

## METHODS

# Storm: Incorporating transient stochastic dynamics to infer the RNA velocity with metabolic labeling information

Qiangwei Peng<sup>1</sup>, Xiaojie Qiu<sup>2,3,4,5\*</sup>, Tiejun Li<sup>1,6\*</sup>

**1** LMAM and School of Mathematical Sciences, Peking University, Beijing, China, **2** Department of Genetics, Stanford University School of Medicine, Stanford, California, United States of America, **3** Basic Sciences and Engineering Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford, California, United States of America, **4** Department of Computer Science, Stanford University, Stanford, California, United States of America, **5** Stanford Cardiovascular Institute, Stanford University, Stanford, California, United States of America, **6** Center for Machine Learning Research, Peking University, Beijing, China

\* [xiaojie@stanford.edu](mailto:xiaojie@stanford.edu) (XQ); [tieli@pku.edu.cn](mailto:tieli@pku.edu.cn) (TL)



## OPEN ACCESS

**Citation:** Peng Q, Qiu X, Li T (2024) Storm: Incorporating transient stochastic dynamics to infer the RNA velocity with metabolic labeling information. PLoS Comput Biol 20(11): e1012606. <https://doi.org/10.1371/journal.pcbi.1012606>

**Editor:** Laleh Haghverdi, Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft, GERMANY

**Received:** January 9, 2024

**Accepted:** November 3, 2024

**Published:** November 18, 2024

**Copyright:** © 2024 Peng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** In this study, we used the following public tscRNA-seq datasets. The primary human HSPCs datasets from scNT-seq are available from the Gene Expression Omnibus (GEO) with accession number GSE193517, and can be downloaded from python package Dynamo (<https://figshare.com/ndownloader/files/47439626>). The neuronal activity under KCl polarization datasets from scNT-seq are available from the GEO with accession number GSE141851, and can be downloaded from python package Dynamo (<https://figshare.com/ndownloader/files/>

## Abstract

The time-resolved scRNA-seq (tscRNA-seq) provides the possibility to infer physically meaningful kinetic parameters, e.g., the transcription, splicing or RNA degradation rate constants with correct magnitudes, and RNA velocities by incorporating temporal information. Previous approaches utilizing the deterministic dynamics and steady-state assumption on gene expression states are insufficient to achieve favorable results for the data involving transient process. We present a dynamical approach, Storm (Stochastic models of RNA metabolic-labeling), to overcome these limitations by solving stochastic differential equations of gene expression dynamics. The derivation reveals that the new mRNA sequencing data obeys different types of cell-specific Poisson distributions when jointly considering both biological and cell-specific technical noise. Storm deals with measured counts data directly and extends the RNA velocity methodology based on metabolic labeling scRNA-seq data to transient stochastic systems. Furthermore, we relax the constant parameter assumption over genes/cells to obtain gene-cell-specific transcription/splicing rates and gene-specific degradation rates, thus revealing time-dependent and cell-state-specific transcriptional regulations. Storm will facilitate the study of the statistical properties of tscRNA-seq data, eventually advancing our understanding of the dynamic transcription regulation during development and disease.

## Author summary

Intricate regulation of RNA biogenesis, such as RNA transcription, splicing and degradation, plays a critical role in most biological processes. Previous approaches have leveraged a deterministic model of spliced and unspliced RNAs to estimate kinetic parameters and quantify RNA velocity, the rate of changes in gene expression states in single cells. However, accurately estimating meaningful kinetic parameters and RNA velocity is hindered

([47439629](#)). The sci-fate dataset are available from the GEO with accession number GSE131351. The PerturbSci-Kinetics datasets are available from the GEO with accession number GSE218566. The scSLAM-seq dataset are available from the GEO with accession number GSE115612. The murine intestinal organoid system dataset and cell cycle datasets from scEU-seq are available from the GEO with accession number GSE128365, and can be downloaded from python package Dynamo (<https://figshare.com/ndownloader/files/47439632> and <https://figshare.com/ndownloader/files/47439641>). Code availability Storm is implemented in Python and is available at <https://github.com/aristoteleo/CSP4ML>.

**Funding:** TL and QP acknowledge the support from NSFC and MSTC under Grant No.s 11825102, 12288101 and National Key R&D Program of China, Grant 2021YFA1003301. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors declare that they have no competing interests.

by biased capture of unspliced RNA and absence of temporal information in conventional scRNA-seq. Significant advances have been made in better measuring RNA kinetics with metabolic labeling enabled scRNA-seq, however, computational tools to analyze them lag far behind. Prior work of Dynamo provides one of the early solutions to properly model RNA metabolic labeling data, but its method still largely uses a deterministic model that only utilizes part of extreme cells and is unable to analyze datasets with significant transient dynamics. To address these challenges, we developed Storm that explicitly models transient stochastic RNA dynamics. Importantly, Storm models RNA kinetics with stochastic differential equations that explicitly account for biological and cell-specific noises. Storm is generally applicable to many metabolic labeling scRNA-seq datasets and we demonstrate excellent performance of Storm in fitting the data to capture the transient dynamics under various noise models.

## Introduction

Cells are dynamic identities that are subject to intricate transcriptional and post-transcriptional regulations. Understanding the tight regulation of the RNA life cycle will shed light on not only the regulatory mechanism of RNA biogenesis, but also cell fate transitions. Based on the observation that most scRNA-seq approaches capture both premature unspliced mRNA and mature spliced mRNA information, La Manno et al. [1] pioneered the concept of RNA velocity or the time derivative of spliced RNA to reveal the local fate of each individual and designed an RNA kinetic parameter inference method called velocyto based on the steady state assumption. In a later work, scVelo [2] relaxed the steady-state assumption and proposed a dynamic RNA velocity model to infer gene-specific reaction rates of transcription, splicing and degradation as well as cell-specific hidden time using the expectation-maximization (EM) algorithm. Li et al. [3, 4] derived a stochastic model of RNA velocity based on the chemical master equation (CME) satisfied by the probability mass function (PMF) rather than the deterministic ordinary differential equation (ODE) satisfied by the mean, and presented a mathematical analysis framework of RNA velocity. More general studies on the analytical solution of the CME for monomolecular reaction systems are also discussed in [5]. In addition, a rigorous and detailed analysis of the entire workflow for RNA velocity is also provided in [6]. MultiVelo [7] extends the dynamic RNA velocity model by incorporating epigenome data that can be jointly measured with emerging multi-omics approaches. Protaccel [8] extends the concept of RNA velocity to protein. UniTVelo [9] uses a top-down design for more flexible estimation of the RNA velocity, as opposed to the usual bottom-up strategy. DeepVelo [10] uses graph convolutional neural networks to infer cell-specific parameters to extend RNA velocity to cell populations containing time-dependent dynamics and multiple lineages which were proven to be challenging in previous methods [11]. Other deep learning-based approaches include VeloAE [12], VeloVI [13], VeloVAE [14], LatentVelo [15], cellDancer [16], and so on. However, due to the absence of physical time information, the above methods usually suffer the issue of scale invariance, that is, amplifying the parameters by an arbitrary amount will not change the solution if the time shrinks with the same amount, e.g., the exact physical time remains undetermined. This issue makes the inferred parameters and the RNA velocity have physical significance only up to a multiplicative constant [3]. In addition, the missing time information enters the model as hidden variables, which makes the parameter inference difficult.

Technological innovations in scRNA-seq now enable us to directly measure the amount of newly synthesized mRNA molecules over a short period of time, either through chemically

introduced mutations in the sequencing reads or direct biotin pull-down of RNA analogs such as 4sU metabolically labeled RNA molecules, which subtly introduces physical time information. These time-resolved metabolic labeling–augmented scRNA-seq (tscRNA-seq) include scSLAM-seq [17], scNT-seq [18], sci-fate [19], NASC-seq [20], scEU-seq [21], PerturbSci-Kinetics [22] and others [23–25]. Qiu et al. [26] recently developed Dynamo to reconstruct analytical vector fields from discrete RNA velocity vectors by taking advantage of tscRNA-seq data to infer more robust and time-resolved RNA velocity, however, they only used the deterministic model and largely relied on the steady-state assumption. CellRank2 [27] also focuses on pulse and chase tscRNA-seq data, but ignores the broader one-shot tscRNA-seq data, and relies on deterministic models for parameter inference.

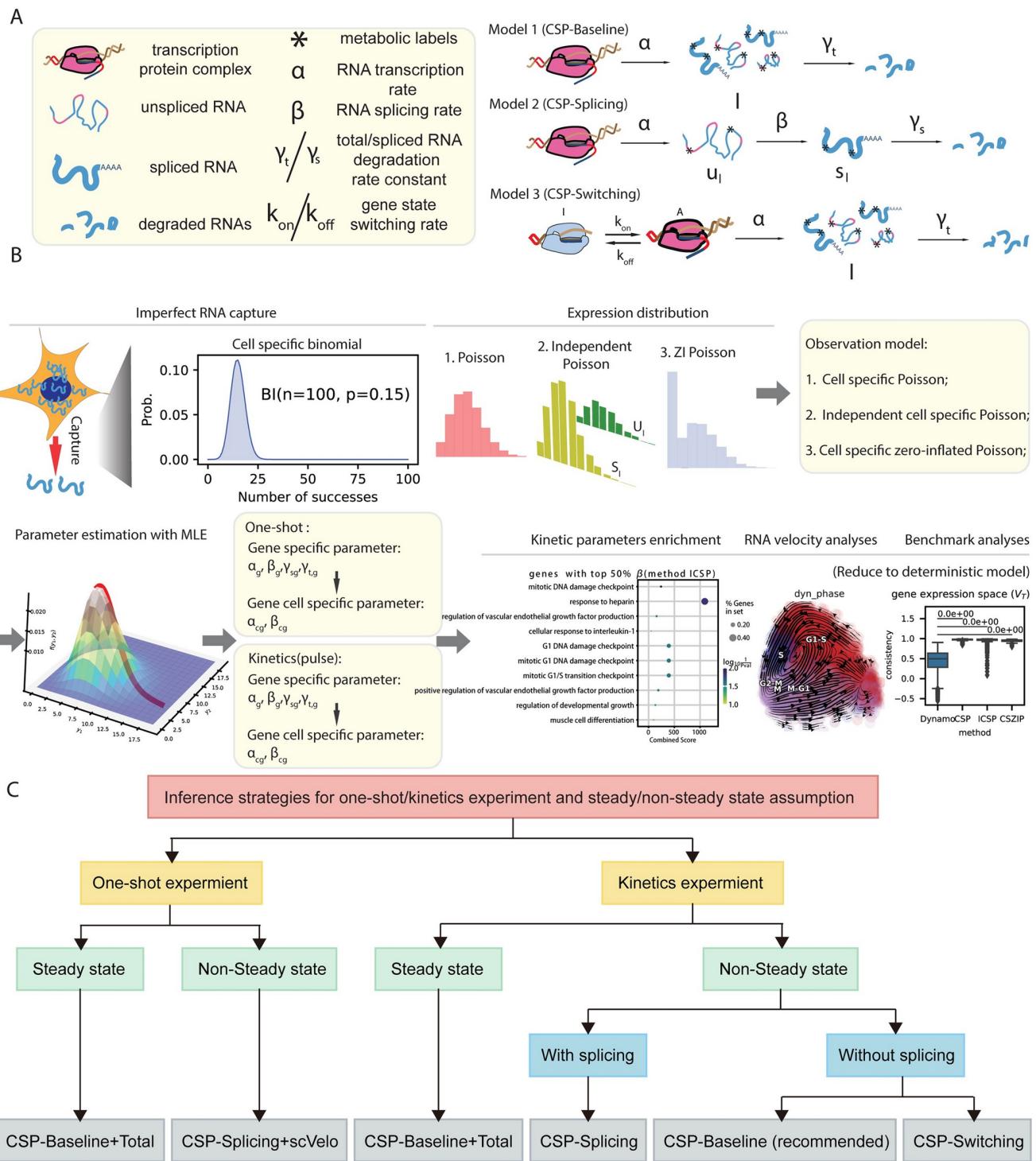
To overcome the shortcomings of Dynamo and fully explore the potential of tscRNA-seq data, we present the Storm approach (Stochastic models of RNA metabolic-labeling) to improve the estimation of RNA kinetic parameters and the inference of the RNA velocity of the metabolic labeling scRNA-seq data by incorporating the transient stochastic dynamics of gene expressions. Importantly, we focus on modeling the kinetics/pulse metabolic labeling data as it follows the RNA synthesis across multiple short time periods and is thus ideal for capturing temporal RNA kinetics. In order to properly model both biological noise and cell-specific technical noise (due to the variations in sequencing depth across individual cells and dropout resulting from imperfect RNA capture in scRNA-seq), we implemented in Storm three stochastic models of new mRNA (or new unspliced and spliced mRNA). Depending on the biological processes considered, Storm indicates that new mRNA sequencing data obeys different types of cell-specific Poisson (CSP) distributions. On this basis, Storm also includes hypothesis testing, parameter inference and goodness of fit evaluation methods for CSP-type distribution. In addition, we analyze the similarities and differences of the model considering RNA splicing or not. For one-shot data containing both unspliced unlabeled (uu), unspliced labeled (ul), spliced unlabeled (su) and spliced labeled (sl) RNA, we devise a two-stage parameter inference method that does not rely on steady-state assumption to infer the absolute magnitude of the kinetic parameters. For one-shot data containing only new RNA and total RNA, we introduce the steady-state assumption to make the parameter inference possible. We verified the effectiveness of Storm in the cell cycle data set of kinetic experiments from the scEU-seq study [21] and several one-shot datasets, including scSLAM-seq, scNT-seq, sci-fate and PerturbSci-Kinetics. Storm is incorporated in Dynamo [26] of the Aristotle ecosystem that facilitates rich downstream analytical vector field modeling.

