaking rate. For these peaking genes and peaking rates, we then checked all the remaining local maximums except the global maximum. If the difference between the local maximum and the larger value of the nearby local minimum was at least 1/4 of the *diff* (*global*), the local maximum was identified as a peak. The total number of peaks was counted by summing up all identified peaks in each profile. Genes with a single expression peak were identified as cycling genes.

Deviation calculation

To check how close the prediction (*pred*; gene expression that was calculated from calculated transcription and degradation rates) to the observation (*obs*; gene expression that was quantified from the sequencing data) is. We defined the mean absolute deviation (*dev*) from prediction to observation over all cells for each gene using the following equation where m is the cell number:

$$dev = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{|pred_i - obsv_i|}{obsv_i} \tag{40}$$

To identify well-predicted genes that showed close prediction to observation, we checked the deviation values for all mature RNA types (labeled, unlabeled, and the total) and took the maximum value as the deviation for each gene.

To check how much prediction changed when transcription or degradation rate was set to a constant value, we calculated the deviation (dev_c) from predictions that were calculated from a constant transcription or a constant degradation rate ($pred$ ($constant\ rate$)) to the original predictions that were calculated from the dynamic model ($pred$) using the following equation where m is the number of cells across the cell cycle:

$$dev_c = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{|pred(constant\ rate)_i - pred_i|}{pred_i} \quad (41)$$

The maximum of dev_c for all mature RNA types (labeled, unlabeled and the total) was taken as the deviation for each gene.

Comparison of scEU-seq cell cycle–related data to SLAM-Drop-seq data

The gene expression and kinetic rates of RPE1-FUCCI cells obtained using the scEU-seq method are taken from the Supplementary Table S1 (Dataset from: Battich *et al*, 2020). The cell cycle procession time sorted gene expression (z-score normalized), transcription rate and degradation rates (normalized to mean rate and log transformed) data are directly taken from the Supplementary Table S1 (Battich *et al*, 2020). To compare the kinetic rates and gene expression profiles between SLAM-Drop-seq and scEU-seq, we took the intersection of cell cycle variable genes between the two datasets. The cell cycle time adjacent HEK293 cells from SLAM-Drop-seq were grouped into 301 bins to match the number of bins in the scEU-seq data. Gene expression values and the kinetic rates were averaged in each bin for SLAM-Drop-seq data and normalized using the same normalization methods used in scEU-seq data. The 98 common cell cycle variable genes between scEU-seq and SLAM-Drop-seq data were compared using heatmaps (Fig EV2A–C).

Comparison of mRNA kinetic rates calculated from sci-fate data to SLAM-Drop-seq data

The 4sU labeled dexamethasone (DEX) treated control (DEX 0 h) A549 samples from sci-fate method (Cao *et al*, 2020) were used in this study. The A549 cells were labeled with 4sU for 2 h followed by IAA alkylation, which led to T->C transitions in the sequencing reads. Similar to SLAM-Drop-seq, both new and old RNAs can be identified in single cells simultaneously by identifying labeled RNAs based on T->C transitions in the reads.

The demultiplexed sci-fate raw sequencing data of A549 cells were obtained directly from the authors (Cao *et al*, 2020). To obtain gene expression count matrices from the sequencing reads, the open-source sci-fate computational pipeline (https://github.com/JunyueC/sci-fate_analysis) for generating full expression and new expression gene counts were used. Same as the sci-fate methods, the public ENCODE (ENCODE Project Consortium, 2004) A549 bulk RNA-seq datasets (sample name: ENCFF542FVG, ENCFF538ZTA, ENCFF214JEZ, ENCFF629LOL, ENCFF149CJD, ENCFF006WNO, ENCFF828WTU, ENCFF380VGD) were used for SNP calling in generating gene expression matrix for new RNAs. In the end, gene counts for every gene in every cell were obtained for both total and new RNAs. We then grouped intronic and exonic reads and generated four RNA count matrices (i.e., labeled_mature, unlabeled_mature, labeled_precursor and unlbaled_precursor).

To sort the cells along the cell cycle, Reveilo was applied to the total (i.e., the sum of new and old) RNA count data of A549 cells. Prior to the cell cycle sorting, Seurat (Macosko *et al*, 2015) was used to cluster the cells and a sub cluster of cells that were not represented by any cell cycle marker genes were removed from the cell population. The remaining cells were ordered and assigned with cell cycle pseudo times using Revelio.

