Unraveling the oscillatory dynamics of mRNA metabolism and chromatin accessibility during the cell cycle through integration of single-cell multiomic data

Maulik K. Nariya[1, 2, 3, 4] ✉, David Santiago-Algarra[1, 2, 3, 4], Olivier Tassy[1, 2, 3, 4], Marie Cerciat[1, 2, 3, 4], Tao Ye[1, 2, 3, 4], Andrea Riba[1, 2, 3, 4, 5], and Nacho Molina[1, 2, 3, 4] ✉

[1] *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Illkirch, France*
[2] *Centre national de la recherche scientifique (CNRS), UMR 7104, Illkirch, France*
[3] *Institut National de santé et de Recherche Médicale (Inserm), UMR-S 1258, Illkirch, France*
[4] *Université de Strasbourg, IGBMC UMR 7104- UMR-S 1258, Illkirch, France*
[5] current affiliation: *Capgemini, Turin, Italy.*
✉ corresponding authors: nariyam@igbmc.fr, molinan@igbmc.fr

Highlights

- Integration of single-cell and single-nucleus multiome profiling with deep learning and biophysical modeling uncovers cell cycle regulation of RNA metabolism
- Genome-wide estimation of mRNA transcription, nuclear export, and degradation rates during the cell cycle in mESCs
- Identification of a prominent wave of transcription during early G1 in unperturbed asynchronously proliferating mESCs
- Characterization of temporal waves of post-transcriptional mRNA regulation controlling both mRNA degradation and nuclear export during the cell cycle
- Chromatin accessibility dynamics within TF footprints reveal key cell cycle regulators

Abstract

The cell cycle is a highly regulated process that ensures the accurate replication and transmission of genetic information from one generation of cells to the next. It is a fundamental biological process and plays a crucial role in development, growth, and maintenance of all living organisms, and its dysregulation can lead to a number of pathologies such as autoimmune diseases, neurodegenerative diseases, and cancer. In this work we present a novel approach to study the gene expression and chromatin accessibility dynamics during the cell cycle in mouse embryonic stem cells. To achieve this, we combined high-depth single-cell multiome sequencing, biophysical modeling, and advanced deep learning techniques. First, we used DeepCycle, a deep learning tool that assigns a cell cycle phase to every cell based on its spliced and unspliced mRNA levels. We then developed a biophysical model that describes the dynamics of gene-specific mRNA transcription, nuclear export, and degradation rates during the cell cycle. Our model obtains genome-wide estimates of mRNA transcription, nuclear retention, and degradation half-lives of genes exhibiting oscillatory dynamics in mESCs. The key feature of single-cell multiome sequencing is that it simultaneously provides readouts for gene expression as well as chromatin accessibility in the same cells. By applying our approach to these data we showcase first-of-its-kind exploration of chromatin accessibility during the cell cycle at high temporal resolution. We believe that our work will be pivotal in formulating a coherent quantitative description of mRNA metabolism.

Introduction

Regulation of cell cycle progression is crucial in many fundamental biological processes such as cell differentiation, DNA synthesis, as well as apoptosis. In somatic cells, the cell cycle consists of four main phases–gap1 phase (G1) during which the cell grows in size and duplicates essential cellular components; synthesis phase (S) during which the cell undergoes DNA replication; gap2 phase (G2) during which the cell develops organelles and proteins for cell division; and mitosis (M) during which the cell segregates the chromosomes, produces two nuclei with identical DNA content, and eventually divides into two daughters. The cells can also exit the cell cycle and enter the quiescent phase (G0), while maintaining the ability to re-enter the cell cycle, typically in G1 [1] (see Figure 1A). For more than half a century, this area of research has been bustling with activity [2]. The cell phase transitions are tightly regulated by checkpoint proteins known as cyclins, which form protein complexes with cyclin dependent kinases (CDKs); the role of CDK-cyclin complexes has been studied in great detail in yeast and mammalian cells [3–8]. Cell differentiation, the process by which cells become specialized, is known to be intimately linked with cell cycle; for instance, it is known that during the early embryonic stages, the cells progress through their cell cycle extremely rapidly. As the cells become more specialized, they tend to increase the duration of the cell cycle, typically by lengthening the G1 phase, and terminally differentiated cells undergo complete cell cycle arrest and reside in G0 phase [9–12].

Despite the significant strides made in the field, there are many open challenges. For instance, it is known that the cell makes most of its cell fate decisions during the G1 phase [9, 13]. However, a thorough understanding of the precise regulation of gene programs is still missing. During mitosis, the chromosome condenses and the global gene expression decreases, yet the daughter cells are able to inherit transcriptional programs from the mother cells and maintain their cellular identity. Mitotic bookmarking has been proposed as a possible explanation for this phenomenon [14], but the relationship between the bookmarking factors and the cell cycle is not fully understood. To address some of these challenges, it is vital to investigate the cell cycle with high temporal resolution. Live imaging using fluorescence microscopy has been extensively used to study the dynamics of various genes and proteins involved in regulating the cell cycle [15]. However, the scope of such methodologies is severely limiting, in that it prevents us from obtaining a systems level description for which simultaneously profiling thousands of genes and proteins is essential. One approach to overcome this is to synchronize the cell population by chemically inducing cell cycle arrest, typically in mitotic phase, and profiling bulk populations at various time intervals [16, 17]. The limitation of this approach is twofold: firstly, the cells tend to desynchronize shortly after the treatment, well before completing one full cycle, limiting the usefulness of this approach; and secondly, the perturbation caused in the cells is substantial and might significantly modify the innate biology of interest.

Recently, RNA velocity was introduced as a framework to obtain the systems-level dynamics of gene expression from static snapshots of transcriptional states measured at single cell resolution [18–21]. RNA velocity relies on the steps involved in mRNA metabolism–inside the nucleus RNA polymerase II uses the double stranded DNA as a guide for transcription of pre-mRNA (unspliced mRNA), the non-coding or the intronic regions are removed by the process of splicing to obtain mature mRNA (spliced mRNA), the mature mRNA is exported into the cytoplasm where fulfills its function of protein synthesis, and after a sufficiently long time is degraded (see Figure 1B). Using the information of unspliced and spliced levels in cells at single-cell resolution La Manno et al. demonstrated how to obtain a dynamical picture of cellular processes in pseudotime. We used the RNA velocity framework to develop DeepCycle, a deep learning tool that uses single-cell RNA sequencing (scRNA-seq), to map the gene expression profiles of every cell to a continuous latent variable, θ, representing the cell cycle phase [22]. We used DeepCycle to study dynamics of gene expression at high temporal resolution in embryonic and somatic cells, revealing major waves of RNA accumulation during the cell cycle and identifying key transcriptional regulators involved in cell cycle progression. This naturally leads to the question: How are the processes that govern RNA accumulation, such as transcription, nuclear export, splicing, and degradation, temporally regulated during the cell cycle on a genome-wide scale?

To address this question, we chose mouse embryonic stem cells (mESCs) as an in vitro model system. We used 10X Genomics technology to perform single-cell RNA sequencing as well as single-nucleus multiome sequencing. Using DeepCycle, we inferred the cell cycle phase of both single cells and single nuclei. This enabled us to develop a biophysical model to mechanistically describe the mRNA metabolism during the cell cycle. Our model predicts the cell cycle dependence of transcription, nuclear export, and degradation rates for every gene, revealing waves of transcriptional and post-transcriptional regulation during the cell cycle. Furthermore, our method allowed us to investigate changes in chromatin accessibility at the gene level and their potential connection to transcription. Remarkably, we identified key transcription factors (TFs) and the dynamics

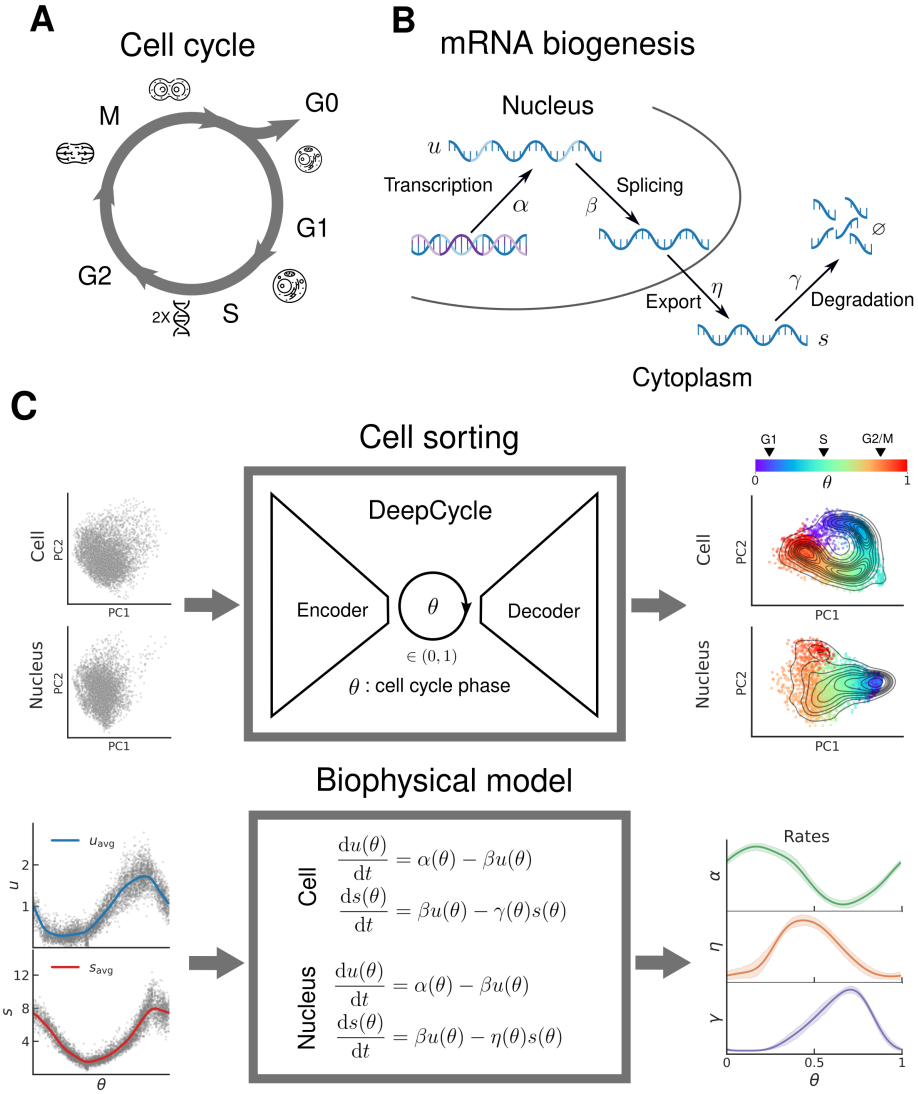

Figure 1. Schematic of cell cycle, mRNA metabolism, and workflow. **A:** Progression through cell cycle phases–G1, S, G2, and M phases. **B:** Schematic diagram of mRNA metabolism: in the nucleus, DNA is transcribed with transcription rate α to unspliced pre-mRNA denoted by u. Introns get spliced out with splicing rate β, giving spliced mature mRNA denoted by s. Spliced mRNA is exported from the nucleus into the cytoplasm with an export rate η and is degraded in the cytoplasm with a degradation rate γ. **C:** Workflow: first, we run DeepCycle to sort mESCs by assigning them a continuous cell cycle phase, θ, based on the unspliced and spliced mRNA profiling of intact cells as well as isolated nuclei at single cell resolution. Note that RNA sequencing of cells and nuclei is performed in separate experiments. Finally, we use a biophysical model of mRNA metabolism to estimate cell cycle-dependent rates of transcription, nuclear export, and RNA degradation.

of their binding activities during the cell cycle by examining changes in chromatin accessibility footprints near their binding sites. In summary, we developed a framework that combines state-of-the-art sequencing technologies with deep learning and biophysical modeling to uncover, for the first time, a global picture of transcriptional and post-transcriptional regulation of mRNA accumulation in mESCs as well as chromatin accessibility dynamics throughout the cell cycle.

Results

DeepCycle sorts single cells and single nuclei according to their cell cycle progression

We cultured mouse embryonic stem cells (mESCs) in leukemia inhibitory factor (LIF), supplemented with an inhibitor of the mitogen-activated protein kinase enzyme MEK, and a glycogen synthase kinase-3 (GSK3) inhibitor (details provided in Methods). This treatment promotes the maintenance of a self-renewal cell state, reducing heterogeneous expression of pluripotency factors and spontaneous differentiation, thereby generating an asynchronous population of proliferating cells [23, 24]. We performed single-cell RNA sequencing (scRNA-seq) using 10X Genomics technology to obtain the transcriptomes of whole mESCs [22]. To complement this dataset with nuclear RNA expression and chromatin accessibility information, we employed the 10X multiome protocol. This involved simultaneously performing single-nucleus RNA sequencing (snRNA-seq) and single-nucleus Assay for Transposase-Accessible Chromatin using sequencing (snATAC-seq) on the same cells [25]. We profiled around 5,000 cells in each experiment at high sequencing depth, revealing the gene expression of roughly 12,000 genes. To obtain an overall view and comparison between the single-cell and single-nucleus data, we calculated the average reads of unspliced and spliced RNA across genes and across cells (or nuclei). Figures 2A and 2B show the distribution of unspliced and spliced reads per cell across all genes. In Figures 2C and 2D, we compare the unspliced and spliced reads in the cell with ones in the nucleus; despite the intrinsic differences in the experimental protocols, we observe a strong correlation between the mRNA reads per cell obtained from the cell and those obtained from the nuclear data—0.86 for unspliced and 0.78 for spliced, suggesting a high degree of coherence in the measurements (see FigureS1 for corresponding figures for replicate 2). The points in the scatterplot in 2C are equally distributed across the $x = y$ line, suggesting that most of the unspliced molecules of the cell reside in the nucleus, whereas most of the points in the scatterplot in 2D are below the $x = y$ line, suggesting that most of the spliced molecules are present in the cytoplasm of cell, which is not surprising.