## Results

### Overall description of Storm

We established three stochastic gene expression models for new mRNA (or new unspliced and spliced mRNA) ([Fig 1A](#)) for the inference of the RNA kinetic parameters and thus the RNA velocity in the Storm approach. In Model 1, only transcription and mRNA degradation were considered. In Model 2, we considered transcription, splicing, and spliced mRNA degradation. And in Model 3, we considered the switching of gene expression states, transcription in the active state, and mRNA degradation.

The complete workflow of Storm is demonstrated in [Fig 1B](#). We first analytically solve the new RNA (or new unspliced and spliced mRNA) stochastic dynamics corresponding to the above three models, which are Poisson distribution, independent Poisson distribution and zero-inflated Poisson distribution, respectively. In addition, we model the technical noise as the cell-specific binomial distribution. By integrating the biological noise and the technical noise together, we obtain the distribution for the measured number of new/



**Fig 1. Schematic overview of Storm.** **A.** Three models of RNA life cycle considering different biological processes: **Model 1 (CSP-Baseline)**: Reaction dynamics model for new RNA  $l(t)$  ignoring the splicing process, where  $\alpha$  is the transcription rate and  $\gamma_t$  is the total mRNA degradation rate. **Model 2 (CSP-Splicing)**: Reaction dynamics model of new unspliced and spliced mRNA ( $u_l(t)$ ,  $s_l(t)$ ) considering the splicing process, where  $\beta$  is the splicing rate,  $\gamma_s$  is the spliced mRNA degradation rate, and  $\alpha$  is the same as **Model 1**. **Model 3 (CSP-Switching)**: Reaction dynamics model of new RNA  $l(t)$  considering gene state switching, where  $\alpha$  and  $\gamma_t$  are the same as in **Model 1**,  $k_{on}$  is the rate at which the gene switches from the inactive state to the active state,  $k_{off}$  is the opposite. **B.** Complete workflow diagram for parameter inference and downstream analysis based on stochastic dynamics of new mRNA considering technical noise. **C.** Specific parameter inference strategies for one-shot/pulse experiments and steady-state/non-steady-state assumption.

<https://doi.org/10.1371/journal.pcbi.1012606.g001>

labeled mRNA molecules (or new unspliced and spliced mRNA molecules), which are cell-specific Poisson distribution, independent cell-specific Poisson distribution and cell-specific zero-inflated Poisson distribution, respectively. Therefore, we call the three models CSP-Baseline, CSP-Splicing, and CSP-Switching for easy memory and use them later to better distinguish these three models. Maximum likelihood estimation (MLE) is used to fit the data and make inferences for the parameters shown in the corresponding models.

To ensure the general applicability of Storm in common nascent RNA labeling schemes, such as one-shot or kinetics/pulse experiments (See Figure 2 of Qiu, et. al [26]), we designed specific estimation strategies for each labeling scheme (Fig 1C). For the one-shot labeling experiments with only new RNA and total RNA data, since there is only one labeling duration and the lack of splicing data, the steady-state assumption under the stochastic dynamics framework is reinvoked to infer parameters (Fig 1C Left). For the one-shot labeling experiments with  $u_u$ ,  $u_l$ ,  $s_u$  and  $s_l$  RNA, we design a two-stage approach that does not rely on the steady-state assumption. More specifically, we first use scVelo [2] to determine the relative size of the kinetic parameters, and then use CSP-Splicing to determine the absolute size (Fig 1C Left). For kinetics/pulse-labeling experiments with multiple labeling durations, the transient stochastic dynamics framework is used without the steady-state assumption (Fig 1C Right). CSP-Baseline is recommended if the concern is RNA velocity, and CSP-Splicing should be used if the concern also includes splicing dynamics.

Although steady-state assumption can also be included, we recommend that non-steady-state approach should be used unless the user has sufficient knowledge of the biological process being studied. Furthermore, the goodness-of-fit index based on deviance commonly used in generalized linear models is utilized to quantify the goodness of fit of our models in kinetics/pulse datasets. The index is then used to select genes that are more consistent with model assumptions for later downstream analysis, such as the enrichment analysis of different gene-specific parameters. Furthermore, we relaxed the previous assumption of constant parameters in genes or cells and assumed that only degradation rates ( $\gamma_t$  in CSP-Baseline and CSP-Switching;  $\gamma_s$  in CSP-Splicing) are constant while the other parameters are cell-specific and depend on the state of gene expression in each cell. This relaxation would be useful for modeling lineage-specific kinetics resulted from hierarchical lineage bifurcation, which is common in cell development. Finally, in order to calculate and visualize the RNA velocity, we reduced the considered stochastic models to derive the deterministic equation for the mean gene expression. The inferred parameters, after filtering with the goodness-of-fit index are then used in RNA velocity analysis and visualization. Notably, to demonstrate Storm's performance, we conducted systematic comparison with the state-of-the-art method Dynamo [26] for processing metabolic labeling scRNA-seq experiment datasets.

In the continued subsections we will present the details of each step in the Storm workflow, starting from the introduction of our mathematical models.

## CSP modeling of counts data with metabolic labeling information

We proposed and analytically solved three aforementioned stochastic gene expression models for the dynamics of new mRNAs (or new unspliced and spliced mRNAs).

For simplicity of modeling, we followed [1, 2] to assume that the genes are independent. In the stochastic gene expression model, the generation of new/labeled mRNA  $\tilde{l}(t)$  (or new unspliced and spliced mRNA  $(\tilde{u}_l(t), \tilde{s}_l(t))$ ) is a stochastic process, and we are interested in the

evolution of its PMF over time, which is denoted by

$$\begin{aligned}\tilde{P}_n(t) &:= \text{Prob}(\tilde{l}(t) = n), \quad n \in \mathbb{N} \\ \tilde{P}_{mn}(t) &:= \text{Prob}((\tilde{u}_l(t), \tilde{s}_l(t)) = (m, n)), \quad (m, n) \in \mathbb{N}^2.\end{aligned}\tag{1}$$

In CSP-Baseline and CSP-Splicing, Since the initial value of  $\tilde{l}(t)$  (or  $(\tilde{u}_l(t), \tilde{s}_l(t))$ ) is 0, we obtained the following closed-form solution (see “[Methods](#)” section).

$$\begin{aligned}\text{CSP-Baseline : } \tilde{P}_n(t) &= \frac{a^n(t)}{n!} e^{-a(t)}, \quad n \in \mathbb{N}, \\ \text{CSP-Splicing : } \tilde{P}_{mn}(t) &= \frac{b^m(t)c^n(t)}{m!n!} e^{-b(t)-c(t)}, \quad (m, n) \in \mathbb{N}^2,\end{aligned}\tag{2}$$

where  $a(t)$ ,  $b(t)$  and  $c(t)$  are solutions to the corresponding deterministic equation, which means that  $\tilde{l}(t)$  obeys the Poisson distribution with mean  $a(t)$  in CSP-Baseline, and  $(\tilde{u}_l(t), \tilde{s}_l(t))$  obey independent Poisson distributions with mean  $b(t)$  and  $c(t)$  in CSP-Splicing. Here  $\alpha$ ,  $\beta$  are the transcription and splicing rates, and  $\gamma_s$ ,  $\gamma_t$  are the spliced and total mRNA degradation rates, respectively. CSP-Switching can be solved similarly analytically under the assumption that the switching rates are much smaller than the transcription and degradation rates (see “[Methods](#)” section).

We also specifically modeled technical noise of the measured number of new RNA (or new unspliced and spliced mRNA) molecules in scRNA-seq experiments. Such noises often lead to dropout of RNA measurements during the sequencing process and generally result in variations in sequencing depth across cells. To account for the noise, in Storm we modeled the dropout process of sequencing technology as cell-specific binomial distributions. Adopting the common practice in many preprocessing pipelines through a size factor to normalize the data [1, 2, 10, 13, 26], we assumed that the total numbers of mRNA molecules across all genes in different cells are close. Probabilistically, this assumption implies that

$$p_j \propto n_j,$$

where  $p_j$  is the probability of mRNA molecules being captured in cell  $j$  and  $n_j$  is the total number of mRNA molecules across all genes in cell  $j$  in scRNA-seq experiments.

Combining the stochastic models for biological and technical noise, we can obtain different formalisms of the distribution for the measured number of new/labeled mRNA molecules  $l(t)$  (or new unspliced and spliced mRNA molecules  $(u_l(t), s_l(t))$ ) in the scRNA-seq experiments (see “[Methods](#)” section) for each model. Specifically, in CSP-Baseline,  $l(t)$  obeys the cell-specific Poisson (CSP) distribution, that is,

$$P_{n,j}(t) = \frac{(p_j a(t))^n}{n!} e^{-p_j a(t)},\tag{3}$$

where  $P_{n,j}(t)$  is the PMF for the measured number of new mRNA molecules in cell  $j$ . In CSP-Splicing,  $(u_l(t), s_l(t))$  obeys the independent cell-specific Poisson (ICSP) distribution, that is,

$$P_{mn,j}(t) = \frac{(p_j b(t))^m}{m!} e^{-p_j b(t)} \frac{(p_j c(t))^n}{n!} e^{-p_j c(t)},\tag{4}$$

where  $P_{mn,j}(t)$  is the joint PMF for the measured number of new unspliced and spliced mRNA molecules in cell  $j$ . The derivation of CSP-Switching is similar (see “[Methods](#)” section). We call

the above distributions as *cell-specific* because different cells obey the distributions with different parameters. Finally for labeling efficiency, we did not model it directly but followed Dynamo using GRAND-SLAM [33] to correct the new RNA data in advance.

Note that Grün et al. also modeled the scRNA-seq data by integrating biological noise and technical noise [34]. Our work is different from them in the following aspects: (1) Our work models the transient dynamics of new mRNA and solves their distribution for the proposed stochastic models analytically. However, in [34], they instead modeled the total mRNA and derived that the biological noise follows a negative binomial distribution as the steady state of the transcriptional bursting model. (2) Our work accurately models the technical noise as a cell-specific binomial distribution, while they approximated the cell-specific binomial distribution with a Poisson distribution and modeled the capture probability as a random variable subject to the Gamma distribution, which finally leads to a negative binomial distribution (Poisson-Gamma mixture distribution) of the technical noise.