The classified gene expression count matrices (i.e., labeled_mature, unlabeled_mature, labeled_precursor and unlbaled_precursor) and the obtained cell cycle pseudo times were used to calculate transcription rates and degradation rates using Eskrate. Due to the sparseness of precursor counts in single cells, we also applied the simplified kinetic model to calculate the transcription and degradation rates. The calculated kinetic rates and the observed mRNA expression levels in A549 cells were then compared to the SLAM-Drop-seq data for the 140 shared cell cycle variable genes. To compare them gene-by-gene using heatmaps, the adjacent HEK293 cells were grouped to have equal number of cells as the sci-fate data along the cell cycle (Fig EV2D–F). RNA expression levels and kinetic rates were normalized as data shown in Fig EV2A–D.

Mean RNA half-lives calculation

RNA half-lives were determined by RNA decay rates. To calculate the mean RNA half-lives, we averaged the calculated time-dependent RNA degradation rates over cells for each gene and then calculated the mean RNA half-lives $t_{1/2}$ from the mean degradation rates ($\bar{\gamma}$) with the following equation:

$$t_{1/2} = \frac{ln(2)}{\bar{\gamma}} \quad (42)$$

Downsampling analysis to identify valid genes for rate estimation

Dropout event is one of the biggest challenges in scRNA-seq data analysis as it leads to a large proportion of zeros in the gene expression matrix. The cell cycle time-dependent transcription and degradation rates are dependent on the temporal profiles of labeled, unlabeled, and total mature RNAs (Equation (25) in Section 2.7.10). Thus, we need to consider the dropout rates of all three types of RNAs along the cell cycle progression time. Since scRNA-seq data are noisy, the normalized gene expression data were smoothed before plugging them into the simplified analytical solutions to calculate the kinetic rates. To know which genes are valid for the rate estimation given the dropout rates, we need to check if the smoothed profiles are valid given the dropout rates. Since the dropout rate of the labeled mature RNAs are the highest, we downsampled the total mature RNA gene expression matrix of a list of selected genes to higher dropout rates that can represent the dropout rates of labeled mature RNAs. The selected genes composed both positive controls genes (i.e., known cell cycle marker genes, Whitfield *et al*, 2002) and negative control genes (i.e., known housekeeping genes for HEK293 cells, Hounkpe *et al*, 2021). They were down sampled to a series of fractions (i.e., 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1) using 'DropletUtils' (Griffiths *et al*, 2018; Lun *et al*, 2019). The downsampling was iterated 100 times for every gene at each downsampling fraction. Afterward, downsampled gene expression data was normalized and smoothed. Based on the fold changes in the smoothed profiles along the cell cycle, these genes were classified as marker genes if at least 95% of the iterative

profiles showed a minimum fold change of 1.5. Otherwise, they were classified as housekeeping genes.

We binned genes to different groups based on their original fold changes (i.e., 0.5, 0.6, 0.7, 0.8, 0.9 and 1) in gene expression of the total mature RNAs. We counted the number of true positives, true negatives, false positives, and false negatives following the rules that shown in the following table:

		Original gene expression profiles	
		Marker	**Housekeeping**
Downsampled gene expression profiles	Marker	True Positive (TP)	False Positive (FP)
	Housekeeping	False Negative (FN)	True Negative (TN)

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{(TP + FN)} \qquad (43)$$

The confusion matrices were generated for marker genes in different groups based on their minimal fold changes. The true-positive rates (Equation 43) were calculated and plotted against the dropout rate cutoffs (Appendix Fig S7D). Genes were expected to be valid for rate estimation if it had at least 95% true positives given its dropout rates.

Data availability

Single cell RNA sequencing: Gene Expression Omnibus GSE197667 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197667). Visualization of cycling gene expression and kinetic rates: MDC (https://shiny.mdc-berlin.de/slam_drop_seq/). Data processing code: GitHub (https://github.com/rajewsky-lab/SLAM-Drop-seq). Eskrate R package: GitHub (https://github.com/rajewsky-lab/Eskrate).

Expanded View for this article is available online.