In a previous work, we developed DeepCycle, a deep learning tool that utilizes gene expression profiles at single cell resolution and RNA velocity to map every cell to a continuous latent phase, θ, representing its progression through the cell cycle [22]. We generated single-cell RNA sequencing (scRNA-seq) data from mESCs and determined the cell cycle phase of each cell using DeepCycle. Our approach effectively reveals the cell cycle manifold embedded in gene expression space, as depicted in the projection onto the two main principal components (PC) in Figure 2E. Single cells are accurately positioned along the dynamic cycling trajectory and color-coded based on their progression through the cell cycle. Unique molecular identifiers (UMIs) create a distinct identity for each input mRNA molecule allowing to correct for amplification bias in the sequencing measurements [26, 27]. Remarkably, the number of UMIs, which in our case can be interpreted as total mRNA content of the cell, show a two-fold change as the cells (gray dots) progress through the estimated cell cycle phase (see Figure 2G). This is the expected behavior, assuming that the concentration of mRNA remains constant during the cell cycle [28]. Sorting the cells by their cell cycle phase reveals gene-specific dynamics of unspliced and spliced levels of mRNA, as exemplified in Figure 2I for Cyclin B1 (Ccnb1) which is crucial in regulating the G2/M transition during the cell cycle [29, 30]. Note, we used the approach described in [22] to obtain the cell cycle phase transitions (grey dotted lines in 2G-I).

We applied the same approach to gene expression profiles obtained from single nuclei. In principle, nuclear RNA expression exhibits higher levels of pre-mRNA, with the accumulation of mature mRNA predominantly influenced by RNA export—a process typically faster than cytoplasmic RNA degradation [31]. Nevertheless, DeepCycle successfully inferred cell cycle phases when applied to snRNA-seq data. While the cell cycle manifold may not appear to be cyclic on the PC projection in Figure 2F, the single nuclei, color-coded according to their progression through the cell cycle, are correctly ordered. It is worth noting that the cells approaching mitosis and the cells in early G1 phase were filtered out as they often appear as extreme outliers in the PC-space (see Figure S1). Importantly, the total mRNA content within the nucleus exhibits a similar two-fold increase as observed in the whole cell during the cell cycle (see Figure 2D). Lastly, the unspliced and spliced RNA nuclear levels of Ccnb1 mirror the dynamics observed in the entire cell (see Figure 2F). Note that in Figures 2H and J, the grey dots get scarce near θ approaches 0 or 1, this is likely because the nuclear envelope breaks down during mitosis and the sequencing technique is unable to profile these nuclei. In summary, we obtained a unique scRNA-seq and snRNA-seq dataset at a high sequencing depth,

and DeepCycle accurately assigns a continuous phase, not only to single cells but also single nuclei, indicating their progression through the cell cycle. This dataset puts us in a unique position to simultaneously study RNA metabolism and chromatin accessibility dynamics throughout the cell cycle.

Kinetic modeling reveals cell cycle-dependent mRNA transcription and degradation rates

DeepCycle reveals oscillatory patterns in mRNA accumulation throughout the cell cycle, prompting an investigation into the underlying processes of mRNA metabolism responsible for these dynamics. To address this question, we constructed a biophysical model of mRNA metabolism. As mentioned earlier, the key steps involved in mRNA metabolism include transcription, splicing, nuclear export, and degradation. These processes were modeled using a system of coupled differential equations with periodic functions to capture cell cycle-dependent rates. We derived a pseudo-analytical solution using Fourier series expansion (details provided in the supporting information). Finally, we used normalized unspliced and spliced reads, along with the cell cycle phase as observable variables, to fit the fundamental rates of RNA metabolism (details in the Methods section).

Figures 3A and 3B show the dynamics of the unspliced and spliced RNA levels as well as the dynamics of the transcription and degradation rates as predicted by our model for the representative genes Ccne1 and Wee1. The grey dots depict the data and the solid lines show the model predictions with 95% confidence intervals indicated by shaded areas (unspliced in blue, spliced in red, transcription in green, degradation in purple). The degradation rates are expressed in hour^{-1} and are obtained assuming that the cell cycle duration in mESCs is approximately 14 hours [22]. We obtained the baseline rates from the zeroth Fourier coefficients; these are constant rates representing the time-average over the cell cycle. Ccne2 encodes the G1/S-specific cyclin-E2 protein, essential for cell cycle control during the late-G1/early-S phase [32, 33]. In Figure 3A, a significant increase in the transcription rate during the G1/S transition is observed, while the degradation rate remains relatively constant, suggesting that the mRNA level of Ccne2 is primarily regulated by transcription. On the other hand, Wee1, which acts as a negative regulator of entry into mitosis [34, 35], shows a sharp increase in mRNA degradation during the M/G1 transition, while the transcription rate does not fluctuate much relative to the baseline (see Figure 3B). This suggests that, in the case of Wee1, mRNA expression is mainly regulated post-transcriptionally.

To determine gene-specific preferences for transcriptional or post-transcriptional regulation genome-wide, we constructed three alternative models: one where only transcription is cell cycle-dependent, with splicing and degradation remaining constant; another where only degradation is cell cycle-dependent, with transcription and splicing remaining constant; and a third where transcription, splicing, and degradation all remain constant during the cell cycle. We fitted these models to the sequencing data and used Bayesian information criterion (BIC) to perform model selection for every gene. Figure 3C shows that for 67.5% of the genes, a full model with cell cycle-dependent transcription as well as degradation is preferred; 1% of the genes prefer the model where only degradation is cell cycle dependent; 19.2% of genes prefer the model where only transcription is cell cycle dependent and, 12.5% of the genes prefer the model with no cell cycle dependence. Due to the high noise level of single-cell sequencing technologies, many genes do not fit any of the models with the desired level of statistical significance. Thus, we used the explained variance, r^2, as a measure of the discrepancy between the full model and the data. By filtering genes with r^2 ¿ 0.2, approximately 1,800 out of the 12,000 profiled genes met this threshold on explained variance (see Figure 3D). All downstream analyses were restricted to these genes. Overall, we were able to fit our biophysical model of mRNA metabolism to thousands of genes.

We investigated the genome-wide synthesis and degradation rates for all the genes satisfying the cutoff mentioned above for the explained variance. As shown in Figure 3E and 3F, the baseline transcription rates and degradation rates in mESCs span a large dynamic range across genes–more than two orders of magnitude. Since the model is fitted to relative quantifications of unspliced and spliced reads, the numerical values of transcription rates in Figure 3E do not reflect the absolute number of molecules synthesized per hour; however, fold changes with respect to the baseline provide information on the transcriptional resources dedicated to a particular gene during cell cycle progression. On the contrary, the degradation half-lives provided in Figure 3F are in units of hours, assuming a typical cell cycle duration of 14 hours [22]. We found that the mean of the baseline degradation half-lives was around 4.5 hours, whereas the mean of the maximum degradation half-lives was 2.7 hours. These values overall agree with previously reported estimates of mRNA degradation rates [36, 37]. In Figure S2C, we show the distribution of splicing rates. Note that by experimental design, the rate β is confounded by intronic degradation, which restricts its interpretation as an effective splicing rate.

We also calculated the difference in time between peak spliced and peak unspliced levels. For the majority of the

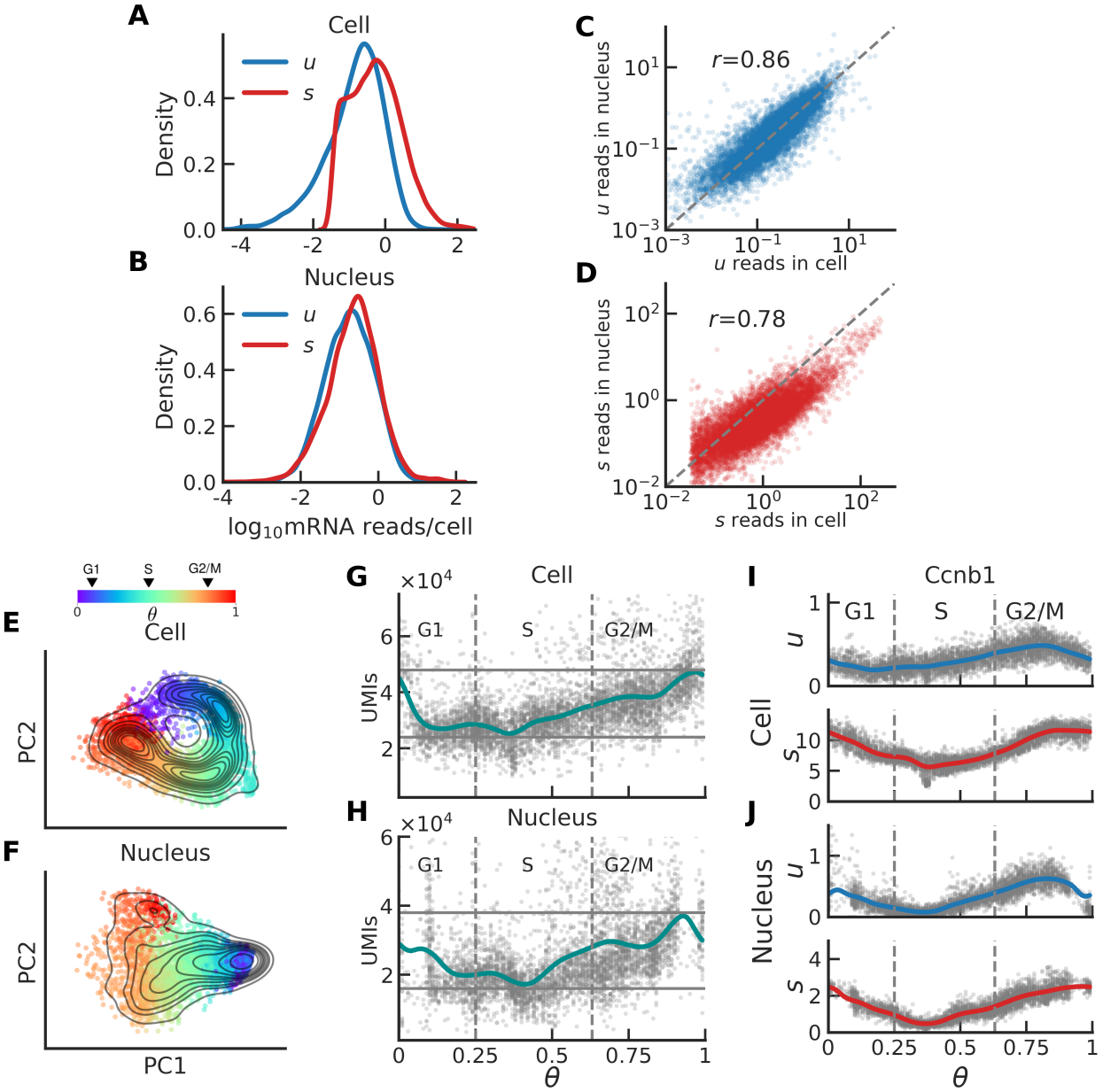

Figure 2. DeepCycle infers the cell cycle phase of both single cells and single nuclei. A and **B:** Distribution of unspliced (blue) and spliced (red) reads per cell, across all genes in the cell and the nucleus respectively. **C** and **D:** Correlation between the reads in the cell versus those in the nucleus for the unspliced (blue) and spliced (red) molecules respectively, every dot represents a gene. **E** and **F:** Visualization of the cycling dynamics based on the sequencing profiles of the cells and isolated nuclei respectively, every dot represents a cell, the color indicates the progression through the cell cycle–early G1 (in purple) through mitotic(in red), **G** and **H:** Unique molecular identifiers (UMIs), a proxy for the total mRNA content in the cell and the nucleus respectively, against the cell cycle phase θ; there is a two-fold change in the total mRNA content, both in the nucleus as well as in the cell, during the cell cycle. **I** and **J:** Expression dynamics of the unspliced and spliced reads for Ccnb1 in the cell and the nucleus respectively.

genes, the time difference is less than one-eight of a cycle, which translates to less than 1.75 hours in real time (see Figure S2). To gain insight into the fold changes in the rate parameters, we calculated the amplitude of the rates, i.e. $(\max - \min)/2$, and compared it to the baseline rates (see Figure 3G). For most of the genes, the fold-change in amplitude for degradation is higher than transcription. Model selection using BIC suggests that mRNA metabolism regulation requires both transcription and degradation to be cell cycle dependent; however, for the majority of genes, the changes in degradation rates contribute more than the changes in transcription to regulating gene expression dynamics. Therefore, mRNA post-transcriptional regulation has a prominent role in shaping mRNA accumulation during the cell cycle.

Waves of transcriptional and post-transcriptional regulation during the cell cycle

After investigating the global trends in transcription and degradation rates, we turned our attention to the timing of peak rates during the cell cycle. Figure 4A is a polar histogram of cell cycle phases corresponding to the maximum transcription for all selected genes. We observed distinct waves of mRNA synthesis during the cell cycle. In particular, 762 genes (44.3% of the 1778 filtered) achieved peak transcription during the G1 phase, 481 (28.0%) genes during the S phase, and 477 (27.7%) genes during the G2/M phase. To better visualize the changes in transcription rate genome-wide during the cell cycle, we normalized the predicted transcription rates using max-min normalization and arranged genes based on the timing of their peak transcription as depicted in Figure 4C. Remarkably, we observed a prominent wave of maximum transcription in the early G1 phase. While this phenomenon has been previously shown in synchronized cell populations [16, 17, 38], to the best of our knowledge, this is the first time it is reported in unperturbed and unsynchronized mESCs. We repeated this analysis for the gene-specific degradation rates, shown in Figures 4B and and 4D. In this case, we found that 820 (47.7%) genes achieved peak degradation during the G1 phase, 707 (41.1%) genes during the S phase, and 193 (11.2%) genes during the G2/M phase. Interestingly, we observed distinct waves of degradation concentrated mainly during G1 and S phases (see Figure 4B and 4D).