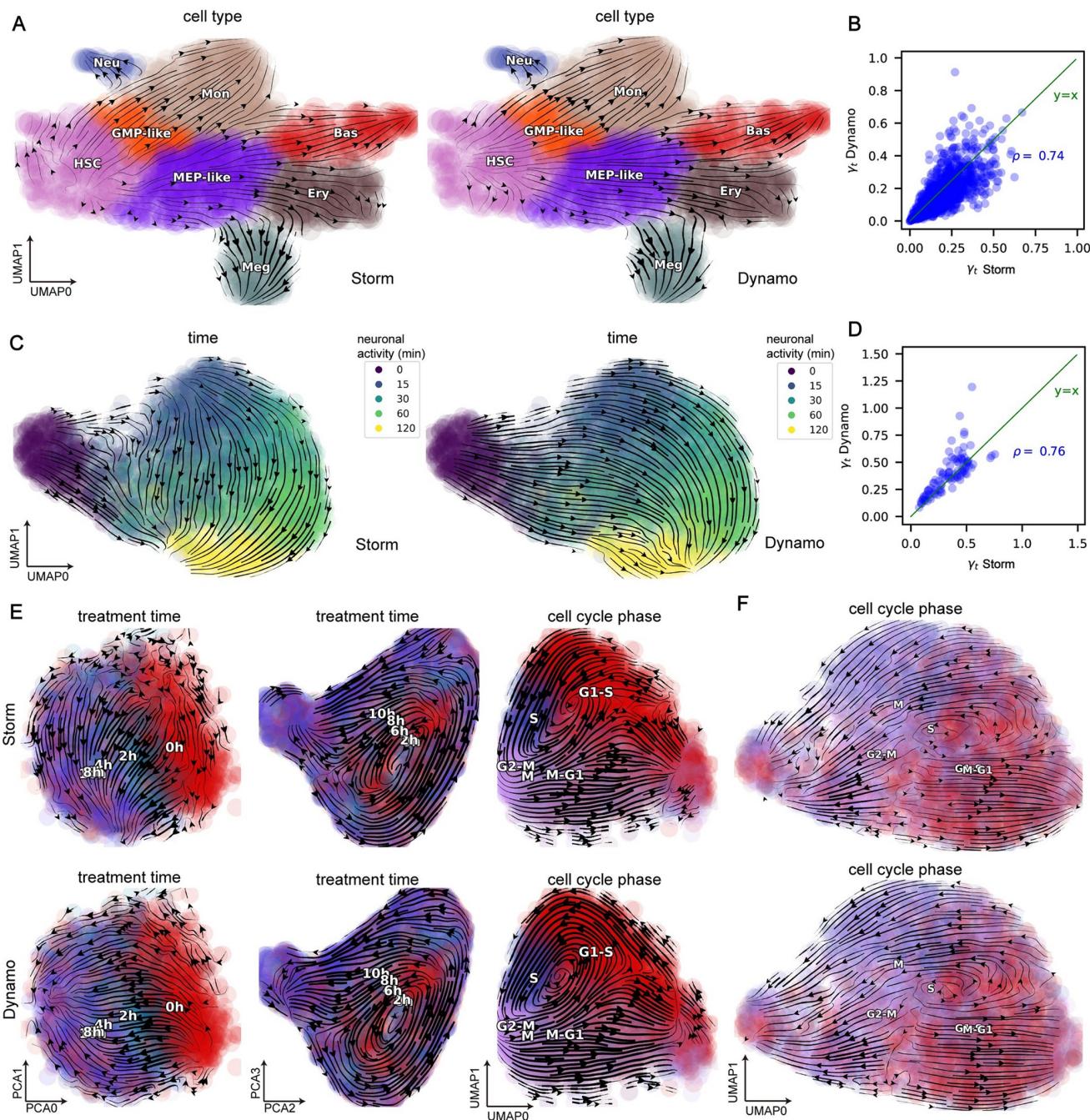
As one-shot labeling experiments are much more convenient than pulse experiments in practice, in the following, we will first demonstrate how Storm can be applied to the one-shot case. We will then extensively show Storm's power in analyzing the pulse datasets.

### Stochastic models combined with steady-state assumptions for one-shot data without splicing information

For one-shot data without splicing information, we designed the corresponding parameter inference method which invokes the steady-state assumption under the stochastic model, focusing specifically on CSP-Baseline (see “[Methods](#)” section). Similar steady-state methods of the stochastic model can also be designed for both CSP-Splicing and CSP-Switching as well, although they are not the focus of this paper.

We validated our method in several one-shot datasets ([Fig 2](#) and [S1 Fig](#)). We first analyzed a primary human HSPCs datasets from scNT-seq [18]. Both Dynamo and Storm reveal a smooth transition of HSCs into MEP-like and GMP-like cells, which further ramify into Meg/Ery/Bas lineages and Mon/Neu lineages, respectively, which is consistent with the established knowledge of hematopoiesis ([Fig 2A](#)). Next, we analyzed the neuronal activity dataset from the scNT-seq study [18] to investigate cellular polarization dynamics after KCl treatment. Dynamo and Storm both revealed a coherent transition that nicely follows the temporal progression from time point 0 to 15, 30, 60 and finally 120 minutes ([Fig 2C](#)). We also analyzed a dataset from the sci-fate study [19] in which cell cycle progression and glucocorticoid receptor (GR) activation were explored. Similar to Dynamo, the RNA velocity flow from our method also revealed a sequential transition of cells following the DEX (dexamethasone) treatment times in the first two principal components (PCs) ([Fig 2E Left](#)). In the second two PCs, we observed an orthogonal circular progression of the cell cycle ([Fig 2E Middle](#)). From the first two UMAP dimensions projected further from the four PCs, we observed a combined dynamics of GR responses and cell cycle progression ([Fig 2E Right](#)). We analyzed a dataset from PerturbSci-Kinetics [22], Dynamo and Storm observed similar results ([Fig 2F](#)). Additionally we analyzed mouse fibroblast cells dataset from scSLAM-seq [17]. We observed that Dynamo and Storm inferred similar velocities, and they both further discriminated infected from non-infected cells ([S1A Fig](#)).

Finally, we quantitatively compared the degradation rates  $\gamma_t$  inferred by the two methods ([Fig 2B and 2C](#) and [S1B and S1D Fig](#)). The inferred results of the two methods are highly correlated, with values of 0.74, 0.76, 0.80, 0.91, and 0.87 respectively on these five datasets. The absolute differences are also low, and the inferred results are mostly distributed around the green line  $y = x$ . We think that the possible reason for this is that the steady-state assumption



**Fig 2. Stochastic model combined with steady-state assumptions for one-shot data without splicing information.** Storm in this figure refers to the inference strategy of CSP-Baseline model combined with the steady state assumption. **A.** Streamline projected in the UMAP space plots of primary human HSPCs datasets from scNT-seq [26]. **B.** Degradation rates  $\gamma_t$  estimated with steady-based method in Storm compared to that of the Dynamo method in the primary human HSPCs datasets. **C.** Streamline projected in the UMAP space plots of neuronal activity under KCl polarization datasets from scNT-seq [18]. **D.** Same as B., but for the neuronal activity datasets. **E.** Streamline plots of the sci-fate dataset [19] reveal two orthogonal processes of GR response and cell-cycle progression. From left to right: streamline plot on the first two PCs, the second two PCs, and the first two UMAP components that are reduced from the four PCs, respectively. The first row is the result of Storm and the second row is the result of Dynamo. **F.** Streamline projected in the UMAP space plots of the dataset from PerturbSci-Kinetics [22].

<https://doi.org/10.1371/journal.pcbi.1012606.g002>

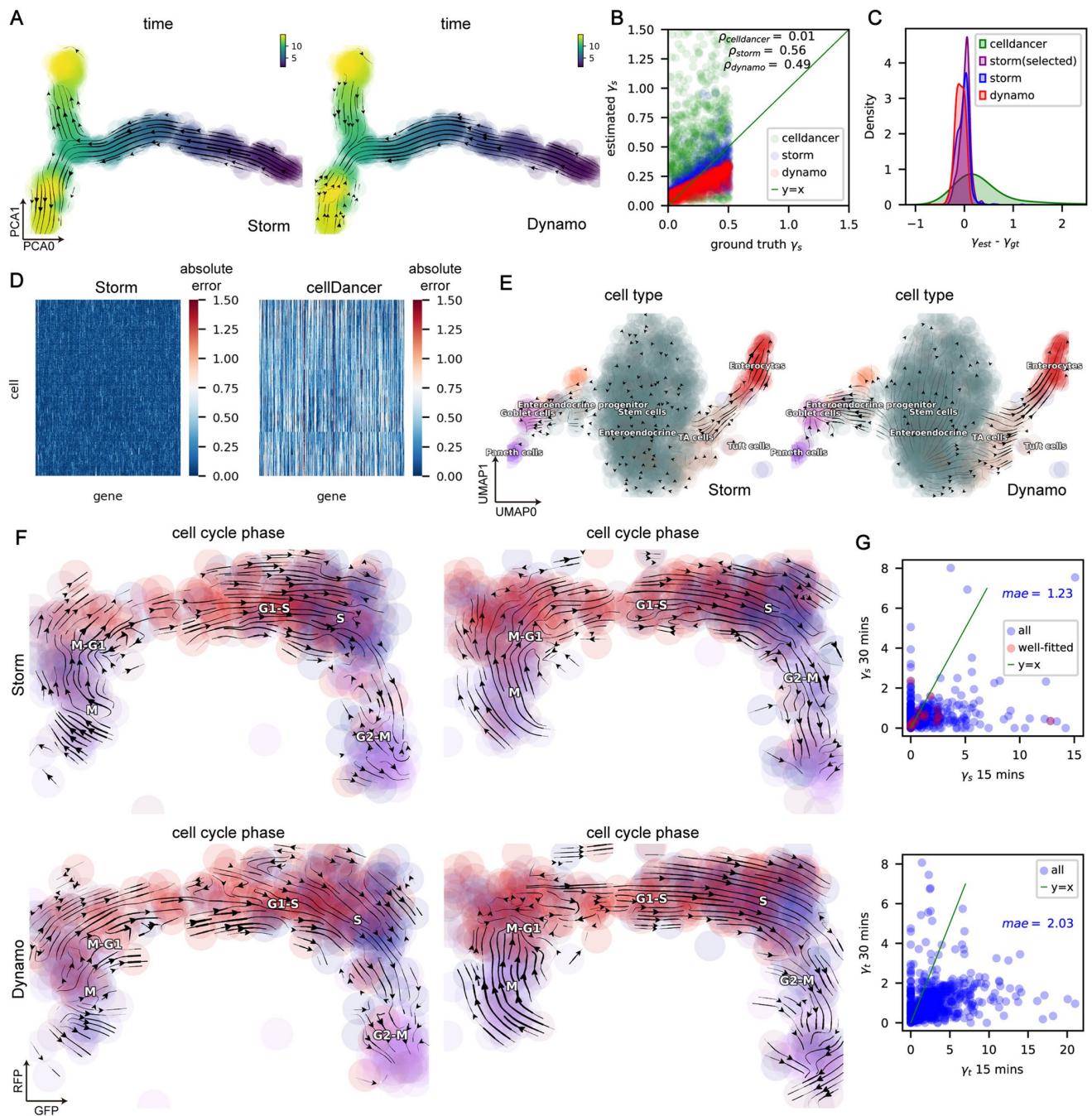
plays a decisive role. Such methods may fail when the steady-state assumption is violated, so it is important to design methods for one-shot experiments that do not rely on the steady state assumption.

### Storm analyzes one-shot data with both splicing and labeling information without steady-state assumption

For one-shot data containing both splicing and labeling information, we designed a two-stage parameter inference method that does not depend on the steady-state assumption by first modeling unspliced and spliced RNAs with the dynamic model in scVelo [2] and then unspliced labeled and spliced labeled RNA counts with CSP-Splicing (see “[Methods](#)” section). Also distinguishing from scVelo [2] in which velocity genes are picked before running the algorithm, we compute goodness of fit  $R^2$  to pick well-fitting genes in scVelo [2] for downstream analysis (see “[Methods](#)” section).