Author contributions

Markus Landthaler: Conceptualization; funding acquisition; project administration; writing – review and editing. **Haiyue Liu:** Software; formal analysis; investigation; methodology; writing – original draft; writing – review and editing. **Roberto Arsiè:** Methodology; writing – original draft; writing – review and editing. **Daniel Schwabe:** Software; investigation; methodology; writing – review and editing. **Marcel Schilling:** Formal analysis; methodology; writing – review and editing. **Igor Minia:** Methodology. **Jonathan Alles:** Methodology. **Anastasiya Boltengagen:** Methodology. **Christine Kocks:** Methodology. **Martin Falcke:** Formal analysis; supervision; writing – review and editing. **Nir Friedman:** Formal analysis. **Nikolaus Rajewsky:** Conceptualization; funding acquisition; project administration; writing – review and editing.

Disclosure and competing interests statement

The authors declare that they have no conflicts of interest. NR is an editorial advisory board member. This has no bearing on the editorial consideration of this article for publication.

References

Alles J, Karaiskos N, Praktiknjo SD, Grosswendt S, Wahle P, Ruffault P-L, Ayoub S, Schreyer L, Boltengagen A, Birchmeier C *et al* (2017) Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol* 15: 44

Alpert T, Herzel L, Neugebauer KM (2017) Perfect timing: splicing and transcription rates in living cells: Splicing and transcription rates in living cells. *WIREs RNA* 8: e1401

Battich N, Beumer J, de Barbanson B, Krenning L, Baron CS, Tanenbaum ME, Clevers H, van Oudenaarden A (2020) Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. *Science* 367: 1151–1156

Blackinton JG, Keene JD (2014) Post-transcriptional RNA regulons affecting cell cycle and proliferation. *Semin Cell Dev Biol* 34: 44–54

Cao J, Zhou W, Steemers F, Trapnell C, Shendure J (2020) Sci-fate characterizes the dynamics of gene expression in single cells. *Nat Biotechnol* 38: 980–988

Carleton M, Mao M, Biery M, Warrener P, Kim S, Buser C, Marshall CG, Fernandes C, Annis J, Linsley PS (2006) RNA interference-mediated silencing of mitotic kinesin KIF14 disrupts cell cycle progression and induces cytokinesis failure. *Mol Cell Biol* 26: 3853–3863

Cheng A, Solomon MJ (2008) Speedy/Ringo C regulates S and G_2 phase progression in human cells. *Cell Cycle* 7: 3037–3047

Cho RJ, Huang M, Campbell MJ, Dong H, Steinmetz L, Sapinoso L, Hampton G, Elledge SJ, Davis RW, Lockhart DJ (2001) Transcriptional regulation and function during the human cell cycle. *Nat Genet* 27: 48–54

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

Dölken L, Ruzsics Z, Rädle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P *et al* (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* 14: 1959–1972

Dynlacht BD (1997) Regulation of transcription by proteins that control the cell cycle. *Nature* 389: 149–152

ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306: 636–640

Erhard F, Baptista MAP, Krammer T, Hennig T, Lange M, Arampatzi P, Jürges CS, Theis FJ, Saliba A-E, Dölken L (2019) scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* 571: 419–423

Erhard F, Saliba A-E, Lusser A, Toussaint C, Hennig T, Prusty BK, Kirschenbaum D, Abadie K, Miska EA, Friedel CC *et al* (2022) Time-resolved single-cell RNA-seq using metabolic RNA labelling. *Nat Rev Methods Primers* 2: 77

Eser P, Demel C, Maier KC, Schwalb B, Pirkl N, Martin DE, Cramer P, Tresch A (2014) Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. *Mol Syst Biol* 10: 717

Evans LC (2010) *Partial differential equations: Graduate studies in mathematics*, p 749. Providence, RI: American Mathematical Society

Griffiths JA, Richard AC, Bach K, Lun ATL, Marioni JC (2018) Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat Commun* 9: 2667

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646–674

Harris ME, Böhni R, Schneiderman MH, Ramamurthy L, Schümperli D, Marzluff WF (1991) Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps. *Mol Cell Biol* 11: 2416–2424

Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* 2: 666–673

Hendriks G-J, Jung LA, Larsson AJM, Lidschreiber M, Forsman OA, Lidschreiber K, Cramer P, Sandberg R (2019) NASC-seq monitors RNA synthesis in single cells. *Nat Commun* 10: 3138

Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, Wlotzka W, von Haeseler A, Zuber J, Ameres SL (2017) Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* 14: 1198–1204

Hounkpe BW, Chenou F, de Lima F, De Paula EV (2021) HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Res* 49: D947–D955

Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J, Boguski MS *et al* (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* 283: 83–87

Jürges C, Dölken L, Erhard F (2018) Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* 34: i218–i226

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993

Liu Y, Chen S, Wang S, Soares F, Fischer M, Meng F, Du Z, Lin C, Meyer C, DeCaprio JA *et al* (2017) Transcriptional landscape of the human cell cycle. *Proc Natl Acad Sci USA* 114: 3473–3478

Lun ATL, Participants in the 1st Human Cell Atlas Jamboree, Riesenfeld S, Andrews T, Dao TP, Gomes T, Marioni JC (2019) EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol* 20: 63

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM *et al* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161: 1202–1214

Manno GL, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A *et al* (2018) RNA velocity of single cells. *Nature* 560: 494–498

Matthews HK, Bertoli C, de Bruin RAM (2022) Cell cycle control in cancer. *Nat Rev Mol Cell Biol* 23: 74–88

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–1303

Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dölken L *et al* (2011) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* 7: 458

Murakawa Y, Hinz M, Mothes J, Schuetz A, Uhl M, Wyler E, Yasuda T, Mastrobuoni G, Friedel CC, Dölken L *et al* (2015) RC3H1 post-transcriptionally regulates A20 mRNA and modulates the activity of the IKK/NF-κB pathway. *Nat Commun* 6: 7367

Pepe-Mooney BJ, Dill MT, Alemany A, Ordovas-Montanes J, Matsushita Y, Rao A, Sen A, Miyazaki M, Anakk S, Dawson PA *et al* (2019) Single-cell analysis of the liver epithelium reveals dynamic heterogeneity and an essential role for YAP in homeostasis and regeneration. *Cell Stem Cell* 25: 23–38.e8

Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL, Mayer G (2018) Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci Rep* 8: 10950

Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* 10: 1096–1098

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9: 171–181

Qiu P (2020) Embracing the dropouts in single-cell RNA-seq analysis. *Nat Commun* 11: 1169

Qiu Q, Hu P, Qiu X, Govek KW, Cámara PG, Wu H (2020) Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. *Nat Methods* 17: 991–1001

Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, Pogson AN, Hein MY, Min KH, Wang L *et al* (2022) Mapping transcriptomic vector fields of single cells. *Cell* 185: 690–711.e45

Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N *et al* (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* 29: 436–442

Riml C, Amort T, Rieder D, Gasser C, Lusser A, Micura R (2017) Osmium-mediated transformation of 4-thiouridine to cytidine as key to study RNA dynamics by sequencing. *Angew Chem Int Ed* 56: 13479–13483

Rubin SM, Sage J, Skotheim JM (2020) Integrating old and new paradigms of G1/S control. *Mol Cell* 80: 183–192

Schafer KA (1998) The cell cycle: a review. *Vet Pathol* 35: 461–478

Schofield JA, Duffy EE, Kiefer L, Sullivan MC, Simon MD (2018) TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat Methods* 15: 221–225

Schwabe D, Formichetti S, Junker JP, Falcke M, Rajewsky N (2020) The transcriptome dynamics of single cells during the cell cycle. *Mol Syst Biol* 16: e9946

Skinner SO, Xu H, Nagarkar-Jaiswal S, Freire PR, Zwaka TP, Golding I (2016) Single-cell analysis of transcription kinetics across the cell cycle. *Elife* 5: e12175

Wang J, Chen W, Yue W, Hou W, Rao F, Zhong H, Qi Y, Hong N, Ni T, Jin W (2022) Comprehensive mapping of alternative polyadenylation site usage and its dynamics at single-cell resolution. *Proc Natl Acad Sci USA* 119: e2113504119

Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO *et al* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* 13: 1977–2000

Wood SN (2017) *Generalized additive models*. Boca Raton, FL: Chapman and Hall/CRC

Zeisel A, Köstler WJ, Molotski N, Tsai JM, Krauthgamer R, Jacob-Hirsch J, Rechavi G, Soen Y, Jung S, Yarden Y *et al* (2011) Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol Syst Biol* 7: 529