To gain insight on the biological processes related to transcriptional and post-transcriptional regulation during the cell cycle, we performed enrichment analysis[39, 40], on the gene sets shown in Figures 4E and 4F. Our analysis revealed that during the G1 phase, the mESCs upregulate genes involved in fundamental processes such as cellular amine metabolism, DNA replication initiation, and cell cycle G1/S phase transition, processes that are essential for cell growth and proliferation. S phase involves peak transcription of genes involved in processes such as nucleosome organization, regulation of cell cycle G2/M transition, establishment of spindle orientation, suggesting that cells are already preparing for mitosis. During the G2/M phase, the mESCs synthesize genes involved in a variety of mitosis related pathways–mitotic sister chromatid segregation, metaphase plate congression, and microtubule cytoskeleton organization involved in mitosis. The enrichment analysis of the degradation waves revealed many of the same processes. The degradation wave during the G1 phase consists of genes that are involved in mitotic regulation, suggesting that these genes were produced during the G2/M phase and are degraded upon entering the G1 phase, when they are no longer needed. Similarly, during the S phase, the mECSs degrade genes involved in processes such as DNA replication initiation, G1/S phase transition, translesion, etc. which achieved peak transcription during the G1 phase. During the G2/M phase, we found peak degradation of genes involved in cellular amine metabolic processes, negative regulation of the G2/M transition, and regulation of hematopoietic stem cell differentiation. Overall, these analyses indicate that our computational methods for performing cell cycle sorting and rate inferences are producing consistent results.

Genome-wide estimates of mRNA nuclear export rates and waves of nuclear export

The 10X Genomics multiome protocol, combined with high-depth sequencing, yielded high-quality quantifications of nuclear unspliced and spliced mRNA in individual nuclei. This allowed us to investigate the accumulation of nuclear mRNA as a dynamic process of gene-specific transcription and export rates that change during the cell cycle. To do this, we used a similar biophysical model to describe the expression changes in the nucleus with one key distinction: instead of degradation, we have nuclear export, denoted by η, as the dominant process leading to depletion in spliced levels inside the nucleus. Note that mRNA stability in the nucleus may also play a role in shaping the accumulation dynamics of spliced nuclear mRNA; however, under the assumption that nuclear export is a faster process than nuclear degradation [41], we can neglect its effects. As we did previously, we used BIC to approximate the posterior probabilities over alternative models and investigate whether the observed mRNA accumulation dynamics support transcription and export rates that change during the cell cycle. Surprisingly, we found that the full model was preferred for the vast majority of the

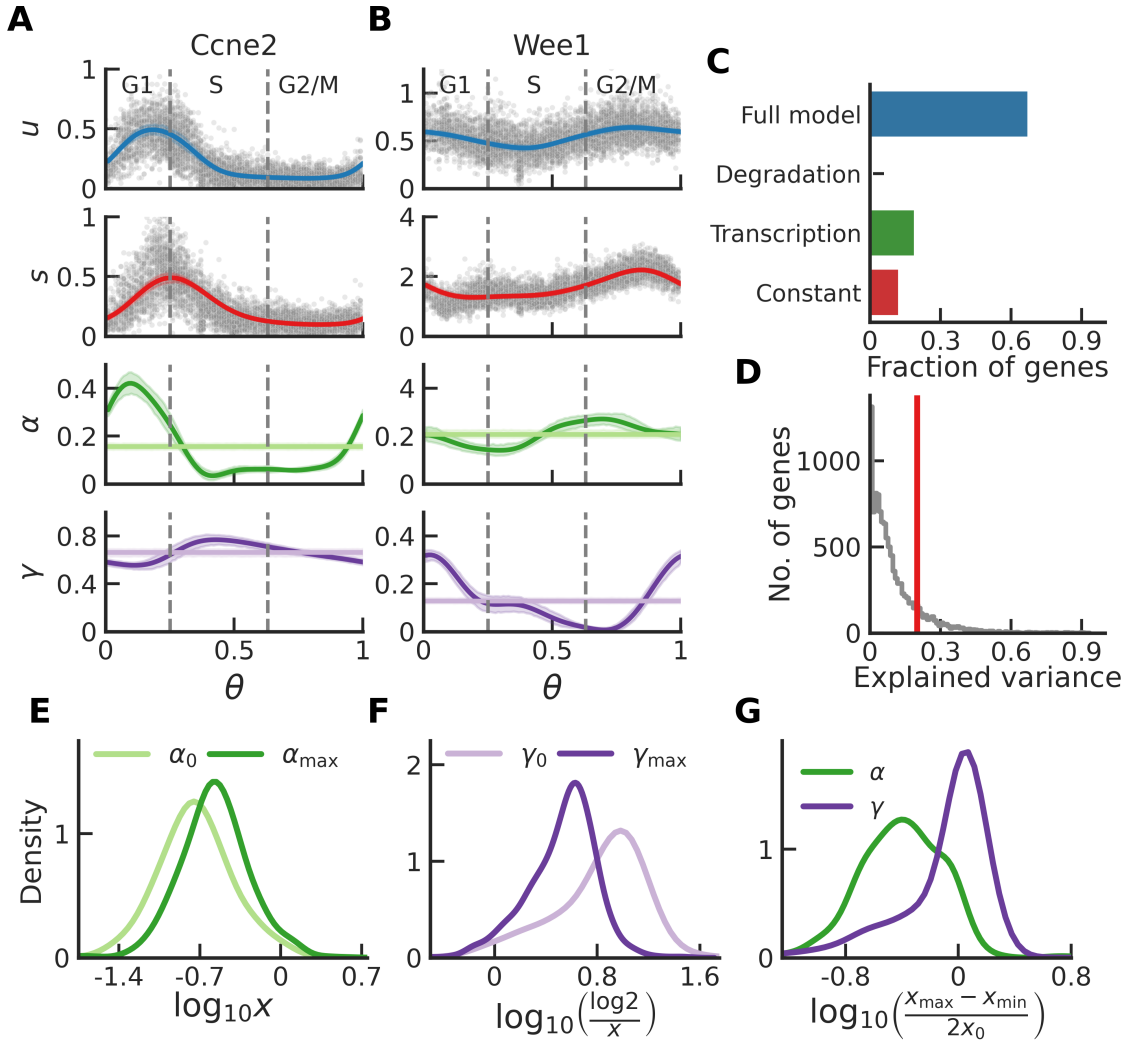

Figure 3. Genome-wide estimation of global transcription and degradation rates and characterization of transcriptional and post-transcriptional regulation. A and **B** (top to bottom): number of unspliced reads (u), number of spliced reads (s), transcription rate (α), degradation rate (γ) as function of the cell cycle phase θ. Grey dots represent single cells, solid lines show model predictions, and shaded regions are 95% confidence intervals from bootstrapping. **C:** Barplot of fraction of genes selecting alternative models based on the Bayesian Information Criterion–full model (blue), model with cell cycle dependent degradation and constant splicing and transcription (purple), model with cell cycle dependent transcription and constant splicing and degradation (green), model with constant transcription, splicing, and degradation (red), for the vast majority of the genes the full model is preferred. **D:** Histogram of the explained variance in spliced counts across all the genes; red line is at $x = 0.2$, the threshold set for filtering genes. **E:** Distribution of transcription rate, in arbitrary units, across filtered genes, α_0 denotes baseline transcription (light green) whereas $\alpha_{\max}$ denotes maximum transcription rate (green), F: Distribution of degradation half lives, in hours, across filtered genes, γ_0 denotes baseline degradation rate (light purple) whereas $\gamma_{\max}$ denotes maximum degradation rate. **G:** Fold change in amplitude of rates, $(x_{\max} - x_{\min})/2$, over the baseline rate for transcription (in green) and degradation (in purple). For most of the genes, the changes in degradation rates are more prominent than those in transcription rates compared to the corresponding baseline rates.

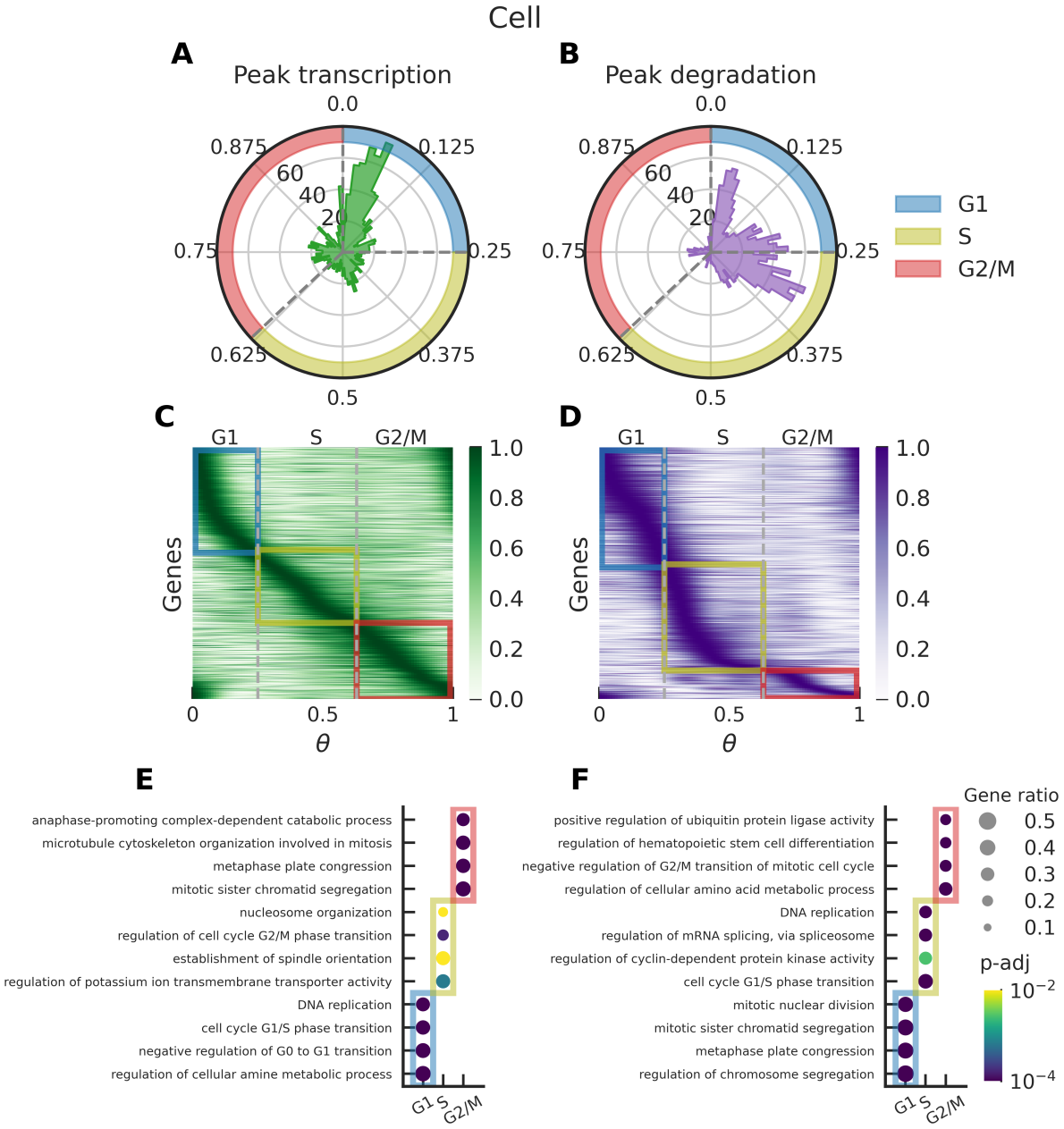

Figure 4. Waves of transcription and degradation and the biological processes involved during the cell cycle. **A** and **B:** Polar histograms of the cell cycle phase at which a gene achieves the highest transcription rate and the highest degradation rate, respectively. **C** and **D:** Heatmaps of transcription and degradation, respectively, during the cell cycle. The rates are normalized using max-min normalization, and the genes are arranged based on the timing of the maximum rate. Genes that achieve their peak transcription and degradation rates during G1, S and G2/M phases are highlighted in light blue, olive, and pink, respectively. **E** and **F:** Gene set enrichment analysis (using GSEApy) for genes that achieve peak transcription and degradation respectively. Gene sets were compared against GO biological processes 2018 (in mouse) size of the dot represents the extent of overlap between the gene set and the GO terms, the color represents the adjusted p-value.