We validate our method on both simulation and real single cell datasets ([Fig 3](#) and [S2 Fig](#)). We first constructed a bifurcated one-shot simulation dataset by following the methods for constructing bifurcated data in SymSim [30] and VeloSim [31] (see “[Methods](#)” section). The correct direction of the streamlines started from the right side of the cells on the PCA embedding and then bifurcated into two branches in the middle. We compare the performance of Storm, Dynamo, and a deep learning-based method cellDancer [16] on this simulated dataset, and the results show that only Storm and cellDancer got the correct streamlines ([Fig 3A](#) and [S2A Fig](#)). In addition we compared the degradation rate values estimated by the different methods with the true values ([Fig 3B](#) and [3C](#)). The results show that cellDancer’s estimated values are not in the same order of magnitude as the true values and have a poor correlation ([Fig 3B](#) and [3C](#)). This is a difficulty inherent in methods that are missing physical time information. Dynamo and Storm’s estimates are in the same order of magnitude as the true values, but Storm has a lower absolute error and a higher correlation with the true values compared to Dynamo ([Fig 3B](#) and [3C](#)). It is also worth noting that the absolute error of the selected genes in Storm is further reduced ([Fig 3C](#)), which indicates that the selection strategy is effective. We also compared gene-cell-wise transcription rates. The results show lower absolute errors in estimated transcription rates for Storm and Dynamo ([Fig 3D](#) Left and [S2C Fig](#)), but much higher for cellDancer ([Fig 3D](#) Right). We analyzed the murine intestinal organoid system dataset from scEU-seq [21]. Both Storm, Dynamo and cellDancer observed a bifurcation ([Fig 3E](#) and [S2B Fig](#)) from intestinal stem cells into the secretory lineage (left) and the enterocyte lineage (right).

To demonstrate the precision and robustness of the Storm method in estimating the one-shot dataset, we benchmarked the estimated kinetic parameters of different subsets of the cell cycle pulse-labeling dataset [21], each with a different duration of labeling. On the 15-minute labeling sub-dataset, Storm recovers a transition that matches well with the cell-cycle progression, while the transition recovered by Dynamo is problematic near the M phase ([Fig 3F](#) Left). On the 30-minute labeling sub-dataset, both methods recover the cell cycle progression correctly, but the streamlines of our method are considerably smoother compared to those of Dynamo ([Fig 3F](#) Right). In addition, we compared the consistency of degradation rates ( $\gamma_s$  in Storm and  $\gamma_t$  in Dynamo) inferred by the two methods between two sub-datasets with different labeling durations ([Fig 3G](#)). The results show that they are similar in terms of correlation, but our method is much smaller in terms of mean absolute error. Notably, although Storm shows higher consistency than Dynamo, it is still not satisfactory, perhaps due to the experimental noises from different labeling durations. Therefore, it is crucial to integrate data of different durations of labeling when a kinetic experiment is available. Furthermore, it is equally



**Fig 3.** Storm analyzes one-shot data with both splicing and labeling information without steady-state assumption. Storm in this figure refers to the inference strategy of CSP-Splicing model combined with scVelo. **A.** Streamline projected in the PCA space plots of one-shot bifurcation simulation data. Left: Storm; Right: Dynamo. **B.** Comparison of the estimated degradation rate with the true degradation rate in one-shot bifurcation simulation data. cellDancer uses the average of cell-wise degradation rates. **C.** Distribution plot of the difference between the estimated degradation rate and the true value, including Storm, Storm (selected) and Dynamo. **D.** Heat map of absolute error between estimated and true gene-cell-wise transcription rates  $\alpha$  of one-shot bifurcation simulation data. Left: Storm; Right: cellDancer. **E.** Streamline plot in the UMAP space of the murine intestinal organoid system dataset from scEU-seq [21]. **F.** Streamline projected in the RFP\_GFP space plots of cell cycle dataset from scEU-seq [21]. On the left is the result of taking only the data labelled with 15 minutes, and on the right is the data labelled with 30 minutes. **G.** Comparison of degradation rates ( $\gamma_s$  in Storm and  $\gamma_t$  in Dynamo) in cell cycle datasets with labeling duration of 15 and 30 minutes.

<https://doi.org/10.1371/journal.pcbi.1012606.g003>

important to design methods that do not rely on the steady-state assumption and integrate data of different durations for parameter inference.

### Statistical analysis of cell cycle dataset based on Storm's stochastic model

Next, we first performed a goodness-of-fit test of the stochastic model proposed in Storm to a cell cycle dataset from scEU-seq [21] with multiple labeling time points to validate our proposals.

When the fixed labeling duration is  $t_{\text{fixed}}$ ,  $a(t_{\text{fixed}})$ ,  $b(t_{\text{fixed}})$  and  $c(t_{\text{fixed}})$  are all fixed constants. We can test whether the number of new mRNA molecules in tscRNA-seq within a fixed labeling duration matches the distribution obtained based on the stochastic models (Eqs (3), (4) and (16)), respectively. A common method of testing whether a dataset obeys a given distribution is the chi-square ( $\chi^2$ ) goodness-of-fit test [35]. However, the usual  $\chi^2$  test is not directly applicable because in our case different cells obey different distributions with different parameters. By inspecting the mathematical analysis procedure of the  $\chi^2$  test [36], we constructed a new asymptotic  $\chi^2$  statistics and proposed a modified  $\chi^2$  test for our cell-specific distributions (see “Methods” section).

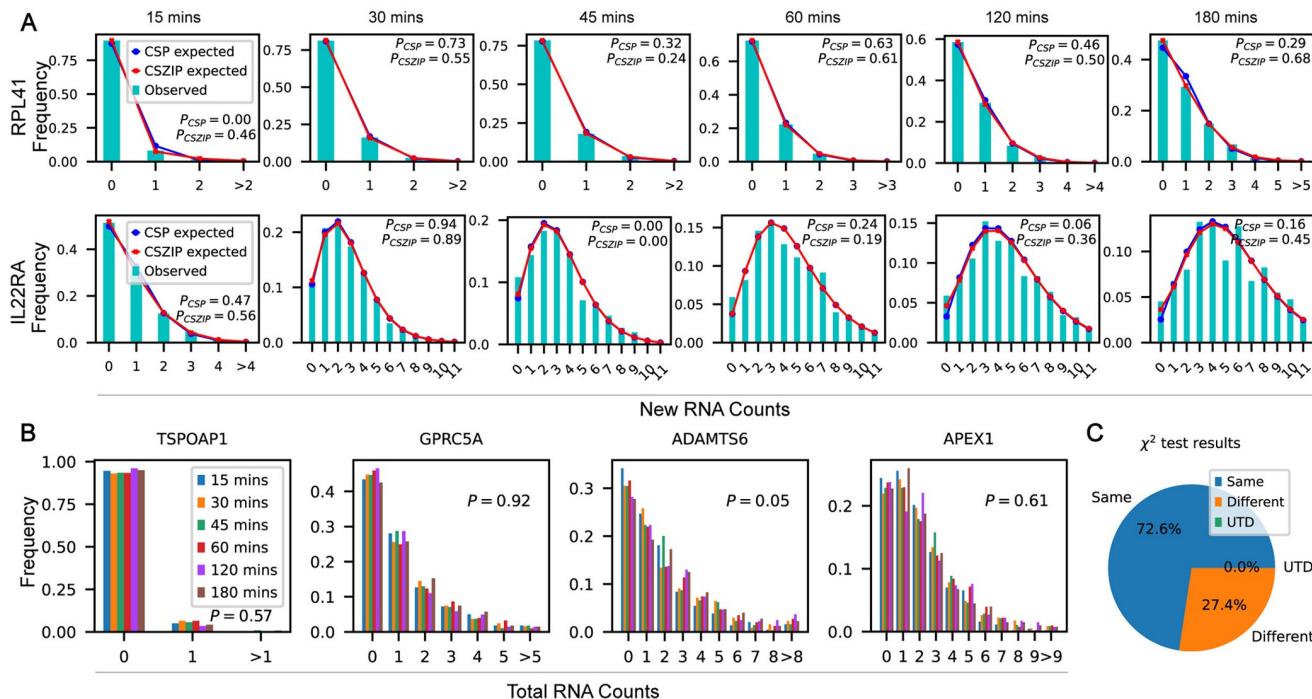
We used the proposed cell-specific  $\chi^2$  test in the cell cycle dataset from the scEU-seq study [21], in which cells were labeled for 15, 30, 45, 60, 120 or 180 minutes. Because the labeled unspliced mRNA counts  $u_i(t)$  were too small to be grouped/binned to create a distribution, hypothesis tests were performed only for CSP and CSZIP distributions and not for ICSP distribution. The results are shown in Table 1. We found that some genes were not well determined (especially for cases with a short duration of labeling) in the sense that these genes had too few new mRNA molecules in the tscRNA-seq experiments, which resulted in very few groupings and perfect fittings. With so few mRNA counts for these genes, we were unable to determine whether they obeyed our proposed distribution or not. Moreover, our results revealed that the CSZIP distribution exhibited a better fit with the data than the CSP distribution when focusing on a fixed time point alone, suggesting that the data are indeed zero-inflated.

We next showed the high goodness-of-fit of the CSP and CSZIP distributions on two characteristic genes, namely *RPL41* and *IL22RA1* with an overall low and high gene expression respectively (Fig 4A). Qualitatively, we found that the expected counts of both the CSP and CSZIP distributions matched well with the observed counts for the gene *RPL41*. Quantitatively, the results of the cell-specific chi-square test also showed that the CSP or CSZIP distribution was well satisfies in most labeling durations (Fig 4A first row). Similar results were observed for the gene *IL22RA1* with significantly higher expression (Fig 4A second row). Therefore, we demonstrated CSP and CSZIP distributions accurately describe these two genes and is thus suitable for modeling the tscRNA-seq datasets.

**Table 1. The proposed sample-specific hypothesis test results on whether the number of new mRNA molecules in the Cell Cycle dataset obeys the CSP and CSZIP distributions.** UTD means that it is unable to determine because there are too few groupings resulting in zero degrees of freedom, when it is always a perfect fit. The significance level is 0.05.

Labeling duration		15mins	30mins	45mins	60mins	120mins	180mins
CSP	Accept	0.116	0.067	0.049	0.062	0.064	0.065
	Reject	0.278	0.568	0.655	0.652	0.695	0.725
	UTD	0.606	0.365	0.296	0.286	0.241	0.210
CSZIP	Accept	<b>0.351</b>	<b>0.467</b>	<b>0.472</b>	<b>0.476</b>	<b>0.459</b>	<b>0.459</b>
	Reject	0.055	0.189	0.266	0.274	0.327	0.344
	UTD	0.594	0.344	0.262	0.250	0.214	0.197

<https://doi.org/10.1371/journal.pcbi.1012606.t001>



**Fig 4. Statistical analysis of cell cycle dataset.** **A.** Observed counts, expected counts of CSP distribution, and expected counts of CSZIP distribution of new mRNA molecules of the two example genes *RPL41* and *IL22RA1*. The first row: Fitting results of the *RPL41* gene with a small number of mRNA molecules; The second row: Fitting results of the *IL22RA1* gene with a higher number of new mRNA molecules (truncated to 11 for better visualization).  $P_{CSP}$  and  $P_{CSZIP}$  refer to the p-values of the cell-specific chi-square tests with the corresponding distributions. **B.** Comparison of the total mRNA counts with different labeling durations of the four example genes *TSPOAP1*, *GPRC5A*, *ADAMTS6* and *APEX1*.  $P$  refers to the p-value of the Chi-square contingency table independence test. **C.** Results of chi-square independence test for total RNA counts (significance level 0.05). “Same” here means accepting the null hypothesis of the chi-square independence test that total RNA counts with different time durations obey the same distribution. “Different” means the opposite.

<https://doi.org/10.1371/journal.pcbi.1012606.g004>

Finally, we found that, for most genes, the number of total mRNA molecules shares the same distribution across different labeling durations. In Fig 4B, we showed the number of total mRNA molecules of four example genes *TSPOAP1*, *GPRC5A*, *ADAMTS6* and *APEX1* is nearly identical across different labeling durations. Quantitatively, we performed a global chi-square independence test on the number of total mRNAs (as distinct from the new mRNAs) with different durations of labeling in all genes and found that, interestingly, there are 72.3% of the genes passed the test at a significance level of 0.05 (Fig 4C). This indicates that a considerable proportion of the number of genes’ total mRNA molecules obeyed the same distribution, consistent with what we observed for the four example genes.

### Storm accurately infers kinetic parameters that leads to rich insights of cell cycle via enrichment analysis

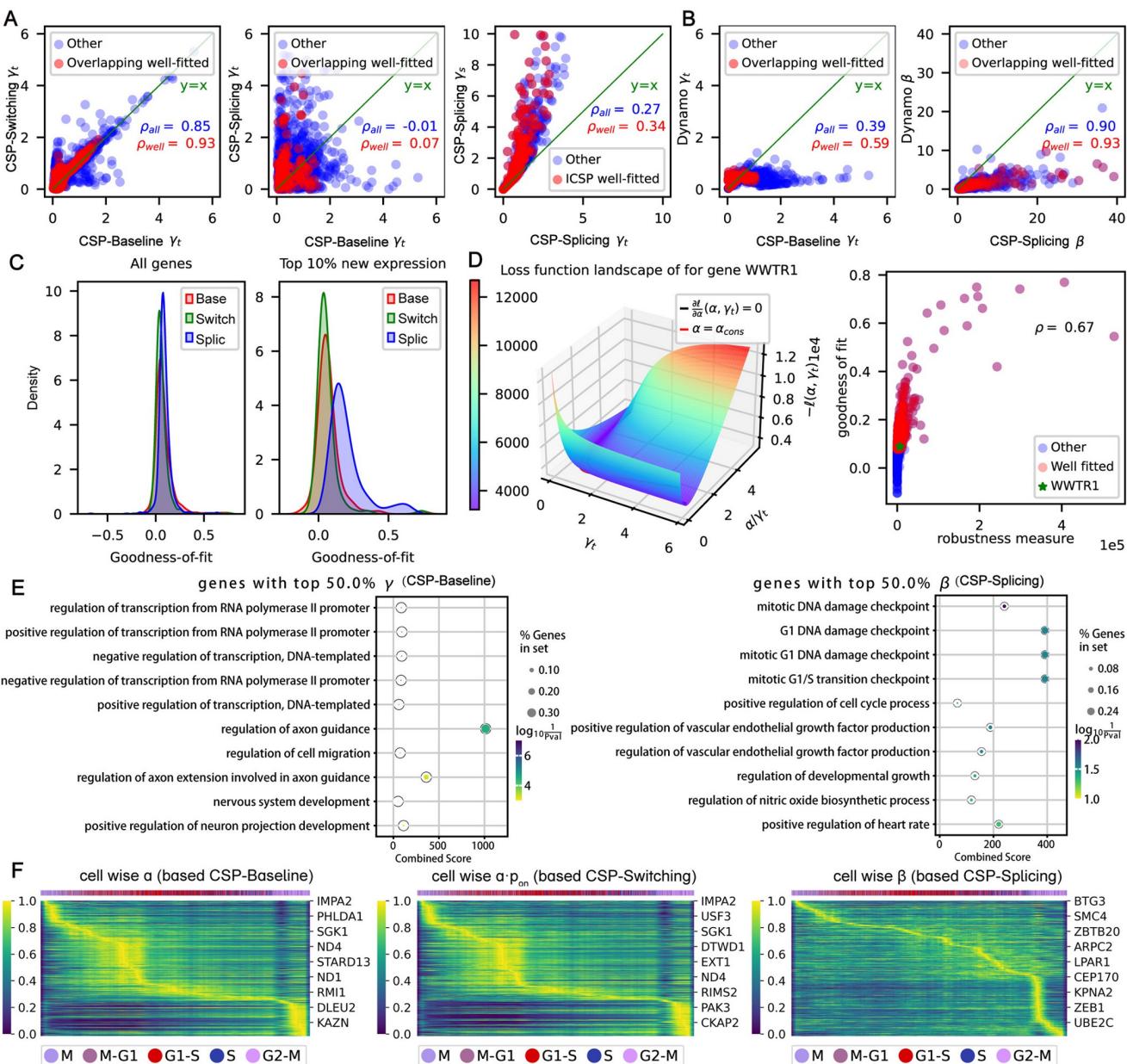
In the kinetic experiments, we relied on three stochastic models without the steady-state assumption to infer different set of kinetic parameters using maximum likelihood estimation (see “Methods” section), namely  $\alpha$  and  $\gamma_t$  for CSP-Baseline,  $\alpha$ ,  $\beta$  and  $\gamma_s$  for CSP-Splicing, and  $\alpha$ ,  $\gamma_t$  and  $p_{off}$  for CSP-Switching. In addition, we defined the goodness-of-fit of each of the three models (see “Methods” section). According to the goodness-of-fit index, we selected genes that were more consistent with the model assumptions for downstream tasks, such as the enrichment analysis and RNA velocity analysis, etc.

Compared with Dynamo [26], the state-of-the-art method for processing tscRNA-seq datasets, our advantages are mainly in the following aspects: (1) Our method does not require steady-state assumptions on the kinetics experiments while Dynamo heavily relies on the steady-state assumptions; (2) Our stochastic model-based approach is more consistent with real biological process, while Dynamo only utilizes the deterministic model of mean value; (3) Our model takes into account all cells in the inference, while the approach based on steady-state assumptions in Dynamo only considers a small number of cells with high expression. In addition, we revealed the difference between the total mRNA degradation rate  $\gamma_t$  and spliced mRNA degradation rate  $\gamma_s$  based on their different physical roles, distinguished them in different models, and finally gave the relationship between these two (see “[Methods](#)” section). We noted that in Dynamo, to infer  $\beta$ ,  $\gamma_t$  was first inferred when the splicing was ignored, then  $\tilde{\gamma} := \gamma_s/\beta$  was inferred using the method based on the steady-state assumption in scVelo [2], and finally  $\gamma_t/\tilde{\gamma}$  was taken as the inference of  $\beta$  upon assuming  $\gamma_t = \gamma_s$ . However,  $\gamma_t/\tilde{\gamma} = \beta\gamma_t/\gamma_s$ , while  $\gamma_t$  and  $\gamma_s$  are generally not equal. This point was overlooked in Dynamo, which causes an inaccurate estimate of  $\beta$ . In fact, under the steady-state assumption,  $\beta$  can be directly estimated by using only  $u_i(t)$  through the formula  $u_i(t) = (1 - e^{-\beta t})\alpha/\beta$ , similar to the two-step method used in Dynamo to estimate  $\gamma_t$  through  $l(t) = (1 - e^{-\gamma_t t})\alpha/\gamma_t$  since they have similar form. However, we don’t use this method in Storm.

With the above inference methods and insights, we studied a cell cycle dataset from the scEU-seq study [21]. We compared the parameter inference results of the three models ([Fig 5A](#)). When splicing was not considered, the inference results based on CSP-Baseline and CSP-Switching were close, with high correlation coefficients, especially in genes with higher goodness of fit ([Fig 5A Left](#)). However, whether or not splicing is considered significantly impacts the inference results. The inference results based on CSP-Baseline and CSP-Splicing were quite different, with low correlation coefficients, even in genes with higher goodness of fit ([Fig 5A Middle](#)). We speculate that this is due to the assumptions of the two models are incompatible: in CSP-Baseline,  $\gamma_t$  is assumed to be a constant; while in CSP-Splicing,  $\gamma_s$  is assumed to be a constant. But these two assumptions can not be held simultaneously for their different roles in the physical modeling and our analysis results (see “[Methods](#)” section). We also compared  $\gamma_t$  and  $\gamma_s$  computed by the CSP-Splicing, and the results showed that  $\gamma_s$  was always greater than  $\gamma_t$ , and the linear correlation between the two was not high ([Fig 5A Right](#)). In summary, we showed that kinetic parameters inferred from CSP-Baseline and CSP-Switching but not CSP-Baseline and CSP-Splicing, are consistent.

The inferred total mRNA degradation rates  $\gamma_t$  from Storm and Dynamo are close in well-fitted genes, while CSP-Splicing’s inferred splicing rates  $\beta$  are always larger than those from Dynamo. We compared the inferred results of  $\gamma_t$  in CSP-Baseline with those in Dynamo ([Fig 5B Left](#)). Although they were not consistent for some genes, they are quite consistent for the genes with better fitting. We also compared the inference of  $\beta$  in CSP-Splicing with those in Dynamo ([Fig 5B Right](#)). The result shows that the inferred  $\beta$  by our approach was usually larger than those in Dynamo, even for the genes with a better fitting. A possible explanation is that the inference of Dynamo ignored the difference between  $\gamma_t$  and  $\gamma_s$ , which made the inferred  $\beta$  smaller. We also compared the goodness-of-fit of the three stochastic models. Overall, they are relatively close ([Fig 5C Left](#)). However, when we focused on genes with higher new mRNA levels (top 10%), CSP-Splicing had a better fit ([Fig 5C Right](#)). We speculate that this is because genes with higher expression are suitable to be fitted with more complex models.

When the parameter  $\gamma_t$  is small, parameter inference may not be robust enough. However, we found that the genes selected by the goodness-of-fit have robust results. We analyzed the robustness of the parameter inference in the simplest CSP-Baseline model (see “[Methods](#)” section). We plotted the landscape of a typical negative log-likelihood loss function based on



**Fig 5. Parameter inference and enrichment analysis for the cell cycle dataset.** The inference strategy involved in this figure is for kinetics/pulse data. **A.** Comparison of parameter inference results of our three stochastic models. From left to right are the comparison of  $\gamma_t$  of CSP-Baseline and CSP-Switching, the comparison of  $\gamma_t$  of CSP-Baseline and CSP-Slicing, the comparison of  $\gamma_t$  and  $\gamma_s$  in CSP-Slicing. The overlapping well-fitted genes were set as the overlap set of genes in the top 40% of the goodness-of-fit for both methods. **B.** Comparison of inferred parameters between our stochastic models and Dynamo's method. **Left:** the comparison of  $\gamma_t$  between CSP-Baseline and Dynamo. **Right:** the comparison of  $\beta$  between CSP-Slicing and Dynamo. **C.** Comparison of the goodness-of-fit of the three stochastic models. **Left:** all highly variable genes. **Right:** genes in the top 10% of average new mRNA expression in highly variable genes. Here Base refers to the CSP-Baseline model, Split to the CSP-Slicing model and Switch to the CSP-Switching model. **D.** Robust analysis. **Left:** Landscape of CSP-Baseline-based loss functions for the a typical gene WWTR1. **Right:** Scatter plot of robustness measure and goodness of fit for parameter inference. **E.** Enrichment analysis results of genes with high gene-wise  $\gamma_t$ ,  $\beta$  (top 50%) in well fitted genes (top 40% of goodness of fit). **F.** Heat map of cell-wise parameters for well-fitted genes. From left to right, cell-wise  $\alpha$  based on the CSP-Baseline, cell-wise  $\alpha \cdot p_{on}$  based on the CSP-Switching and cell-wise  $\beta$  based on the CSP-Slicing, respectively. Across all three heatmaps, the X-axis is the relative cell cycle position while the order of genes in the Y-axis is arranged such that the peak time of each gene increases from the top left to bottom right.

<https://doi.org/10.1371/journal.pcbi.1012606.g005>

CSP-Baseline for gene *WWTR1* (Fig 5D Left), with the black line corresponding to  $\partial\ell(\alpha, \gamma_t)/\partial\alpha = 0$  (Eq (83) in the “Methods” section) and blue line corresponding to  $\alpha = \alpha_{\text{cons}}$  when  $1 - e^{-\gamma_t t} \sim \gamma_t t$  holds (Eq (84) in the “Methods” section). The landscape of the loss function shows a fairly flat area around  $\partial\ell/\partial\alpha = 0$ , and the two lines almost coincide when  $\gamma_t$  is small, which is consistent with our previous argument. We design a quantitative index to measure the robustness of parameter inference (see “Methods” section) and analyzed the relationship between the robustness measure and the goodness-of-fit  $\bar{R}_D^2$  (Fig 5D Right). We found that parameter robustness was positively correlated with the goodness of fit and the correlation coefficient was as high as 0.69. Though the reason for this high correlation is not clearly understood in theory, we can utilize this fact to select the genes with high goodness of fit for downstream analysis, which also ensures the results are relatively robust.

We selected the well-fitted genes (top 40%  $\bar{R}_D^2$ ) and performed enrichment analysis on this fraction according to the magnitude of gene-wise parameters  $\gamma_t, \beta, \alpha$  and  $p_{\text{off}}$  (Fig 5E and S3 Fig). The results of the enrichment analysis showed that these genes were highly correlated with the cell cycle progression.