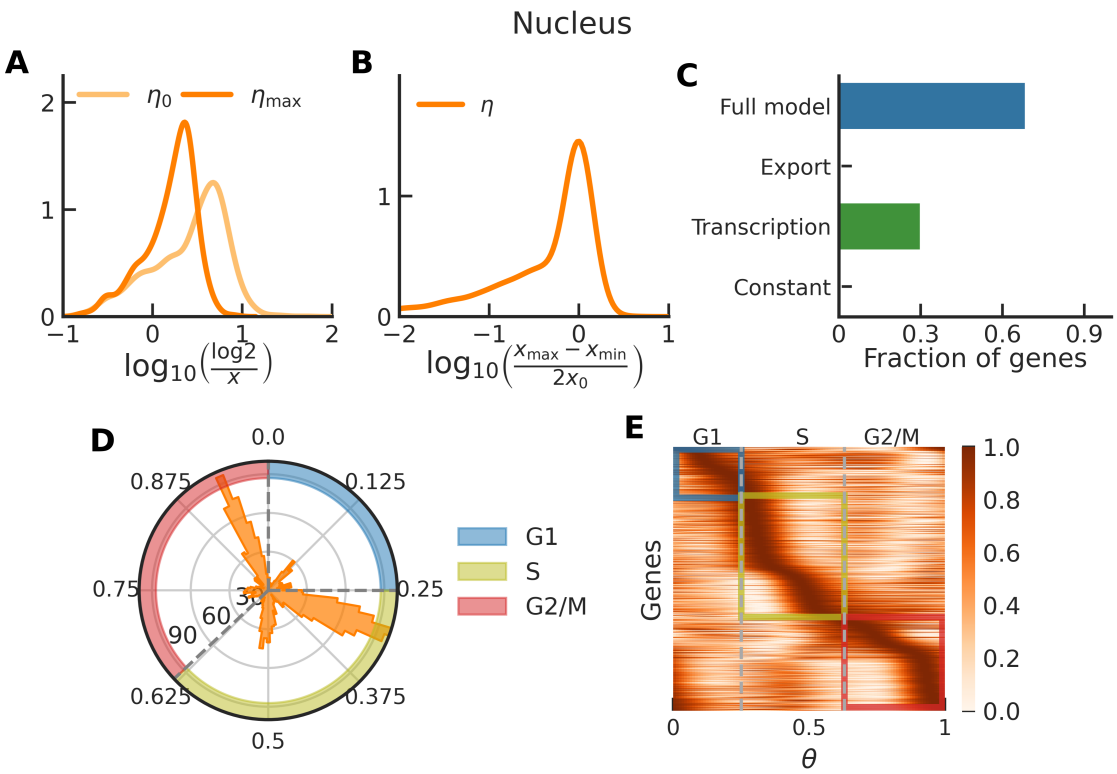

Figure 5. Genome-wide estimation of global export rates and characterization of waves of export during the cell cycle. A: Distribution of export half-lives, in per hour, for baseline export rate (light orange) and maximum export rate (orange), export rates span a large range, mean baseline export half-life is 1.5 hours. **B:** Distribution of fold-change in amplitude over baseline. **C:** Barplot of fraction of genes selecting alternative models based BIC–full model (blue), model with cell cycle dependent export and constant splicing and transcription (purple), model with cell cycle dependent transcription and constant splicing and export (green), model with constant transcription, splicing, and export (red), for vast majority of the genes the full model is preferred. **D:** Polar histogram of the cell cycle phase at which a gene achieves the highest export rate. **E:** heatmaps of nuclear export during the cell cycle, the rates are normalized using max-min normalization, and the genes are arranged based on the timing of maximum export rate.

genes (67.6%), indicating that nuclear mRNA retention may be regulated during the cell cycle instead of depending on a passive nuclear export process of mRNA (see Figure 5). For 19.0% of the genes, we found that only transcription was cell cycle-dependent, and 12.5% of the genes preferred a model with constant rates. We also calculated the explained variance and filtered genes with $r^2 < 0.2$, out of 12,000 genes that were profiled, roughly 2,000 satisfied the criterion for explained variance (see Figure S3F and S3I). The framework of RNA-velocity relies heavily on the quality of reads, particularly on how well we can map them to exonic and intronic regions. We believe that in the case of the snRNA-seq the capture efficiency of intronic reads (unspliced mRNA) is superior to that obtained in scRNA-seq.

In Figure 5A, we show the distribution of baseline and maximum export half-lives across genes that satisfy the explained variance criterion. As seen in the case of transcription and degradation, nuclear retention half-lives also span a large dynamic range spanning two orders of magnitude, with the mean baseline half-life of 1.5 hours and the mean maximum half-life of 36 mins. These times are slightly longer than the values that have been previously reported for a limited number of genes [42]; however, we believe this is the first time that these rates are estimated genome-wide in mESCs. Figure 5B shows the distribution of amplitude fold-change over the baseline for export rates ranging from 0.001 to 4.11 times the baseline. Indicating that for some genes, the change in export rate can play a significant role in the dynamics of mRNA accumulation during the cell cycle.

We also examined the timing of peak transcription and export rates. Figure 5D shows the polar histogram of the cell cycle phase corresponding to peak export rates (polar histogram for peak transcription in Figure S5). Strikingly, our method uncovers very prominent waves of nuclear export during the cell cycle. Of the 2170 genes that satisfied the explained variance criterion, 331 (15.3%) genes achieved peak export during the G1 phase, 1101 (50.7%) genes during the early-S phase, and 738 (34%) genes during the G2/M phase. The heat map in figure 5E shows the export rate normalized using max-min normalization, highlighted with their corresponding cell cycle phases.

Dynamics of chromatin accessibility within genes during the cell cycle

The novelty of multiome profiling is that it provides information about chromatin accessibility in addition to gene expression data for the same cells. This feature gave us the unique opportunity to investigate chromatin accessibility dynamics throughout the cell cycle by combining cell cycle phase estimation obtained based on gene expression using DeepCycle, and chromatin accessibility readout obtained by collecting all ATAC fragments mapped to a specific genomic region. We first examined the replicate correlation in the accessibility profiles for every gene (see Figure S6) and filtered out those genes for which the correlation was weak ($\rho > 0.75$). Figure 6A shows the heat map of changes in raw accessibility throughout the cell cycle for the filtered genes, and Figure 6B shows the average accessibility profile across the genes. Interestingly, we found that chromatin accessibility was highest for most genes in cells undergoing the S/G2 transition. This may be a consequence of DNA synthesis in combination with chromatin condensation during mitosis, i.e. genes that underwent DNA replication would contribute twice as much material, which serves as a readout for accessibility, followed by a global reduction of accessibility during mitosis. Furthermore, the heatmap in 6A also suggests a preference in the order in which genes undergo DNA replication during the cell cycle. To account for the ploidy, we normalized the accessibility readout by the total number of ATAC fragments detected in the nucleus. We found that the normalized accessibility profile revealed gene-specific patterns of accessibility during the cell cycle. Averaging the normalized accessibility across all genes in Figure 5D, we observed that there are two global peaks in relative accessibility, one during the G1/S and the second during the S/G2 transition. The normalized accessibility profile shows a prominent dip in global accessibility during mitosis, as expected due to chromatin condensation. In Figure 6E, we compare the z-scored moving averages of unspliced, spliced, and normalized chromatin accessibility profiles for select genes. Ccnb2, Nusap1, and Kif2c are known to achieve peak expression during the G2/M transition; however, their accessibility profiles vary significantly from one another (see Figure S7 for corresponding panels for replicate 2), indicating that the relationship between chromatin accessibility and transcription might not be straightforward.

Chromatin accessibility dynamics within TF footprints reveal key cell cycle regulators

ATAC-seq data has been instrumental in characterizing TF footprints within accessible chromatin regions [43]. Our single-nuclei multiome experiments, together with computational cell cycle sorting, allowed us to investigate the dynamic changes of TF footprints throughout the cell cycle. To that end, we first used a set of known Position Weight Matrices (PWMs) of 678 TFs to identify putative binding sites on the DNA within ATAC peaks found by CellRanger [44]. We then calculated TF footprints by combining all the observed Tn5 cuts at a given distance with respect to the predicted binding site

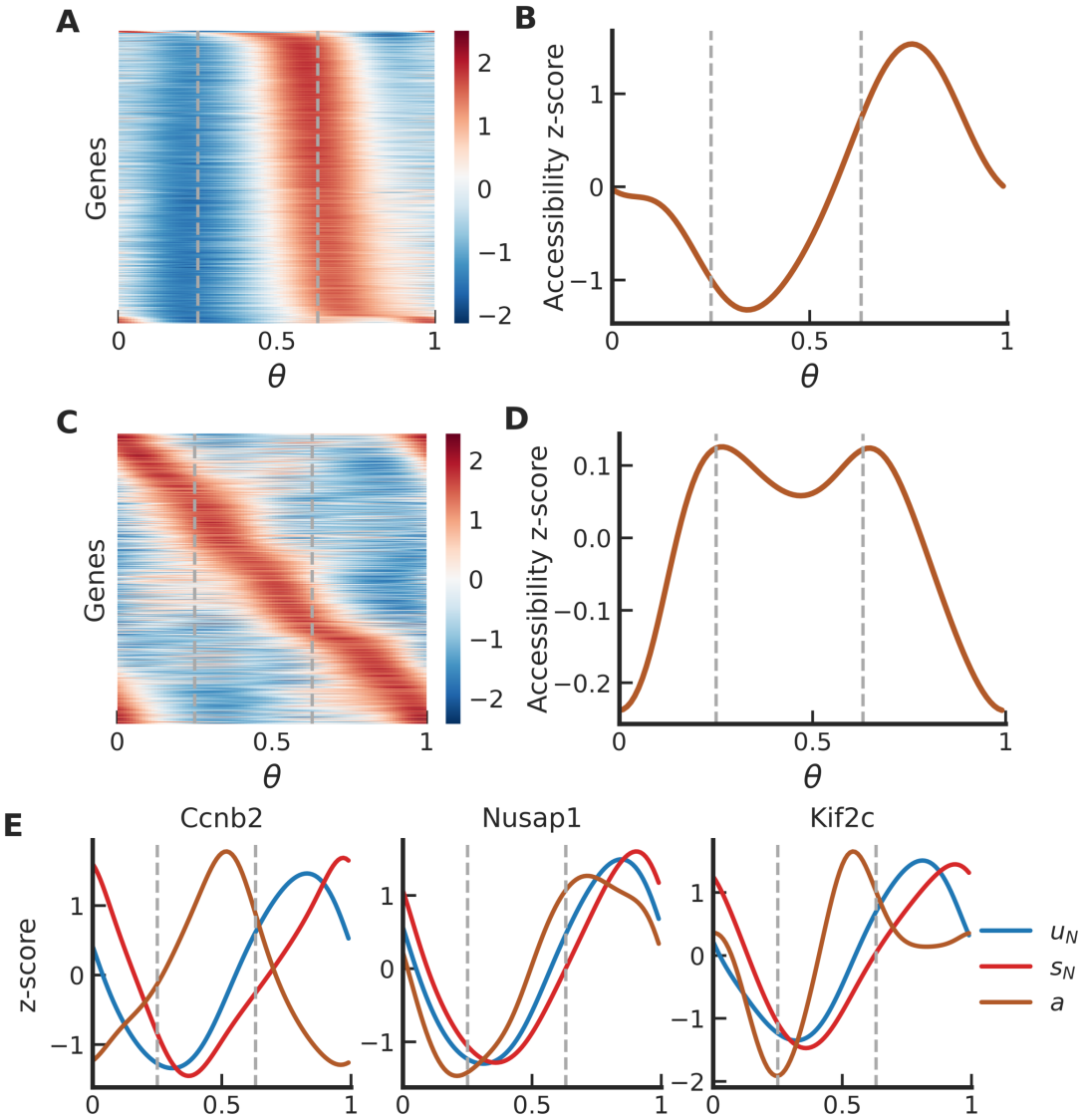

Figure 6. Chromatin accessibility during the cell cycle. A: Unnormalized chromatin accessibility against the cell cycle phase, θ, for genes with high replicate correlation, chromatin accessibility is maximum during the S/G2 transition which is consequence of the fact that there is twice as much DNA in these cells compared to that have not undergone DNA synthesis, the heatmap also suggests a preference in the order of which the genes undergo DNA replication. **B:** Average chromatin accessibility, across the genes show in panel A, is lowest at the beginning of S phase and highest at the beginning of G2 phase. **C:** Chromatin accessibility normalized for the total genetic material in cell, normalized accessibility of the genes is more evenly distributed during the cell cycle. **D:** Average accessibility is highest at the beginning of S and G2 phases and there is a global dip in accessibility during mitosis, likely due to chromatin compactification. **E:** Comparison of the dynamics of accessibility, spliced and unspliced profiles for select genes during the cell cycle; it is known that Ccnb2, Nusap1, and Kif2c expression levels peak in late G2/M phase, however, their accessibility peaks at different times during the cell cycle.

(see Methods). In Figure 7A, we show the characteristic average footprints of two representative factors with distinct profiles, Tfcp2l1 and Pml. By repeating the same analysis considering nuclei in a specific cell cycle phase bin, we could calculate how the footprints fluctuate with respect to the average throughout the cell cycle (see Figure 7B). Interestingly, we observe an increase in chromatin accessibility around Tfcp2l1 binding sites during the G1-S transition, reaching a minimum during the G2-M transition (see Figures 7C and 7D). In contrast, the chromatin accessibility around binding sites of Pml exhibited a pronounced maximum during the G2-M transition (see Figures 7C and 7D). These changes in the TF footprints suggest potential variations in the binding activity of these TFs across different cell cycle phases. Interestingly, previous studies have established that Tfcp2l1 and Pml play an important role in controlling cell proliferation [45, 46], which supports the functional relevance of our findings. We then ranked all the investigated TFs based on the significance of their accessibility footprint changes (see Figure 7E). Remarkably, among the top 40 TFs, several have been previously reported as regulators of the cell cycle (e.g: Chd1, E2f1, Nfya and Nrf1). This further supports the validity of our approach and highlights its potential to identify novel TFs involved in regulating transcription during the cell cycle (see Figure 7E). In summary, our method unveiled distinct cell cycle-dependent changes in TF footprints of chromatin accessibility, generating predictions of the specific phases when TFs are likely to actively interact with DNA and potentially regulate transcription during the cell cycle.