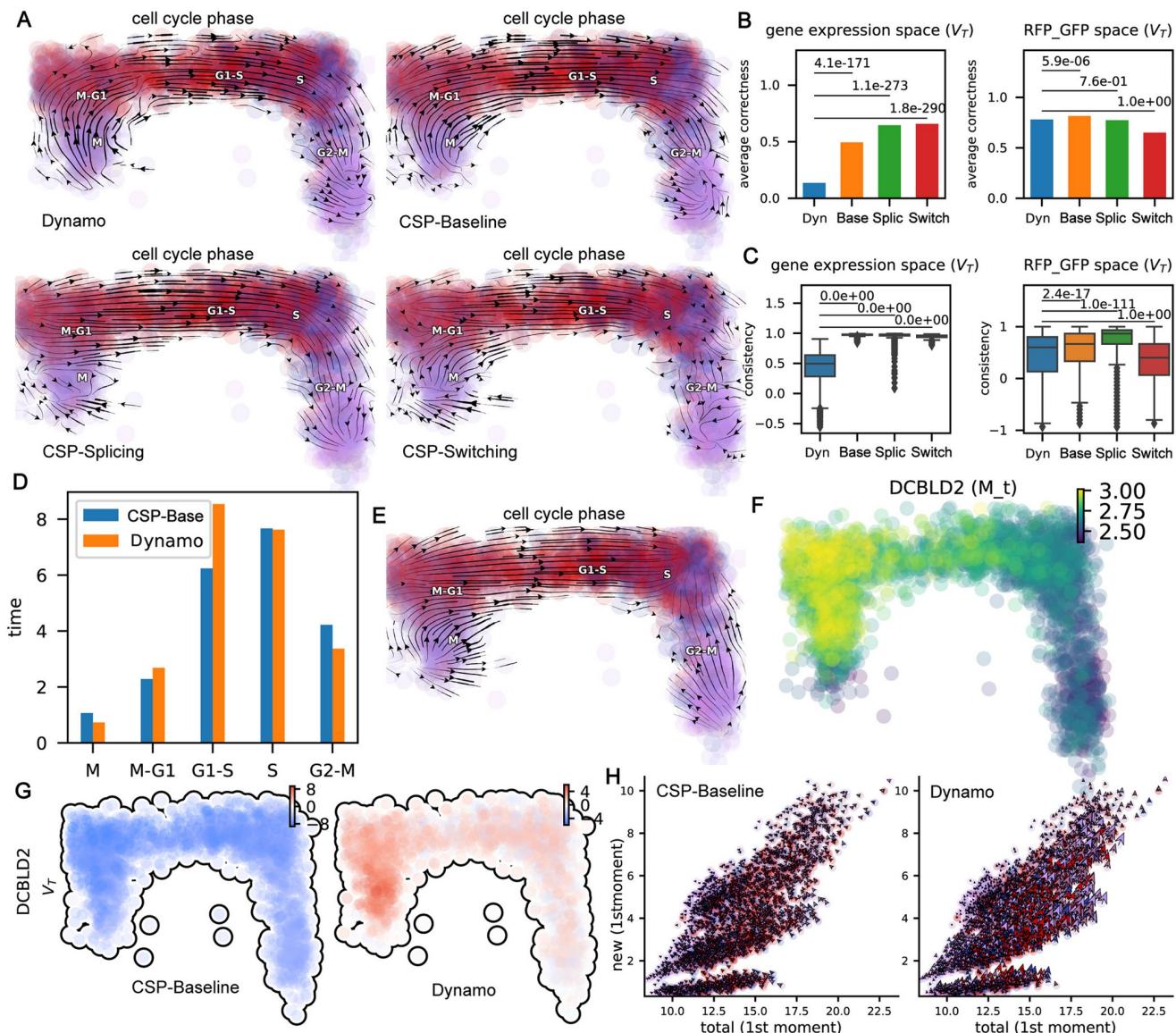
The assumption of constant coefficients is often violated because of the time-dependent kinetics and multiple lineages [11]. Many works relaxed the constant coefficient assumption and inferred cell-specific parameters to overcome this issue [10, 13, 16, 26]. In our proposal, we take a post-processing step to get the cell-specific parameters after inferring all parameters through previous procedures. We relaxed the constant coefficient assumption and proposed a method to infer cell-specific parameters except the constant degradation rate  $\gamma_t$  or  $\gamma_s$ , i.e., we inferred cell-specific  $\alpha$  in CSP-Baseline, cell-specific  $\alpha \times p_{\text{on}}$  in CSP-Switching, and the cell-specific  $\alpha$  and  $\beta$  in CSP-Splicing (see “Methods” section). This partial constant coefficient assumption had support from the study in [21], which showed that the degradation rate of most genes was independent of time. Finally, We plotted heat maps of the cell-wise  $\alpha$  (based on CSP-Baseline),  $\alpha \times p_{\text{on}}$  (based on CSP-Switching) and  $\beta$  (based on CSP-Splicing) for the well-fitted genes (Fig 5F). The results show that cells in the same cell cycle phase usually have closer kinetic parameters.

### Storm improves the robustness and accuracy of time-resolved RNA velocity analysis

Our three stochastic models described the evolution of the PMF (or joint PMF) of the number of new mRNA (or new unspliced and spliced mRNA) molecules over time for different settings. To estimate RNA velocity of single cells, only the evolution of the mean value over time will be considered, which requires us to reduce the stochastic models to the corresponding deterministic models (see “Methods” section).

Based on the deterministic model derived for the mean corresponding to the three stochastic models, we inferred the relevant parameters for computing different types of RNA velocity for different models. In Models 1 and 3, we computed the total RNA velocity  $d\langle\tilde{r}(t)\rangle/dt$  because the splicing process was ignored. In CSP-Splicing, we calculated both total RNA velocity  $d\langle\tilde{r}(t)\rangle/dt$  and spliced RNA velocity  $d\langle\tilde{s}(t)\rangle/dt$  (see “Methods” section). Note that because the new RNA velocity mostly reflects the metabolic labeling process of RNA and does not reveal RNA biogenesis, it is thus not used. In addition, a derived relationship between  $\gamma_t$  and  $\gamma_s$  suggests that the total RNA velocity can be computed based on either  $d\langle\tilde{r}(t)\rangle/dt = \alpha - \gamma_s\langle\tilde{s}(t)\rangle$  or  $d\langle\tilde{r}(t)\rangle/dt = \alpha - \gamma_t\langle\tilde{r}(t)\rangle$ . In practice, we used the former approach by default.

We compared the streamlines of the total RNA velocity of our three models with that of Dynamo on the cell cycle scEU-seq dataset (Fig 6A). Almost all streamlines from our models correctly reflect the cell cycle progression, except that part of them from CSP-Splicing had a



**Fig 6. RNA velocity analysis of the cell cycle dataset. The inference strategy involved in this figure is for kinetics/pulse data.** **A.** Comparison of total RNA velocity streamline visualizations between three stochastic methods and Dynamo in cell cycle dataset. **B.** Comparison of average correctness of total velocity in gene expression space and RFP\_GFP space. The p-values are given by the one-sided Wilcoxon test. Here Base refers to the CSP-Baseline model, Splic to the CSP-Splicing model and Switch to the CSP-Switching model. **C.** Similar to **B**, comparison of velocity consistency. **D.** The duration time (unit: hour) of each cell cycle phase of the human RPE1-FUCCI system based on Storm's CSP-Baseline and Dynamo. **E.** Total RNA velocity streamlines calculated using Storm's CSP-Baseline with gene-wise parameters (instead of using gene-cell-wise parameters except for the degradation rate). **F.** The smoothed expression of *DCBLD2* in different cells. **G.** Comparison of total RNA velocity in *DCBLD2* between CSP-Baseline and Dynamo. **H.** Phase portraits of new-total RNA planes of *DCBLD2* of CSP-Baseline and Dynamo. Quivers correspond to the total (x-component) or new (y-component) RNA velocity calculated by the different methods.

<https://doi.org/10.1371/journal.pcbi.1012606.g006>

minor flaw in the M phase and CSP-Switching in the S phase. In addition, we found both CSP-Splicing and Dynamo's spliced RNA velocity (S4A Fig) did not get entirely correct streamline results. The streamlines of CSP-Splicing were problematic in the M-G1 phase, while the streamlines of Dynamo were problematic in the S phase. We speculate that this is probably due to the fact that new unspliced mRNAs have rather low expression levels,

frustrated with many dropouts and very sparse data, resulting in unreliable inferences of the parameter  $\beta$  and inaccurate RNA velocities.

We also quantitatively benchmarked the average correctness and consistency of the velocities in different methods in the original gene expression space and low-dimensional space (here the RFP\_GFP space is used which corresponds to the Geminin-GFP and Cdt1-RFP-corrected signals of RPE1-FUCCI cells) ([Fig 6B and 6C](#) and [S4B and S4C Fig](#)). The definition of correctness and consistency of velocity is given in the “[Methods](#)” section. In the gene expression space, the average correctness and consistency of the total RNA velocity of CSP-Baseline, CSP-Splicing, and CSP-Switching are significantly better than that of Dynamo ([Fig 6B and 6C](#) Left), while the spliced RNA velocity of CSP-Splicing has slightly lower consistency than that of Dynamo ([S4B and S4C Fig](#) Left). In the RFP\_GFP space, the average correctness of total RNA velocity of all methods are significantly higher compared to that in the gene expression space, and simpler methods tend to improve more. The average correctness of CSP-Baseline is highest at this time ([Fig 6B](#) Right). However, the average correctness of the CSP-Splicing’s spliced RNA velocity still perform slightly worse than Dynamo’s ([S4B Fig](#) Right). In contrast, the total RNA velocity consistency of CSP-Baseline and CSP-Splicing is significantly better than that of Dynamo ([Fig 6C](#)) and the spliced RNA velocity consistency of CSP-Splicing is also significantly better than that of Dynamo ([S4C Fig](#) Right). Overall, the CSP-Baseline-based total RNA velocity has the highest average correctness and consistency, and significantly outperforms Dynamo, while the CSP-Splicing-based spliced RNA velocity was close to Dynamo quantitatively.

To demonstrate the significance of inferring time-resolved velocities with physical units, we calculated the duration time of each cell cycle phase of the human RPE1-FUCCI system based on the total RNA velocities (see “[Methods](#)” section, [Fig 6D](#)). Indeed, the human RPE1-FUCCI system has a cell-cycle time of about 21 hours (about 6 hours for G1-S phase, 8 hours for S phase, 4 hours for G2-M phase, 1 hour for M phase and 2 hours for M-G1 phase) [37].

To demonstrate the value of using gene-cell-wise parameters (except degradation rates), we visualized the streamlines of total RNA velocity based on gene-cell-wise parameters and those based only on gene-wise parameters ([Fig 6E](#) and [S4D Fig](#)). We observed that the streamlines of CSP-Baseline and CSP-Switching in the S to G2-M phase are incorrectly reversed ([Fig 6E](#) and [S4D Fig](#) Right), and the streamlines of CSP-Splicing are also less smooth and accurate than those when gene-cell-wise parameters are used ([S4D Fig](#) Left).

We now illustrate the advantages of our method in the estimation of kinetic parameters and the calculation of RNA velocity with two example genes: *DCBLD2* and *HIPK2*. In gene *DCBLD2*, the cells at M and M-G1 have the highest overall expression and the correct RNA velocity should be negative ([Fig 6F](#)). However, Dynamo returned the positive velocity, which is problematic ([Fig 6G](#) Right). In contrast, CSP-Baseline, CSP-Switching and CSP-Splicing all returned negative velocities ([Fig 6G](#) Left and [S4E Fig](#)). We speculated one possible explanation is that the expression of the gene *DCBLD2* has not yet reached a steady state. Consistent results were also observed from phase portraits of new-total RNA planes of *DCBLD2* ([Fig 6H](#) and [S4F Fig](#)). For gene *HIPK2*, similarly, cells in phase M and M-G1 have the highest expression overall and the correct velocity should be negative ([S4I Fig](#)), but Dynamo and CSP-Baseline both returned positive velocities while CSP-Switching got the correct results ([S4G and S4H Fig](#)). We speculated one possible explanation for this is that the expression switch plays an important role in *HIPK2*.

Finally we generated a simulated non-steady-state pulse data ([S5 Fig](#)). Both the CSP-Baseline and Dynamo produced the correct streamlines ([S5A Fig](#)). However, the error between the degradation rate estimated by CSP-Baseline and the true value was lower compared to

Dynamo, and the error was further reduced for the well-fitted genes selected by goodness-of-fit ([S5B](#) and [S5C Fig](#)).

## Discussion

Storm utilizes three stochastic models for the dynamical description of new mRNAs and allows the estimation of the RNA velocity for kinetics experiments and one-shot data with splicing information without the need for the steady-state assumption. It can also generally handle one-shot data without splicing information when the steady-state assumption is enforced. One possible limitation of our model is that it does not fully utilize the total mRNA information in kinetics experiments. According to the results of the chi-square independence test, the number of total mRNA molecules of most genes obeys the same distribution. Noting that the old mRNA molecules with a labeling duration of zero are the total mRNA molecules, we think that it is a feasible direction to establish the stochastic dynamics of old mRNA and use the Wasserstein distance in optimal transport approach [38, 39] to measure the differences between discrete distributions. Therefore, the optimal transport modeling of old RNAs may be integrated with Storm to obtain more robust RNA velocity inference. In addition, it is also worth exploring stochastic models that consider switching of gene expression states, transcription in the active state, splicing and spliced mRNA degradation simultaneously (i.e., integration of CSP-Splicing and CSP-Switching).

Some recent works, such as MultiVelo [7], Chromatin Velocity [40], and protaccel [8], extend RNA velocity to multi-omics. It is expected that the combination of metabolic labeling technology with other multi-omics measurements will bring new opportunities, which allows for simpler parameter inference and more accurate results.

Although Storm was able to infer cell-specific transcription and splicing rates through the post-processing steps, it still assumes that degradation rates are consistent across cells, which may introduce a potential bias. In addition, Storm like many methods assumes that genes are independent when inferring kinetic parameters (e.g. velocyto [1], scVelo [2], cellDancer [16], and Dynamo [26]), which is biologically implausible. Deep neural networks are expected to solve these problems by directly inferring cell-specific kinetic parameters and vector fields end-to-end in situations where gene regulation is considered. For example, DeepVelo [10] claims to achieve this goal for unspliced/spliced data. How to introduce deep neural networks to scRNA-seq data with metabolic labeling information is a direction worth exploring, and Storm may be able to provide some insights (e.g., loss function design) to achieve this goal.

Finally, Storm, like many other existing methods, first infers the RNA velocity in the high-dimensional gene expression space, then selects an appropriate two-dimensional embedding, and finally visualizes the RNA velocity by projecting it into the low-dimensional space. The two-step process was criticized and may lead to specious results [6, 41–43]. Nevertheless, a large number of efforts have been proposed to compensate for the shortcomings of the two-step process of projecting gene-specific RNA velocities from high-dimensional space to low-dimensional embeddings. For example, UnitVelo [9] supports the inference of a unified latent time across the transcriptome, GraphDynamo [44] maps the cellular dynamics onto a discrete graph representation, PAGA [45] generates a much simpler abstracted directed graph of partitions by using RNA velocity in raw space, and LatentVelo [15], DeepCycle [46] and VeloCycle [47] simultaneously infer the hidden space of gene expression and the dynamics on the hidden space, where DeepCycle [46] and VeloCycle [47] are also specifically designed for cell cycle processes. We tested Storm's inference results using PAGA. The results show that Storm's stochastic modeling strategy is effective ([S6 Fig](#)). However, we would also like to mention that the determination of which visualization is the best choice is not an easy problem. A thorough

discussion about this issue deserves independent publications and detailed comparisons and studies in the future, which is not the main concern of Storm. The key contribution of Storm is the design of parameter inference methods for the scRNA-seq data with metabolic labeling that does not rely on steady-state assumptions.

## Conclusion

We present Storm for estimating absolute kinetic parameters and inferring the time-resolved RNA velocity of metabolic labeling scRNA-seq data by incorporating the transient stochastic dynamics of gene expressions. Storm establishes three stochastic models of new mRNA which take into account both biological noise and cell-specific technical noise, and makes inference to the gene-specific degradation rates and other gene-cell-specific parameters without relying on the steady-state assumption in kinetics experiments and one-shot data with splicing information. It can also handle one-shot data without splicing information when the steady-state assumption is adopted. Numerical results show that Storm is able to accurately fit the kinetic cell cycle dataset and many one-shot experimental datasets. In addition, our numerical experience suggests that CSP-Baseline outperforms the other two models when splicing dynamics is not of interest, and CSP-Splicing is the valid choice if the data contains both labeling and splicing information and splicing dynamics is of interest. However, further applications and performance evaluations for more challenging datasets with temporal information are desired and it will be studied in the future. We hope the developed method will become increasingly important when more metabolic labeling data are available.

## Methods

### Derivation of three stochastic dynamical models

Here we developed three stochastic models for the dynamical description of new mRNAs: Model 1 (CSP-Baseline): a stochastic dynamical model of new mRNA involving only metabolic-labeling transcription and degradation; Model 2 (CSP-Splicing): a stochastic dynamical model of new unspliced and spliced mRNA involving metabolic-labeling transcription, splicing and spliced mRNA degradation; and Model 3 (CSP-Switching): a stochastic dynamical model of new mRNA involving gene state switching, metabolic-labeling transcription and degradation.

**Model 1 (CSP-Baseline): Stochastic dynamical modeling of new mRNA.** Following [21, 26], we made the following assumptions: (1) Genes are independent. (2) Both the transcription rate  $\alpha$  and the degradation rate of total mRNA  $\gamma_t$  are constants.

The chemical master equation (CME) for the new/labeled mRNA  $\tilde{l}(t)$ , corresponding to the chemical reactions shown in the first row of Fig 1A, is given by

$$\frac{d\tilde{P}_n}{dt} = -(\alpha + n\gamma_t)\tilde{P}_n + \alpha\tilde{P}_{n-1} + \gamma_t(n+1)\tilde{P}_{n+1}, \quad (5)$$

where  $\tilde{P}_n(t) = \text{Prob}(\tilde{l}(t) = n)$ . The initial value of new mRNA count is zero, i.e.,  $\tilde{P}_n(0) = \delta_{0n}$ , where

$$\delta_{mn} = \begin{cases} 1, & \text{if } m = n \\ 0, & \text{otherwise} \end{cases}$$

is the Kronecker's delta function. The solution of Eq (5) is

$$\tilde{P}_n(t) = \frac{a(t)^n}{n!} e^{-a(t)}, \quad n \in \mathbb{N}, \quad (6)$$

where  $a(t) = \alpha(1 - e^{-\gamma_t t})/\gamma t$ . This means that  $\tilde{l}(t)$  obeys the Poisson distribution with mean  $a(t)$ .

The above stochastic model only describes the true expression count of new mRNA  $\tilde{l}(t)$  in a cell with labeling duration  $t$ , but the measured sequencing data is different from this count due to technical noise. Denote by  $l(t)$  the number of measured new mRNA molecules, and assume that  $l(t)$  is associated with  $\tilde{l}(t)$  through a dropout process, which we modeled as a binomial distribution:

$$\text{Prob}(l(t) = n | \tilde{l}(t) = N) = C_N^n p^n (1-p)^{N-n} := B_n(N, p), \quad (7)$$

where  $p$  is the capture probability of a single mRNA molecule. We further assume that the total number of mRNA molecules across all genes in different cells are close, which was commonly adopted in the preprocessing step [1, 2, 26]. Denote by  $n_j$  the total number of mRNA molecules across all genes in cell  $j$ , i.e.,  $n_j = \sum_i r_{ij}$ , where  $r_{ij}$  refers to the number of mRNA molecules in gene  $i$  of cell  $j$  in the scRNA-seq measurements. This assumption implies that the capture probability of mRNA molecules in different cells is different, and  $p_j \propto n_j$ . In our computation, we took  $p_j = n_j/n_0$ , where  $n_0 = n_{\text{med}}$  is the median of  $n_j$ . Note that  $p_j$  here is what is commonly referred to as the size factor, which is chosen to be consistent with the deterministic approach. Such choice might make  $p_j > 1$  for some  $j$ . However, this artifact can be easily avoided by taking  $n_0$  larger, e.g.,  $\tilde{n}_0 := \max_j n_j$ . This alternative choice does not affect the inference of the degradation and splicing rates except that the transcription rate  $\alpha$  will be rescaled by the multiple  $\tilde{n}_0/n_{\text{med}}$  from Eq (9) and the form of  $a(t)$ . In this case, the direction of the inferred RNA velocity is not affected up to a common multiplicative constant, and the whole approach is still valid.

We denoted the PMF of new mRNA sequencing result  $l_j(t)$  of cell  $j$  with labeling duration  $t$  by

$$P_{n_j}(t) := \text{Prob}(l_j(t) = n). \quad (8)$$

Then

$$P_{n_j}(t) = \sum_{n=n}^{\infty} \tilde{P}_N(t) B_n(N, p_j) = \frac{(p_j a(t))^n}{n!} e^{-p_j a(t)}, \quad (9)$$

which means that  $l_j(t)$  obeys the Poisson distribution with mean  $p_j a(t)$ .

In summary, the former derivation shows that the number of new mRNA molecules in different cells in scRNA-seq measurements obeys Poisson distribution with cell-specific parameters, and these parameters were proportional to  $p_j$ , i.e., proportional to  $n_j$ . We call this distribution the *cell-specific Poisson distribution*.

**Model 2 (CSP-Splicing): Stochastic dynamical modeling of new unspliced and spliced mRNAs.** Compared with CSP-Baseline, we distinguished whether an mRNA molecule is spliced or not and incorporated the splicing process, which was shown in the second row of Fig 1A. Again we assumed that the genes are independent. In addition, we further assumed that the transcription rate  $\alpha$ , splicing rate  $\beta$ , and spliced mRNA degradation rate  $\gamma_s$  are all constants.

The CME for the new/labeled unspliced and spliced mRNAs ( $\tilde{u}_l(t), \tilde{s}_l(t)$ ), corresponding to the considered chemical reactions shown in the second row of Fig 1A, is given by

$$\begin{aligned}\partial_t \tilde{P}_{mn} &= \alpha(\tilde{P}_{m-1,n} - \tilde{P}_{mn}) + \beta[(m+1)\tilde{P}_{m+1,n-1} - m\tilde{P}_{mn}] \\ &\quad + \gamma_s[(n+1)\tilde{P}_{m,n+1} - n\tilde{P}_{mn}],\end{aligned}\quad (10)$$

where  $\tilde{P}_{mn}(t) = \text{Prob}((\tilde{u}_l(t), \tilde{s}_l(t)) = (m, n))$ . The initial distribution of new unspliced and spliced mRNA is  $\tilde{P}_{mn}(0) = \delta_{m0}\delta_{n0}$ . The solution of Eq (10) is

$$\tilde{P}_{mn}(t) = b(t)^m c(t)^n e^{-b(t)-c(t)} / m!n!, \quad (m, n) \in \mathbb{N}^2, \quad (11)$$

where

$$\begin{aligned}b(t) &= \alpha(1 - e^{-\beta t})/\beta, \\ c(t) &= \begin{cases} \frac{\alpha}{\gamma_s}(1 - e^{-\gamma_s t}) + \frac{\alpha}{\gamma_s - \beta}(e^{-\gamma_s t} - e^{-\beta t}), & \beta \neq \gamma_s, \\ \frac{\alpha}{\beta}(1 - e^{-\beta t}) - \alpha t e^{-\beta t}, & \beta = \gamma_s, \end{cases}\end{aligned}\quad (12)$$

which means that  $\tilde{u}_l(t)$  and  $\tilde{s}_l(t)$  obey independent Poisson distributions with mean  $b(t)$  and  $c(t)$ , respectively. We refer interested readers to [3] for derivation details.

Denote by  $(u_l(t), s_l(t))$  the number of measured new unspliced and spliced mRNA molecules in the scRNA-seq experiments with labeling duration  $t$ . By assuming that the dropout processes for new unspliced and spliced mRNAs are independent and the capture probability is independent of whether they are spliced or not, we modeled the dropout process for  $\tilde{u}_l(t)$  and  $\tilde{s}_l(t)$  as independent binomial distributions with the same parameter  $p$ . So we got

$$\begin{aligned}\text{Prob} &((u_l(t), s_l(t)) = (m, n) \mid (\tilde{u}_l(t), \tilde{s}_l(t)) = (M, N)) \\ &= C_M^m p^m (1-p)^{M-m} C_N^n p^n (1-p)^{N-n} := B_m(M, p) B_n(N, p).\end{aligned}\quad (13)$$

For the same reason as CSP-Baseline, we take  $p_j$  proportional to  $n_j$ . And we took  $p_j = n_j/n_{\text{med}}$  in the computation.