Discussion

The RNA-velocity framework, which relies on the steps involved in mRNA metabolism and allows to perform trajectory inference based on single-cell sequencing data, has been useful for characterizing gene expression dynamics in various biological contexts [20, 21]. DeepCycle uses the RNA-velocity framework to infer the cell cycle phase based on the unspliced and spliced levels of pre-selected cycling genes. Thus, DeepCycle reveals the oscillatory dynamics of gene expression at a very high temporal resolution [PMID: 35606383] (ref your previous paper). DeepCycle uses the RNA-velocity framework to infer the cell cycle phase based on the unspliced and spliced levels of pre-selected cycling genes. Thus, DeepCycle reveals the oscillatory dynamics of gene expression at a very high temporal resolution [22]. In this work we developed a biophysical model that explains the dynamics of the underlying process of mRNA metabolism regulating gene expression and applied it to single-cell sequencing data as well as multiome data for proliferating mESCs. Our model makes genome-wide predictions for transcription, nuclear export, and degradation rates in mESCs and gives insight about the timing of aup or down regulation of cellular biological processes during the cell cycle. And Llastly, combining the cell cycle phase information along with the chromatin accessibility allows us to investigate the dynamics of chromatin accessibility within genes and regulatory sites during the cell cycle.

DeepCycle was developed based on scRNA-seq data; we found that the methodology extends compatibility with snRNA-seq data as well. It was previously known that the number of mRNA molecules in the cytoplasm increases during the cell cycle. Using UMIs as a proxy for the amount of mRNA, we found that this is true even inside the nucleus, suggesting that the local mRNA concentration stays constant during the cell cycle in both the cytoplasm as well as the nucleus. We also found extremely strong correlations in reads from scRNA-seq and snRNA-seq even though these are two completely different experimental protocols, giving us confidence in the quality of the measurements.

One of key features of our model is that it makes genome-wide predictions for transcription, nuclear export, and degradation rates. Predictions for the baseline rates, i.e. the cell cycle independent rates, were in agreement with the available literature [36, 37], validating our approach. It is noteworthy that the measurements for the rate parameters in the literature are derived from a variety of experiments, our biophysical model allows us to make predictions for all of these rate parameters based on a single experiment highlighting the resourcefulness of this approach. Comparing the BIC score for alternate models we found that for most of the oscillating genes, the full model with cell cycle-dependent transcription, export, and degradation is preferred. Based on the distributions of the fold-change in amplitude compared baseline rates we found that cell cycle-dependent changes in post-transcriptional regulation contribute to gene expression to an even greater extent than cell cycle-dependent transcription in controlling the gene expression dynamics in proliferating mESCs.

We examined the timing of peak rates and found prominent waves of transcription, export, and degradation during the cell cycle. We found that nearly half of the oscillating genes achieved peak transcription the early-G1 phase, this finding is inline with the fact that cells tend to make most of their cell-fate decisions during G1 phase. The peak transcription of the remainder of the genes is more evenly distributed throughout the cell cycle. The dynamics of nuclear export remains extremely poorly characterized, primarily due to lack of reliable transcriptomic measurements for the nucleus. In this work we debut a genome-wide estimation of nuclear

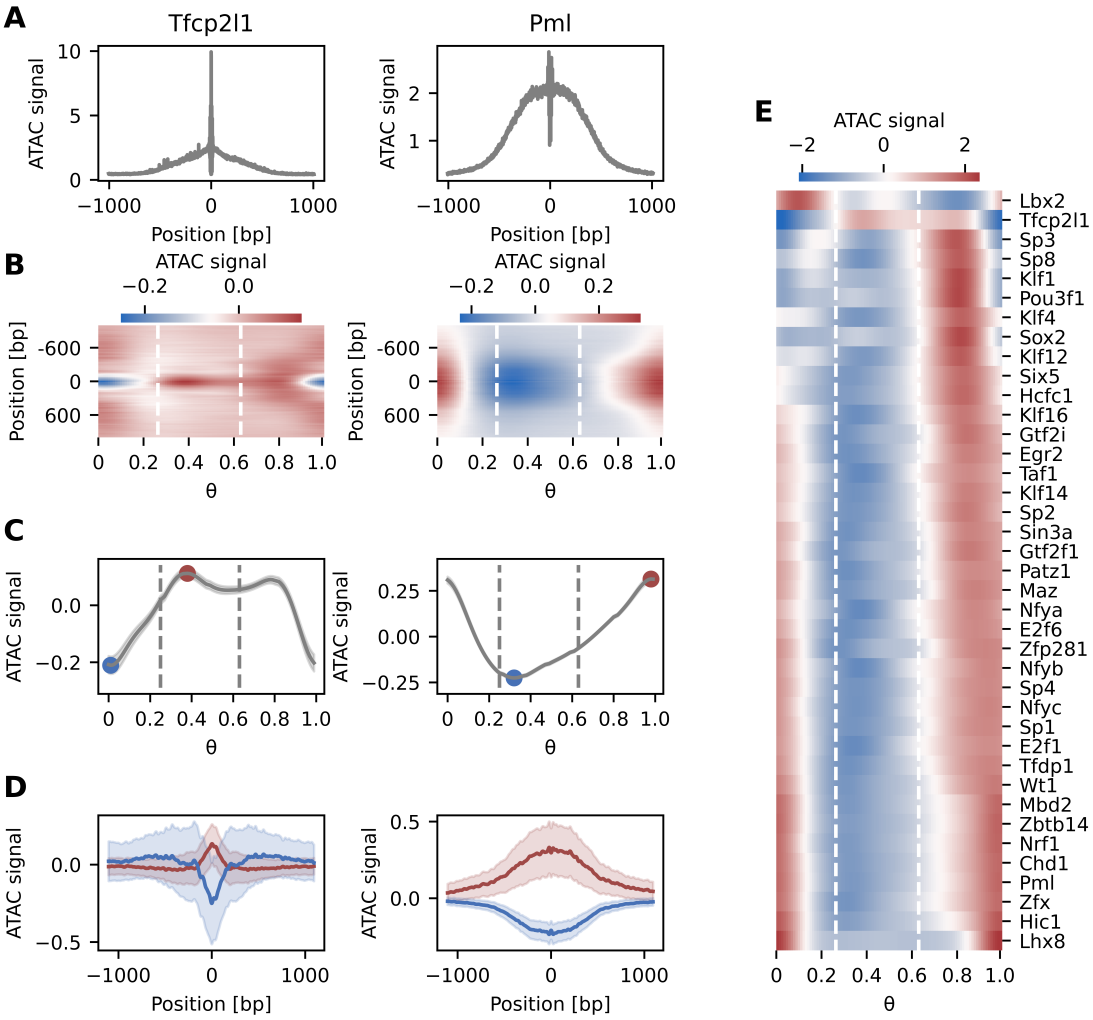

Figure 7. TF footprint dynamics during the cell cycle uncover key cell cycle regulators. A: Average ATAC-seq footprints of TFcp2l1 and Pml showing the normalized number of Tn5 cuts centered relative to predicted binding sites within ATAC peaks. **B:** Cell cycle dynamics of the Tfcp2l1 and Pml footprints as a function of the cell cycle phase θ relative to the average footprint depicted in A. White dashed lines delimit cell cycle phases in the following order: G1, S and G2/M. **C:** Changes in chromatin accessibility during the cell cycle averaged within a 100bp window centered at the TF binding site. Maximum and minimum accessibility are indicated in red and blue, respectively. Gray dashed lines delimit cell cycle phases in the following order: G1, S and G2/M. **D:** Relative TF footprint at the maximum (red) and minimum (blue) of accessibility as indicated in C. Shaded areas represent standard deviations. **E:** Changes in averaged chromatin accessibility around binding sites of the top 40 TFs for which the change is most significant. White dashed lines delimit cell cycle phases in the following order: G1, S and G2/M.

export rates. We found that only about 15% of the genes achieve peak export during the G1 phase, and nearly 50% of the gene achieve maximal export in S phase, hinting towards preferential nuclear retention of mRNA transcripts (see figure S5 for distribution of difference between peak transcription and peak export). Recent work by Krenning et al. reports two distinct waves of mRNA degradation in RPE1 cells–an immediate wave within 20 mins after metaphase and a delayed wave roughly 80 mins after metaphase [47]. We observed a prominent degradation wave in mESCs at around 25 mins of reentry into G1; in addition to this we also observed two additional waves of degradation–a late G1 wave and an early S wave. [Concluding statement]

The advent of multiome single cell profiling has made it possible to simultaneously profiling gene expression as well as chromatin accessibility at single cell resolution. This provided us with a unique opportunity of pioneering the investigation of accessibility dynamics during the cell cycle progression. One of the surprising findings from this work was that genome-wide the raw chromatin accessibility was highest during the S/G2 transition, one would expect that it would be highest during the S phase, given the DNA is being duplicated. On the other hand, normalized accessibility does show a peak in accessibility during G1/S transition, the accessibility stays relatively high during S phase, and there is a global dip during mitosis, as one would expect due to chromatin condensation. We were also able to probe gene-wise chromatin accessibilities and compare them with the gene expression readouts from snRNAseq data, from the example of Ccnb2, Nusap1, and Kif2c, we found hints of differential mechanisms involved in regulating the steps between chromatin accessibility and active transcription of associated genes. The scope of our approach is not limited to in vitro systems, it could be easily extended to others, such as organoids or patient samples, and could help better understand mechanisms in developmental and disease biology.

Limitations of the study

Single-cell sequencing technologies are inherently noisy, more so in the case of multiome sequencing where the genetic material from a single nucleus is shared between two experiments. We would like to highlight there was extremely good concurrence in analysis between the two replicates snRNA-seq. Recent work of Gorin et al. highlight the shortcomings of RNA velocity framework concerning both the underlying assumption of the biophysical model implemented popular softwares well as the actionable reliability, in part arising due to the large number of user-set hyperparameters [48]. We believe that future research in this field will mitigate these shortcomings and will enhance the impact of studies based on the RNA velocity framework. In its current implementation, DeepCycle is most effective for cells that are largely proliferative, but we know that cells often become quiescent, undergo irreversible cell cycle arrest and become senescent, or activate a variety of differentiation programs. In future work we would like to develop tools that unify some, if not all, of these biological phenomena.

Methods

Cell culture

Mouse ES cells E14 were cultivated without feeders on 0.1% gelatine-coated dishes in DMEM high glucose (4.5 g/L) with GlutaMAX supplemented with 15% fetal calf serum heat inactivated, 0.001% B-mercaptoethanol, 1X non-essential amino acids, 1X NaPyr, 1X LIF, 1 mM MEK Inhibitor PD 0325901, and 3 mM GSK3 inhibitor CHIR99021.

Nuclei isolation

Nuclei were isolated from 4X105 mouse ES cells using 100 μL of cold lysis buffer (10 mM Tris-HCl (pH7.4), 10mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 0.1% IGEPAL CA-630, 0.01% Digitonin, 1% BSA, 1mM DTT, 1 U/μL RNase inhibitor). Cells were incubated for 5 min at 4°C, then resuspended in 1 mL chilled wash buffer (10 mM Tris-HCl (pH7.4), 10mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 1% BSA, 1mM DTT, 1 U/μL RNase inhibitor) and centrifuged at 500 rcf for 5 min at 4°C. Pellet was rinsed two more times with 1 mL wash buffer, for a total of 3 washes. The nuclei pellet was resuspended in 40 μL Diluted Nuclei Buffer (1X Nuclei buffer (10X Genomics, PN-1000280), 1 mM DTT, 1 U/μL RNase inhibitor), and passed through a 40 μm Flowmi Cell Strainer (Merck).

Single nuclei library preparation and sequencing

Library preparation was performed at the GenomEast platform at the Institut de Génétique et de Biologie Moléculaire et Cellulaire. Transposition was performed using the 10X Genomics kit (PN-1000280) and nuclei were loaded into the 10X Genomics Chromium Controller with a target capture of 6,000 single nuclei. Libraries for the ATAC and 3′ mRNA were produced according to the Chromium Next GEM instructions (10X Genomics, Document CG000338 Rev F), and sequenced using the Illumina NextSeq 2000. Libraries were sequenced on an Illumina NextSeq 2000 sequencer as paired-end 28 + 85 base reads for 3′ mRNA, 50 base reads for ATAC. Image analysis and base calling were performed using RTA version 2.7.7 and BCL Convert version 3.8.4.

Data Processing

Raw data was processed using CellRanger (version 7.0.1), cellranger count option with the 10X reference genome mm10-2020-A, for demultiplexing, barcoded processing, and gene counting. Next we used velocyto (version 0.17.17, python implementation), run10x function along with the 10X reference genome to map the reads to intronic and exonic regions to obtain the unspliced and spliced quantification. We then used scVelo (version 0.2.3) for smoothing the unspliced and spiced quantifications with these hyperparameters: `min_shared_counts=20`, `n_pcs=30`, `n_neighbors=30`. Lastly We ran DeepCycle on the processed data to infer the cell cycle phases.

Biophysical model

We described the dynamics of the unspliced and the spliced mRNA for every gene with respect to the transcriptional phase during the cell cycle using a system of coupled ordinary differential equations,
Cell:

$$\frac{\mathrm{d}u(\theta)}{\mathrm{d}\theta} = \alpha(\theta) - \beta u(\theta)$$
$$\frac{\mathrm{d}s(\theta)}{\mathrm{d}\theta} = \beta u(\theta) - \gamma(\theta) s(\theta)$$

Nucleus:

$$\frac{\mathrm{d}u(\theta)}{\mathrm{d}\theta} = \alpha(\theta) - \beta u(\theta)$$
$$\frac{\mathrm{d}s(\theta)}{\mathrm{d}\theta} = \beta u(\theta) - \eta(\theta) s(\theta)$$

where α denotes the synthesis rate, β denotes the processing or the splicing rate, η denotes the nuclear export rate, and γ denotes the degradation rate. For every gene, we allow the synthesis, export, and degradation rates to explicitly depend on the cell cycle phase, θ. Since the dynamics in question is cyclic in nature, u, s, α , η, and γ will be periodic functions of θ. We used the Fourier approximation to solve this system of ordinary differential equations (details in supporting information).

Model fitting

The optimization was performed using the `optimize.minimize` package from `scipy` module in python. We used loss function to be minimized was,

$$\begin{aligned}\text{Loss} = \lambda_1 \frac{N}{2} &\left[\log \sum_{i=1}^{N} (u_i - u_i^{\text{mod}})^2 + \log \sum_{i=1}^{N} (s_i - s_i^{\text{mod}})^2\right] \\ &+ \lambda_2 \left[\sum_{i=1}^{N} \Theta(-\alpha)\cdot\alpha^2 + \sum_{i=1}^{N} \Theta(-\gamma)\cdot\gamma^2\right] \\ &+ \lambda_3 \left[\sum_{i=1}^{N} \Delta\theta_i \sqrt{1 + \left(\frac{\mathrm{d}\alpha}{\mathrm{d}\theta}\right)_i^2} + \sum_{i=1}^{N} \Delta\theta_i \sqrt{1 + \left(\frac{\mathrm{d}\gamma}{\mathrm{d}\theta}\right)_i^2}\right]\end{aligned}$$

where u_i and s_i denote the observed unspliced and spliced reads, whereas u_i^{mod} and s_i^{mod} denote the model predictions of the unspliced and spliced reads in the ith cell.
For every gene we calculated the explained variance,

$$r^2 = 1 - \frac{\sum_i^N (s_i - s_i^{\text{mod}})^2}{(s_i - \langle s \rangle)^2}$$

where $\langle s \rangle$ denotes the average spliced molecules across all the cells. We set a threshold of 20% and excluded all genes below this value.

Chromatin accessibility analysis

To compute chromatin accessibility dynamic profiles for each gene, we used the 10x gene annotation (Ensembl98) and extracted genomic regions from 2000 bp upstream of the TSS to the end of the gene. Then, The 10X "is-cell" barcode fragments were extracted and intersected with the above genomic regions with bedtools coverage tool (v2.30.0). This resulted in two matrices, one for each replicate, containing the number of ATAC fragments mapped to each gene in every cell. Sorting cells according to their cell cycle phase, θ, and smoothing the number of ATAC fragments across cells produced chromatin accessibility profiles for each gene during the cell cycle. Only genes showing high replicate correlation ($\rho > 0.75$) between the chromatin accessibility profiles in each replicate were retained.

To determine TF footprints within accessible chromatin, we initially identified putative TF binding sites in ATAC peaks. To achieve this, we run MoteEvo [49] with a set of known Position Weight Matrices (PWMs) of 678 TFs obtained from the SwissRegulon database [50]. We then calculated TF footprints for each cell by aggregating all observed Tn5 cuts at a given distance from the predicted binding site within a window of ± 1 Kbp. Cell cycle-dependent TF footprints were computed by smoothing single-cell data sorted according to their cell cycle phase, θ. Additional smoothing along the relative genomic coordinates with respect to the TF binding sites was performed. Finally, relative TF footprints were calculated by subtracting the average TF footprint across the cell cycle.

Resources

Processed data: 10.5281/zenodo.10462840
Github:
`https://github.com/MolinaLab-IGBMC/FourierCycle`
Web server:
`https://molina.igbmc.science/FourierCycle-Home.html`

Acknowledgments

We are grateful to Manuel Mendoza, Bertrand Séraphin, Georg Stoecklin, and Jingkui Wang for their thorough review and constructive criticism, which greatly improved the overall quality of this manuscript. This work of the Interdisciplinary Thematic Institute IMCBio+, as part of the ITI 2021-2028 program of the University of Strasbourg, CNRS and Inserm, was supported by IdEx Unistra (ANR-10-IDEX-0002), and by SFRI-STRAT'US project (ANR-20-SFRI-0012) and EUR IMCBio (ANR-17-EURE-0023) under the framework of the France 2030 Program. This work was also supported by a fellowship from L'Institut d'Études Avancées de l'Université de Strasbourg (USIAS). Library preparation and sequencing were performed by the GenomEast platform, a member of the France Génomique consortium (ANR-10-INBS-0009). Icons used in Figure 1 made by dDara (cell), Dimitry Mirolibov (DNA), Mayor Icons (anaphase), and Freeplk (cytokinesis) from `www.flaticon.com`.

Author contributions

MKN and NM conceptualized the study, DS and MC performed the experiments, MKN, AR, TY, and NM performed bioinformatics analyses, MKN and AR developed the code for biophysical modeling, OT developed the web server, MKN, DS, and NM wrote the manuscript, and NM led the research, secured funding, and provided supervision.

Author ORCiDs

Name	ORCiD
Maulik K. Nariya	0000-0001-6646-2353
David Santiago-Algarra	0000-0003-4697-7047
Olivier Tassy	0000-0002-2088-155X
Marie Cerciat	0000-0002-8085-1942
Tao Ye	0000-0002-3394-2083
Andrea Riba	0000-0002-6920-4592
Nacho Molina	0000-0003-0233-3055

References

[1] B. Alberts et al., *Molecular biology of the cell*, 6th edition (Garland Science, 2002).

[2] P Nurse, "A long twentieth century of the cell cycle and beyond", Cell **100** (2000).

[3] L. H. Hartwell et al., "Genetic control of the cell division cycle in yeast", Science **183** (1974).

[4] P. Nurse, "Genetic control of cell size at cell division in yeast", Nature **256** (1975).

[5] A. B. Pardee, "A restriction point for control of normal animal cell proliferation", Proceedings of the National Academy of Sciences **71** (1974).

[6] L. Hartwell, "Cell division from a genetic perspective", The Journal of cell biology **77** (1978).

[7] P. Nurse and Y. Bissett, "Gene required in g1 for commitment to cell cycle and in g2 for control of mitosis in fission yeast", Nature **292** (1981).

[8] E Rosenthal, "Selective translation of mrna controls the pattern of protein synthesis during early development of the surf clam, *Spisula solidissima*", Cell **20** (1980).

[9] S. Dalton, "Linking the cell cycle to cell fate decisions", Trends in Cell Biology **25** (2015).

[10] K. A. U. Gonzales et al., "Deterministic restriction on pluripotent state dissolution by cell-cycle pathways", Cell **162** (2015).

[11] L. Liu et al., "The cell cycle in stem cell proliferation, pluripotency and differentiation", Nature Cell Biology **21** (2019).

[12] S. Jirawatnotai et al., "Role of cyclins and cyclin-dependent kinases in pluripotent stem cells and their potential as a therapeutic target", Seminars in Cell & Developmental Biology **107** (2020).

[13] S. Pauklin and L. Vallier, "The cell-cycle state of stem cells determines cell fate propensity", Cell **155** (2013).

[14] N. Festuccia et al., "Mitotic bookmarking in development and stem cells", Development **144** (2017).

[15] N. Zielke and B. A. Edgar, "Fucci sensors: powerful new tools for analysis of cell proliferation", WIREs Developmental Biology **4** (2015).

[16] K. C. Palozola et al., "Mitotic transcription and waves of gene reactivation during mitotic exit", Science **358** (2017).

[17] S. Sarnataro et al., "Regulation of transcription reactivation dynamics exiting mitosis", PLOS Computational Biology **17**, edited by A. Csikász-Nagy (2021).

[18]A. Zeisel et al., "Coupled pre-mrna and mrna dynamics unveil operational strategies underlying transcriptional responses to stimuli", Molecular Systems Biology **7** (2011).

[19]D. Gaidatzis et al., "Erratum: analysis of intronic and exonic reads in rna-seq data characterizes transcriptional and post-transcriptional regulation", Nature Biotechnology **34** (2016).

[20]G. La Manno et al., "Rna velocity of single cells", Nature **560** (2018).

[21]V. Bergen et al., "Generalizing rna velocity to transient cell states through dynamical modeling", Nature Biotechnology **38** (2020).

[22]A. Riba et al., "Cell cycle gene regulation dynamics revealed by rna velocity and deep-learning", Nature Communications **13** (2022).

[23]P. Navarro, "2i, or not 2i: the soliloquy of nanog-negative mouse embryonic stem cells", Stem Cell Reports **11** (2018).

[24]S. Hastreiter et al., "Inductive and selective effects of gsk3 and mek inhibition on nanog heterogeneity in embryonic stem cells", Stem Cell Reports **11** (2018).

[25]K. Conklin, "10x genomics single-nucleus multiome (rna + atac) assay for profiling adult human tissues", (2022).

[26]S. Islam et al., "Quantitative single-cell rna-seq with unique molecular identifiers", Nature Methods **11** (2013).

[27]R. Kou et al., "Benefits and challenges with applying unique molecular identifiers in next generation sequencing to detect low frequency mutations", PLOS ONE **11**, edited by J. Wang (2016).

[28]O. Padovan-Merhar et al., "Single mammalian cells compensate for differences in cellular volume and dna copy number through independent global transcriptional mechanisms", Molecular Cell **58** (2015).

[29]N. R. Brown et al., "Cyclin b and cyclin a confer different substrate recognition properties on cdk2", Cell Cycle **6** (2007).

[30]E. T. Petri et al., "The crystal structure of human cyclin b", Cell Cycle **6** (2007).

[31]K. Bahar Halpern et al., "Nuclear retention of mrna in mammalian tissues", Cell Reports **13** (2015).

[32]D. J. Lew et al., "Isolation of three novel human cyclins by rescue of g1 cyclin (cln) function in yeast", Cell **66** (1991).

[33]M. Ohtsubo et al., "Human cyclin e, a nuclear protein essential for the g1-to-s phase transition", Molecular and Cellular Biology **15** (1995).

[34]N. Watanabe et al., "Regulation of the human wee1hu cdk tyrosine 15-kinase during the cell cycle.", The EMBO Journal **14** (1995).

[35]C. R. Elbæk et al., "Wee1 kinase limits cdk activities to safeguard dna replication and mitotic entry", Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **819–820** (2020).

[36]E. Yang et al., "Decay rates of human mrnas: correlation with functional characteristics and sequence attributes", Genome Research **13** (2003).

[37]L. V. Sharova et al., "Database for mrna half-life of 19, 977 genes obtained by dna microarray analysis of pluripotent and differentiating mouse embryonic stem cells", DNA Research **16** (2009).

[38]C. C.-S. Hsiung et al., "A hyperactive transcriptional state marks genome reactivation at the mitosis–g1 transition", Genes & Development **30** (2016).

[39]E. Y. Chen et al., "Enrichr: interactive and collaborative html5 gene list enrichment analysis tool", BMC Bioinformatics **14** (2013).

[40]Z. Fang et al., "Gseapy: a comprehensive package for performing gene set enrichment analysis in python", Bioinformatics **39** (2022).

[41]B. M. Smalec et al., "Genome-wide quantification of rna flow across subcellular compartments reveals determinants of the mammalian transcript life cycle", (2022).

[42]D. Lefaudeux et al., "Kinetics of mrna nuclear export regulate innate immune response gene expression", Nature Communications **13** (2022).

[43]L. Minnoye et al., "Chromatin accessibility profiling methods", Nature Reviews Methods Primers **1** (2021).

[44]A. T. Satpathy et al., "Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral t cell exhaustion", Nature Biotechnology **37** (2019).

[45]J. Heo et al., "Phosphorylation of tfcp2l1 by cdk1 is required for stem cell pluripotency and bladder carcinogenesis", EMBO Molecular Medicine **12** (2019).

[46]R. D. Everett et al., "Cell cycle regulation of pml modification and nd10 composition", Journal of Cell Science **112** (1999).

[47]L. Krenning et al., "Time-resolved single-cell sequencing identifies multiple waves of mrna decay during the mitosis-to-g1 phase transition", eLife **11** (2022).

[48]G. Gorin et al., "Rna velocity unraveled", PLOS Computational Biology **18** (2022).

[49]P. Arnold et al., "Motevo: integrated bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of dna sequences", Bioinformatics **28** (2011).

[50]M. Pachkov et al., "Swissregulon: a database of genome-wide annotations of regulatory sites", Nucleic Acids Research **35** (2007).

Supplementary figures

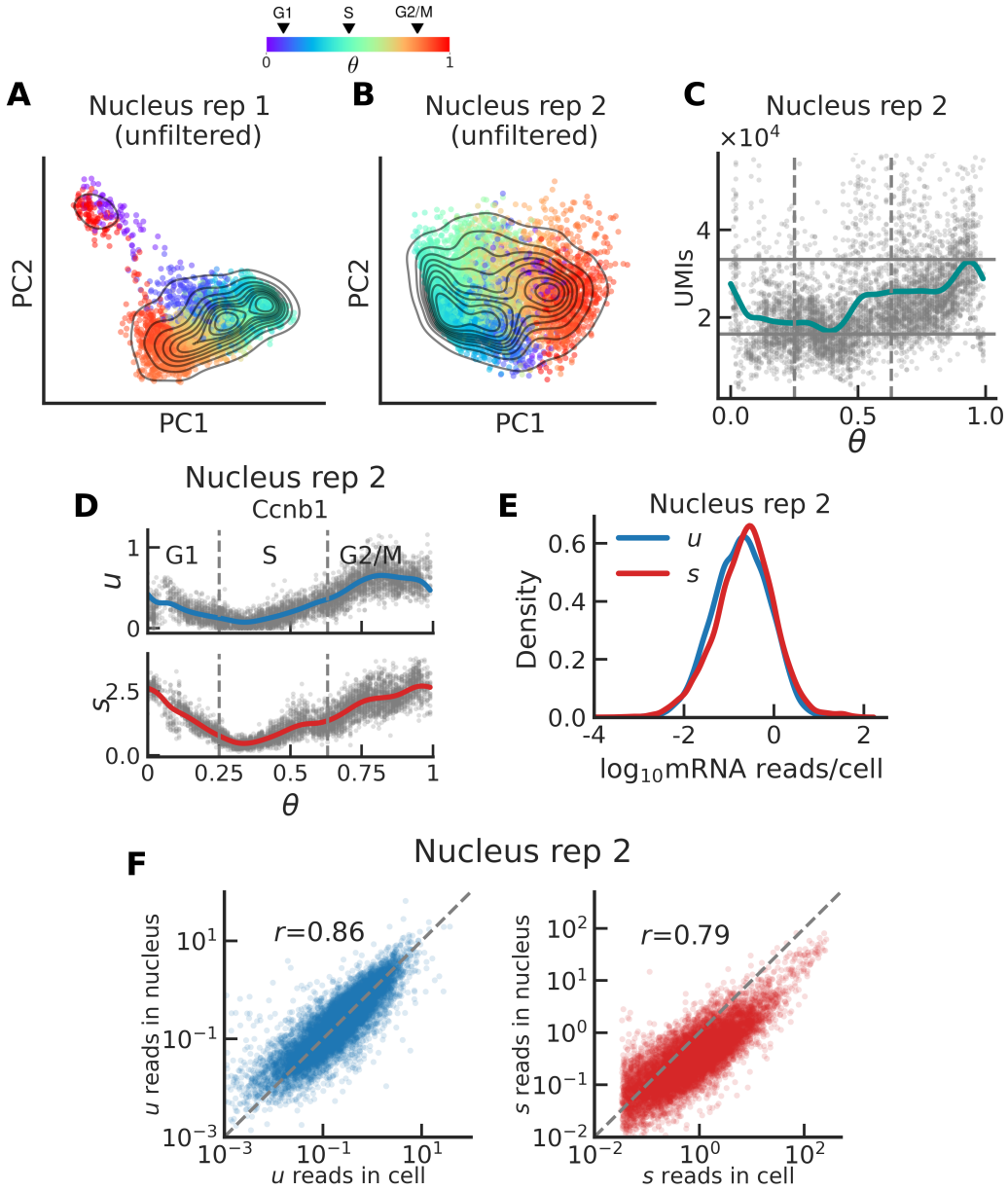

Figure S1 Data characteristics of single-nucleus sequencing data. **A** and **B:** Principal components analysis of unfiltered snRNAseq data. **C:** UMIs as a proxy for the total mRNA content in the nucleus (replicate 2), against the cell cycle phase θ. **D:** Expression dynamics of the unspliced and spliced reads for Nusap1 in the nucleus (replicate 2). **E:** Distribution of unspliced (blue) and spliced (red) reads per cell, across all genes in the nucleus (replicate 2). **F:** Correlation between the reads in the cell versus those in the nucleus (replicate 2) for the unspliced (blue) and spliced (red) molecules respectively.

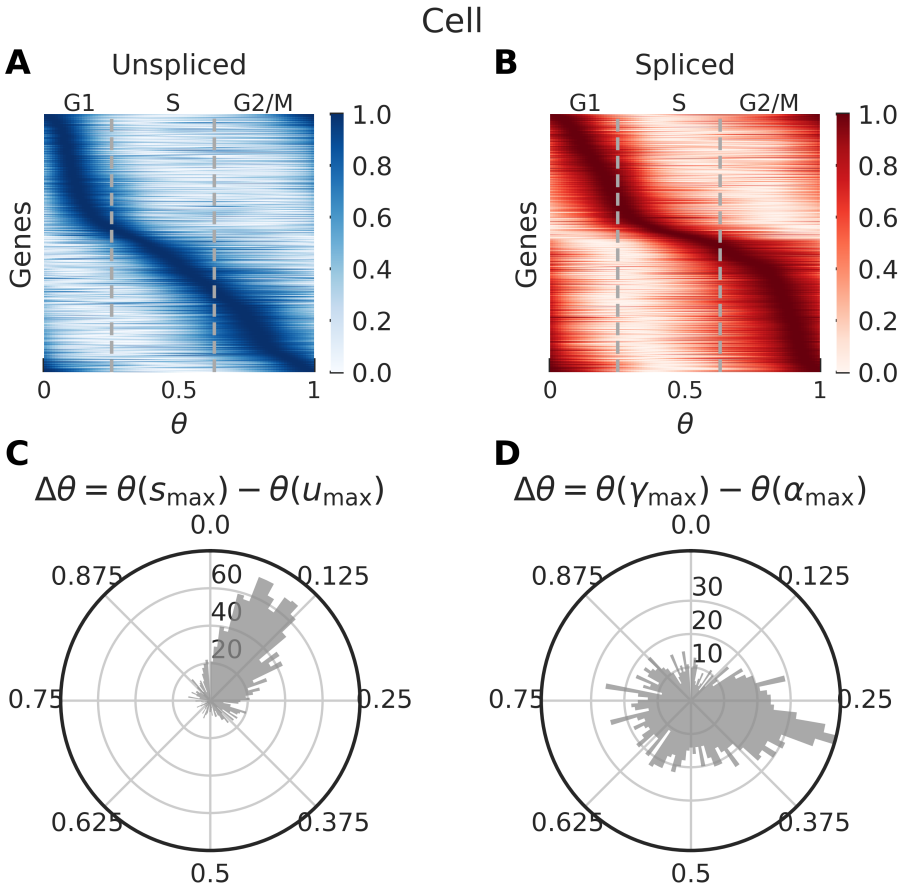

Figure S2 Heatmaps of gene expression and histograms of time differences for the cell. A and **B:** Heatmaps of normalized, using max-min normalization, of unspliced and spliced quantifications of the genes filtered based on the explained variance cutoff. **C:** Polar histograms of time difference between maximum unspliced and maximum spliced **D:** Polar histograms of time difference between maximum transcription and maximum degradation for scRNA-seq data.

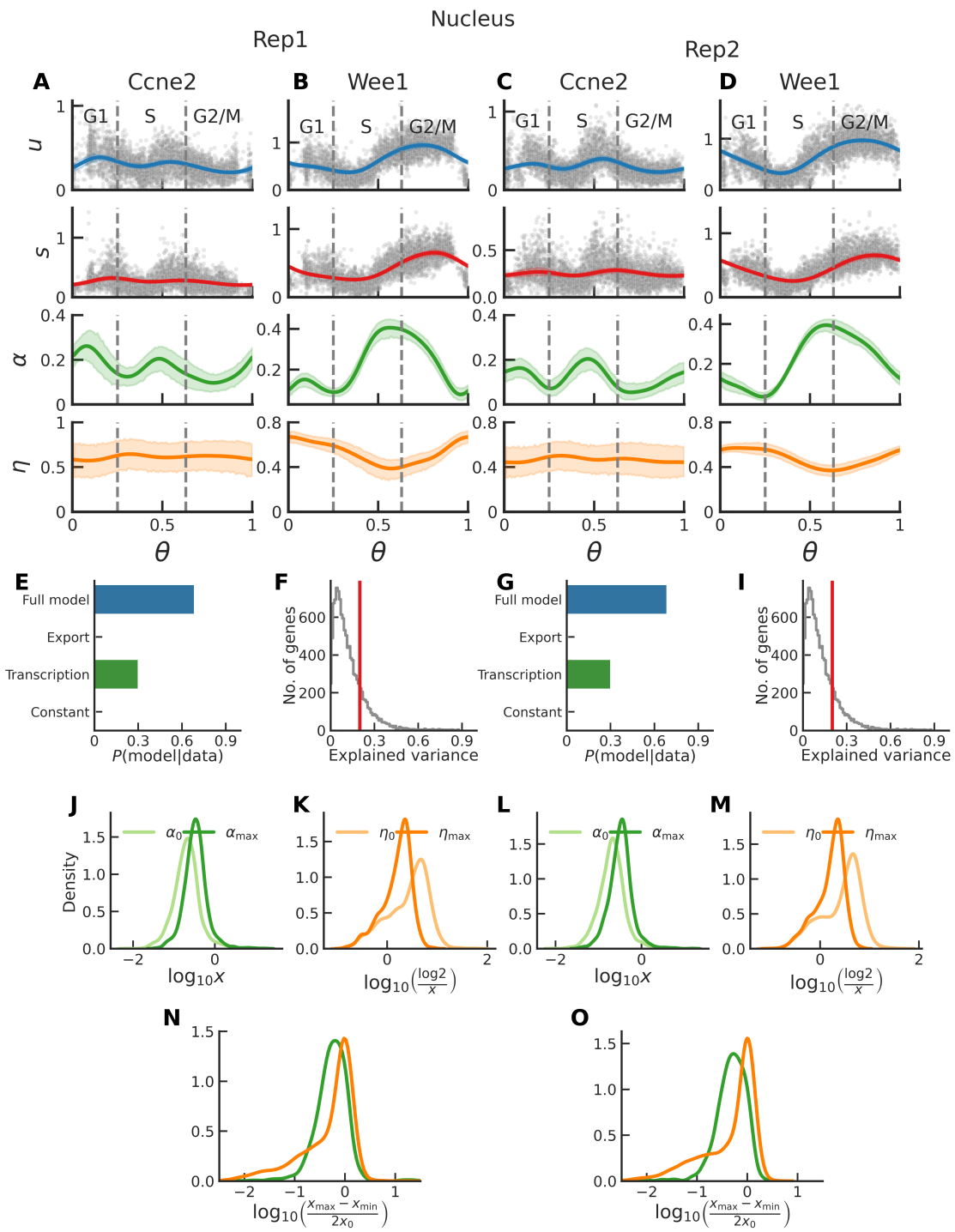

Figure S3 Kinetic modeling results for single-nucleus sequencing data. **A-D:** Number of unspliced reads (u), number of spliced reads (s), transcription rate (α), export rate (η) as function of the cell cycle phase θ for representative genes Ccne2 and Wee1 (multiome snRNAseq data replicates 1 and 2). **E** and **G:** Barplot of fraction of genes selecting alternative models based on the Bayesian Information Criterion for replicates 1 and 2 respectively. **F** and **H:** Histogram of the explained variance in spliced counts across all the genes for replicate 1 and 2 respectively; red line is at $x = 0.2$, the threshold set for filtering genes. **J** and **L** Distribution of transcription rate, in arbitrary units, acrsr/share/texlive/texmf-dist/fonts/type1/public/amsfonts/cm/cmtt10.pfb¿¡/usr/share/texmf/fonts/type1/public/cm-super/sfrm1000.pfb¿ Output written on fouriercycle.pdf (26 pages, 27381514 bytes). Transcript written on fouriercycle.log. (oss filtered genes, α_0 denotes baseline transcription (light green) whereas $\alpha_{\max}$ denotes maximum transcription rate (green) for replicates 1 and 2 respectively. **K** and **M:** Distribution of export half lives, in hours, across filtered genes, η_0 denotes baseline export rate (light orange) whereas $\eta_{\max}$ denotes maximum export rate (orange) for replicates 1 and 2 respectively.

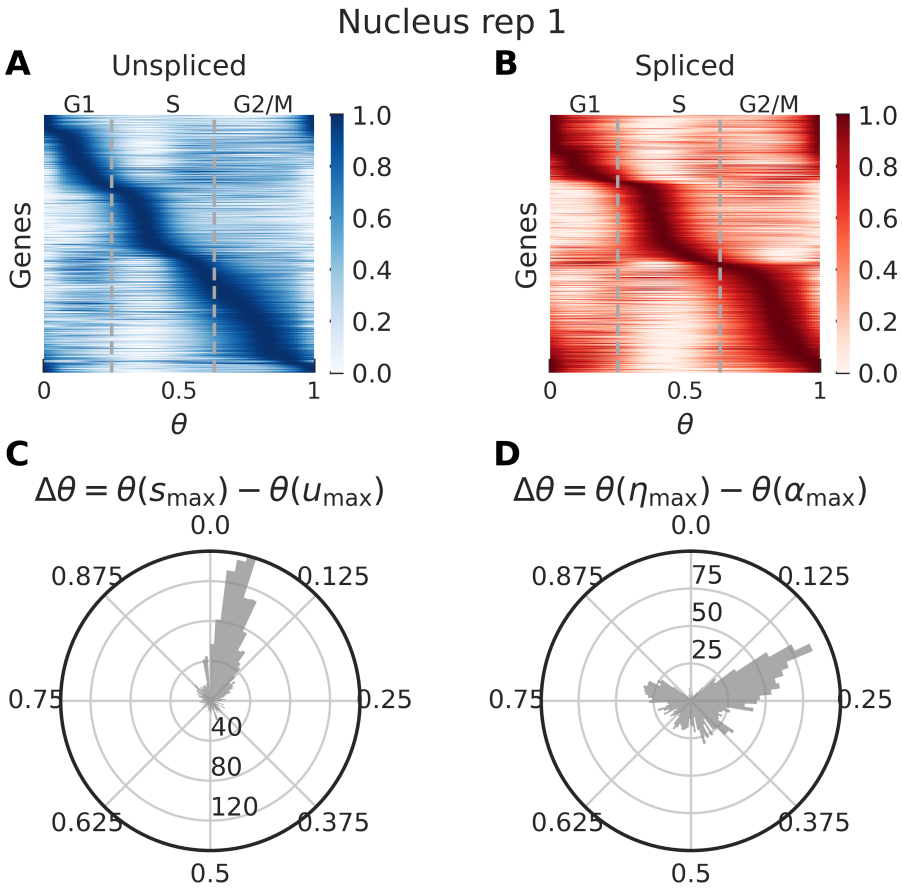

Figure S4 Heatmaps of gene expression and histograms of time differences for the cell. A and **B:** Heatmaps of normalized, using max-min normalization, unspliced and spliced of the genes filtered based on the explained variance cutoff. **C:** Polar histograms of time difference between maximum unspliced and maximum spliced. **D:** Polar histograms of time difference between maximum transcription and maximum export for snRNA-seq data.

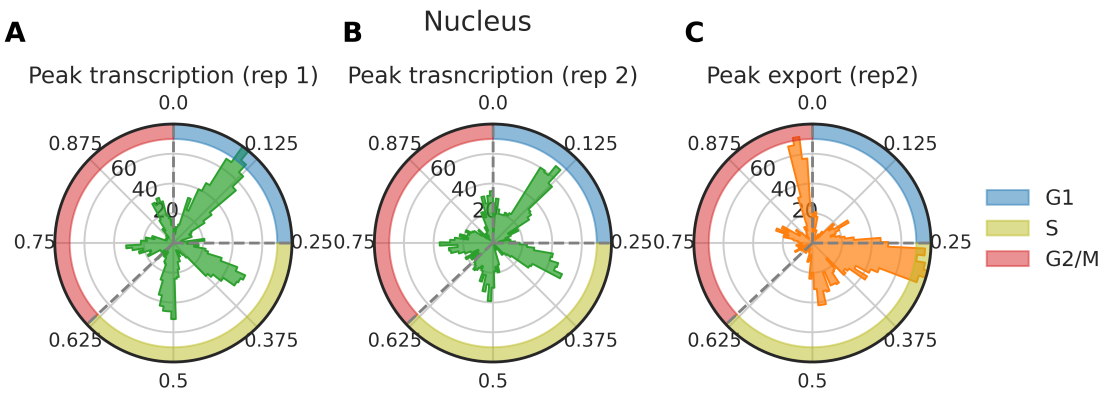

Figure S5 Waves of transcription and export in the nucleus. A and **B:** Polar histograms for timing of peak transcription for snRNAseq data (replicates 1 and 2) and **C:** Polar histograms of timing of peak export (replicate 2, replicate 1 is in Figure 5D)

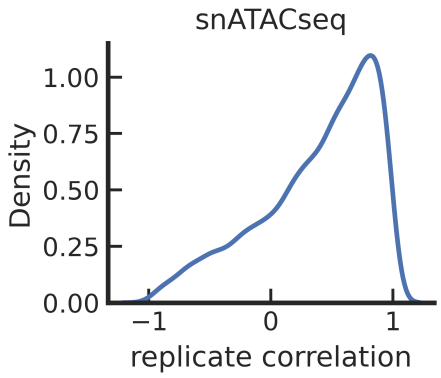

Figure S6 Replicate correlation for snATACseq data

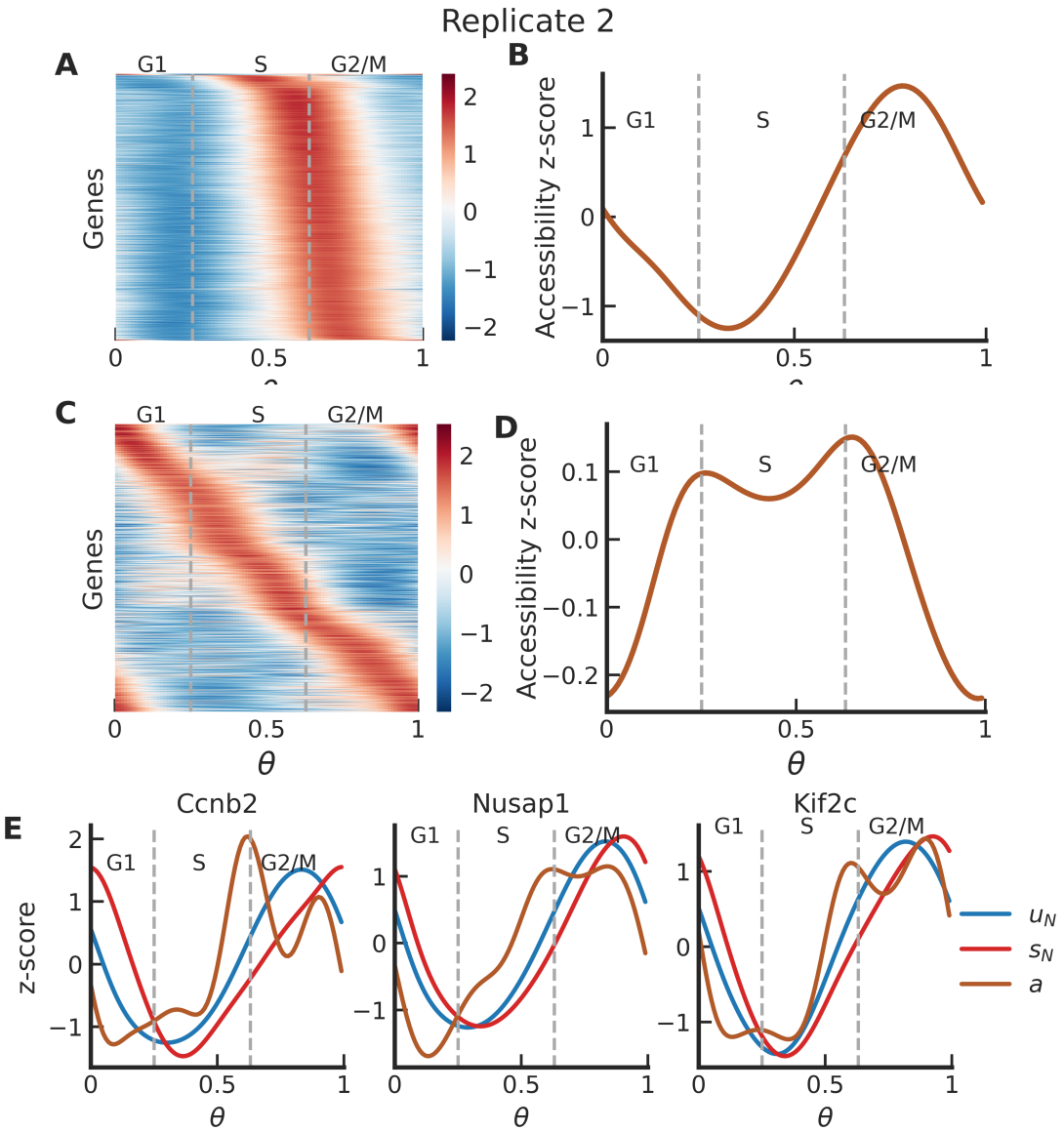

Figure S7 Chromatin accessibility during the cell cycle. Same as panels in figure 6 for replicate 2.

Supplementary information: Fourier solution for mRNA metabolism

Model

We model the dynamics of mRNA metabolism by expressing the steps involved—transcription, splicing, degradation—as a system of coupled ordinary differential equations.

$$\frac{\mathrm{d}u(t)}{\mathrm{d}t} = \alpha(t) - \beta u(t) \tag{1}$$

$$\frac{\mathrm{d}s(t)}{\mathrm{d}t} = \beta u(t) - \gamma(t)s(t) \tag{2}$$

α: synthesis rate
β: splicing rate
γ: degradation rate

Unspliced solution

We assume that $u(t)$ and $\alpha(t)$ are periodic functions. Thus, we can express $u(t)$ and $\alpha(t)$ as sums of Fourier series.

$$u(t) = \sum_{n=-\infty}^{+\infty} u_n e^{\frac{i2\pi nt}{P}} \tag{3}$$

$$\alpha(t) = \sum_{n=-\infty}^{+\infty} \alpha_n e^{\frac{i2\pi nt}{P}}, \tag{4}$$

wehre P : the period of oscillation. Let $\theta = \frac{2\pi t}{P}$, such that $\theta \in (0, 1)$

Thus, we have

$$u(\theta) = \sum_{n=-\infty}^{+\infty} u_n e^{in\theta}$$

$$\frac{\mathrm{d}u(\theta)}{\mathrm{d}t} = \sum_{n=-\infty}^{+\infty} inu_n e^{in\theta}$$

$$\alpha(\theta) = \sum_{n=-\infty}^{+\infty} \alpha_n e^{in\theta}$$

Using these in equation (1), we get,

$$\sum_{n=-\infty}^{+\infty} inu_n e^{in\theta} = \sum_{n=-\infty}^{+\infty} \alpha_n e^{in\theta} - \beta \sum_{n=-\infty}^{+\infty} u_n e^{in\theta}$$

$$\sum_{n=-\infty}^{+\infty} \left(iu_n n - \alpha_n + \beta u_n\right) e^{in\theta} = 0$$

$$iu_n n - \alpha_n + \beta u_n = 0$$

$$u_n = \frac{\alpha_n}{\beta + in}$$

The Fourier approximation for $u(\theta)$

$$u(\theta) = \sum_{n=-\infty}^{+\infty} \left(\frac{\alpha_n}{\beta + in}\right) e^{in\theta} \tag{5}$$

When performing numerical optimization, it is convenient to have real-valued parameters, *i.e.* Fourier coefficients

for $\alpha(\theta)$ and $\gamma(\theta)$. Fourier series for α with real-values co-efficients can be written as

$$\alpha(\theta) = \alpha_0 + \sum_{n=1}^{N} \alpha_n^A + \cos(n\theta) + \alpha_n^B \sin(n\theta)$$

We can recover the complex co-efficients

$$\alpha_n = \frac{\alpha_n^A + i\alpha_n^B}{2} \qquad \text{(for } n < 0\text{)}$$
$$\alpha_n = \frac{\alpha_n^A - i\alpha_n^B}{2} \qquad \text{(for } n > 0\text{)}$$

Spliced solution

Akin to $u(\theta)$ and $\alpha(\theta)$, we expect $s(\theta)$ and $\gamma(\theta)$ also to be periodic functions. The Fourier series expansion of s and γ (as a function of θ) is,

$$s(\theta) = \sum_{n=-N}^{N} s_n e^{in\theta}$$
$$\gamma(\theta) = \sum_{n=-N}^{N} \gamma_n e^{in\theta}$$

Using these expressions in (2),

$$\sum_{n=-N}^{N} ins_n e^{in\theta} = \beta \sum_{n=-N}^{N} u_n e^{in\theta} - \sum_{n=-N}^{N} \sum_{m=-N}^{N} s_n \gamma_m e^{i(m+n)\theta} \tag{6}$$

In the last term in the above equation, let $k = m + n$

$$\begin{aligned}
\sum_{n=-N}^{N} \sum_{m=-N}^{N} s_n \gamma_m e^{i(m+n)\theta} &= \sum_{n=-N}^{N} \sum_{m=-N}^{N} \sum_{k=-N}^{N} s_n \gamma_m \delta_{n+m,k} e^{ik\theta} \\
&= \sum_{k=-N}^{N} \sum_{m=-N}^{N} \sum_{n=-N}^{N} s_k \gamma_m \delta_{k+m,n} e^{in\theta} \\
&= \sum_{n=-N}^{N} \sum_{m=-N}^{N} \sum_{k=-N}^{N} s_k \gamma_m \delta_{k+m,n} e^{in\theta}
\end{aligned}$$

Using this in (6)

$$\begin{aligned}
&\sum_{n=-N}^{N} ins_n e^{in\theta} = \beta \sum_{n=-N}^{N} u_n e^{in\theta} - \sum_{n=-N}^{N} \sum_{m=-N}^{N} \sum_{k=-N}^{N} s_k \gamma_m \delta_{k+m,n} e^{in\theta} \\
&ins_n = \beta u_n - \sum_{m=-N}^{N} \sum_{k=-N}^{N} s_k \gamma_m \delta_{k+m,n} \\
&\sum_{k=-N}^{N} \left[\delta_{k,n} in + \sum_{m=-N}^{N} \gamma_m \delta_{k+m,n} \right] s_k = \beta u_n \\
&\sum_{k=-N}^{N} G_{nk} s_k = \beta u_n
\end{aligned}$$

We can express this as a matrix equation

$$\mathbf{s} = \beta \mathbf{G}^{-1} \mathbf{u} \tag{7}$$

where $\mathbf{G}$ is a $(2N+1) \times (2N+1)$ matrix:

$$\begin{pmatrix} \gamma_0 - Ni & \gamma_{-1} & \cdots & \gamma_{-N} & 0 & \cdots & 0 & 0 \\ \gamma_1 & \gamma_0 - (N-1)i & \cdots & \gamma_{-(N-1)} & \gamma_{-N} & \cdots & 0 & 0 \\ \gamma_2 & \gamma_1 & \cdots & \gamma_{-(N-2)} & \gamma_{-(N-1)} & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \gamma_{N-2} & \gamma_{N-3} & \cdots & \gamma_{-1} & \gamma_{-2} \\ 0 & 0 & \cdots & \gamma_{N-1} & \gamma_{N-2} & \cdots & \gamma_0 + (N-1)i & \gamma_{-1} \\ 0 & 0 & \cdots & \gamma_N & \gamma_{N-1} & \cdots & \gamma_1 & \gamma_0 + Ni \end{pmatrix} \tag{8}$$

Model fitting

The optimization was performed using `scipy.optimize.minimize`, we used the following loss function for the optmization,

$$\begin{aligned} \text{Loss} = \lambda_1 \frac{N}{2} \left[\log \sum_{i=1}^{N} (u_i^{\text{obs}} - u_i^{\text{mod}})^2 + \log \sum_{i=1}^{N} (s_i^{\text{obs}} - s_i^{\text{mod}})^2 \right] + \lambda_2 \left[\sum_{i=1}^{N} \Theta(-\alpha) \cdot \alpha^2 + \sum_{i=1}^{N} \Theta(-\gamma) \cdot \gamma^2 \right] \\ + \lambda_3 \left[\sum_{i=1}^{N} \Delta\theta_i \sqrt{1 + \left(\frac{\mathrm{d}\alpha}{\mathrm{d}\theta} \right)_i^2} + \sum_{i=1}^{N} \Delta\theta_i \sqrt{1 + \left(\frac{\mathrm{d}\gamma}{\mathrm{d}\theta} \right)_i^2} \right] \end{aligned}$$