We denoted the joint PMF of new unspliced and spliced mRNA sequencing counts  $(u_{l,j}(t), s_{l,j}(t))$  of cell  $j$  with labeling duration  $t$  by

$$P_{mn,j}(t) := \text{Prob}((u_{l,j}(t), s_{l,j}(t)) = (m, n)).$$

Then

$$\begin{aligned}P_{mn,j}(t) &= \sum_{M=m}^{\infty} \sum_{N=n}^{\infty} \frac{b(t)^M c(t)^N}{M!N!} e^{-b(t)-c(t)} B_M(m, p_j) B_N(n, p_j) \\ &= \sum_{M=m}^{\infty} \frac{b(t)^M}{M!} e^{-b(t)} B_M(m, p_j) \sum_{N=n}^{\infty} \frac{c(t)^N}{N!} e^{-c(t)} B_N(n, p_j) \\ &= \frac{(p_j b(t))^m}{m!} e^{-p_j b(t)} \frac{(p_j c(t))^n}{n!} e^{-p_j c(t)},\end{aligned}\quad (14)$$

which means that  $u_{l,j}(t)$  and  $s_{l,j}(t)$  are independently Poisson distributed with mean  $p_j b(t)$  and  $p_j c(t)$ , respectively.

In summary,  $(u_l(t), s_l(t))$  obeys independent cell-specific Poisson distribution.

**Model 3 (CSP-Switching): Stochastic dynamical modeling of new mRNA considering switching.** In CSP-Switching, we further considered the on/off gene state switching shown in the third row of Fig 1A. We assumed that the genes are independent as well, and the transcription rate  $\alpha$ , mRNA degradation rate  $\gamma_b$ , the gene on-to-off rate  $k_{\text{off}}$  and off-to-on rate  $k_{\text{on}}$  are all constants. Furthermore, following [32] we assumed that  $k_{\text{on}}$  and  $k_{\text{off}}$  are significantly smaller than  $\alpha$  and  $\gamma_b$ , which implies that the gene expression is either always on or always off during the transcription/degradation period. From Eq (5), it is known that cells in the on state obey a Poisson distribution with mean  $a(t)$ , while cells in the off state do not express. Define  $p_{\text{off}} = k_{\text{off}}/(k_{\text{on}} + k_{\text{off}})$ . Then  $\tilde{l}(t)$  obeys the zero-inflated Poisson distribution

$$\begin{aligned}\tilde{P}_0(t) &= (1 - p_{\text{off}})e^{-a(t)} + p_{\text{off}}, \\ \tilde{P}_n(t) &= (1 - p_{\text{off}}) \frac{a(t)^n}{n!} e^{-a(t)}, \quad n \geq 1.\end{aligned}\quad (15)$$

Similarly, by taking into account the technical noise in scRNA-seq experiments, the PMF of  $l_j(t)$  is

$$\begin{aligned}P_{0,j}(t) &= (1 - p_{\text{off}})e^{-p_j a(t)} + p_{\text{off}}, \\ P_{n,j}(t) &= (1 - p_{\text{off}}) \frac{(p_j a(t))^n}{n!} e^{-p_j a(t)}, \quad n \geq 1.\end{aligned}\quad (16)$$

In summary, different cells obey the ZIP distribution with different parameters as shown in Eq (16), which we called cell-specific zero-inflated Poisson distribution.

### Chi-square goodness-of-fit test for cell-specific distributions at a fixed time

We would construct an asymptotic  $\chi^2$  statistic for the data with common distribution type but sample-specific parameters. This goodness-of-fit test is to assess whether the null hypothesis that the considered data, at a fixed labeling duration, obeys the proposed distribution can be accepted.

We first divided the value range of the considered data into  $c$  classes. According to the range that the samples fall in, we got  $n$  independent categorically distributed random samples  $X_i \in \{1, 2, \dots, c\}$  for  $i = 1, 2, \dots, n$  with sample dependent parameter  $p_i$ , respectively. An equivalent representation for the categorical variable  $X_i$  is to denote  $X_i = (X_{ij})_{j=1, \dots, c} \in \{e_1, \dots, e_c\}$ , where  $e_j = (\delta_{jk})_{k=1, \dots, c}$  is the indicator vector for  $j = 1, \dots, c$ . Correspondingly, the parameter  $p_i = (p_{i1}, \dots, p_{ic})^T$  is a  $c$ -dimensional vector with non-negative elements and sums to one, which is defined as

$$p_{ij} := \text{Prob}(X_{ij} = 1) = 1 - \text{Prob}(X_{ij} = 0), \quad j = 1, \dots, c. \quad (17)$$

This implies that  $\text{Var}(X_{ij}) = p_{ij}(1 - p_{ij})$  and  $\text{Cov}(X_{ij}, X_{il}) = \mathbb{E}[X_{ij}X_{il}] - p_{ij}p_{il} = -p_{ij}p_{il}$  for  $j \neq l$ . Therefore, the covariance matrix of random vector  $X_i$  is

$$\Sigma_i = \begin{pmatrix} p_{i1}(1 - p_{i1}) & -p_{i1}p_{i2} & \cdots & -p_{i1}p_{ic} \\ -p_{i1}p_{i2} & p_{i2}(1 - p_{i2}) & \cdots & -p_{i2}p_{ic} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{ic}p_{i1} & -p_{ic}p_{i2} & \cdots & p_{ic}(1 - p_{ic}) \end{pmatrix}. \quad (18)$$

For sample  $i$ , we defined the truncated random vector  $X_i^* = (X_{i1}, \dots, X_{ic-1})^T$  and truncated vector  $p_i^* = (p_{i1}, \dots, p_{ic-1})^T$ , which is the first  $c - 1$  components of  $X_i$  and  $p_i$ , respectively. The covariance matrix of  $X_i^*$  is the submatrix consisting of the upper-left  $(c - 1) \times (c - 1)$  block of  $\Sigma_i$ , denoted by  $\Sigma_i^*$ , which can be written as

$$\Sigma_i^* = \text{diag}(p_i^*) - p_i^*(p_i^*)^T, \quad (19)$$

where  $\text{diag}(p_i^*)$  is the diagonal matrix formed by the components of  $p_i^*$ .

Define  $\bar{X}^* := (\sum_{i=1}^n X_i^*)/n$ ,  $\bar{p}^* := (\sum_{i=1}^n p_i^*)/n$  and  $\bar{\Sigma}^* := (\sum_{i=1}^n \Sigma_i^*)/n$ , and let

$$\chi^2 := n(\bar{X}^* - \bar{p}^*)^T (\bar{\Sigma}^*)^{-1} (\bar{X}^* - \bar{p}^*). \quad (20)$$

Below we would show that  $\chi^2$  is an asymptotic chi-square statistic with degrees of freedom  $c - 1$ . First note that

$$E[\bar{X}^*] = E\left[\frac{1}{n} \sum_{i=1}^n X_i^*\right] = \frac{1}{n} \sum_{i=1}^n E[X_i^*] = \frac{1}{n} \sum_{i=1}^n p_i^* = \bar{p}^*, \quad (21)$$

then the covariance

$$D[\bar{X}^*] = D\left[\frac{1}{n} \sum_{i=1}^n X_i^*\right] = \frac{1}{n^2} \sum_{i=1}^n D[X_i^*] = \frac{1}{n} \left( \frac{1}{n} \sum_{i=1}^n \Sigma_i^* \right) = \frac{1}{n} \bar{\Sigma}^*. \quad (22)$$

Let  $Y_n = \sqrt{n}(\bar{\Sigma}^*)^{-1/2}(\bar{X}^* - \bar{p}^*)$ . When  $n$  goes to infinity,  $Y_n$  converges in distribution to the normal distribution  $N(0, I_{c-1})$  according to the central limit theorem for the independent sum of random variables. Thus,  $\chi^2 = Y_n^T Y_n$  converges in distribution to a chi-square distribution with degrees of freedom  $c - 1$ .

In summary, we proposed a new asymptotic  $\chi^2$  statistic for sample-specific distributions. For a fixed labeling duration  $t_{\text{fixed}}$ ,  $a(t_{\text{fixed}})$ ,  $b(t_{\text{fixed}})$  and  $c(t_{\text{fixed}})$  are all constants, the proposed  $\chi^2$  statistics can be used to test whether the new mRNA sequencing data are consistent with the CSP, ICSP and CSZIP distributions based on Models CSP-Baseline, CSP-Splicing and CSP-Switching, respectively. In addition, since there are one, two and two parameters to be inferred in CSP, ICSP and CSZIP distributions, respectively, the same number of degrees of freedom should be subtracted. Following [28], we ensured that the expected count  $np_j \geq 0.25$  in each group when determining the group value ranges. Finally, we take  $p$ -value as 0.05 in the computation.

## Parameter inference in one-shot experiments with steady state assumption

In the one-shot experiments where we only observe new RNA  $l_j(t)$  and total RNA  $r_j(t)$  data for one labeling duration  $t$ , we had to invoke the steady-state assumption for the total RNA.

When the dynamics of total RNA in CSP-Baseline is at steady state, i.e.,

$$0 = \frac{d\tilde{P}_{r,n}}{dt} = -(\alpha + n\gamma_t)\tilde{P}_{r,n} + \alpha\tilde{P}_{r,n-1} + \gamma_t(n+1)\tilde{P}_{r,n+1}, \quad (23)$$

where  $\tilde{P}_{r,n} := \text{Prob}(\tilde{r} = n)$  is the invariant PMF of the true expression of total RNA. At this point  $\tilde{P}_{r,n}$  is a Poisson distribution with  $\alpha/\gamma_t$  as the mean and from [3] we know that at this point the PMF of the true expression of old RNA  $\tilde{P}_{u,n}$  is a Poisson distribution with  $(\alpha/\gamma_t)e^{-\gamma_t t}$  as the mean. From Eq (9) we know that when technical noise is considered, the observed old

RNA counts obey a similar CSP distribution

$$P_{u,n_j} = \frac{(p_j(\alpha/\gamma_t) e^{-\gamma_t t})^n}{n!} e^{-p_j(\alpha/\gamma_t) e^{-\gamma_t t}}. \quad (24)$$

At this point, we obtained the distributions of the new RNA and old RNA observations so that parameter inference can be performed using the MLE. Notice that the distribution of new RNA counts is not independent of total RNA counts, whereas the distribution of new RNA and old RNA counts is independent. We want to maximize the log-likelihood function

$$\begin{aligned} \ell(\alpha, \gamma_t) &= \sum_{j=1}^n \log (\text{Prob}(r = r_j, l = l_j)) \\ &= \sum_{j=1}^n \log (\text{Prob}(u = r_j - l_j, l = l_j)) \\ &= \sum_{j=1}^n \log (\mathcal{P}(p_j a(t))|_{l_j}) + \log (\mathcal{P}(p_j(\alpha/\gamma_t) e^{-\gamma_t t})|_{r_j - l_j}) \end{aligned} \quad (25)$$

where  $\mathcal{P}(\lambda)|_n := \text{Prob}(X = n) = e^{-\lambda} \lambda^n / n!$  is the probability of  $X = n$  for a Poisson-distributed random variable  $X$  with mean  $\lambda$ . When  $\partial \ell / \partial \alpha = 0$  and  $\partial \ell / \partial \gamma_t = 0$ , the likelihood function is maximized and it can be solved analytically

$$\gamma_t = -\frac{1}{t} \log \left( 1 - \frac{\langle l_j \rangle}{\langle r_j \rangle} \right), \quad \alpha = \gamma_t \frac{\langle r_j \rangle}{\langle p_j \rangle}, \quad (26)$$

where  $\langle \cdot \rangle$  means the population average defined by

$$\langle \cdot \rangle = \left( \sum_{k=1}^K \sum_{j=1}^{n_k} (\cdot) \right) / \left( \sum_{k=1}^K n_k \right). \quad (27)$$

Since here it is for the one-shot data set,  $K = 1$ . Note that Eq (26) is similar to the formula in Dynamo [26] for estimating the parameters for one-shot data. The difference is that this formula averages the raw counts, while the method in Dynamo averages the smoothed data.

### Parameter inference in one-shot experiments without steady state assumption

In one-shot dataset containing both labeling and splicing information, i.e. unspliced unlabeled RNA  $u_{u,j}$ , unspliced labeled RNA  $u_{l,j}$ , spliced unlabeled RNA  $s_{u,j}$  and spliced labeled RNA  $s_{l,j}$  information is observed, we can make parametric inference without relying on the steady-state assumption.

The method is divided into two steps; in the first step, we sum unspliced unlabeled RNA and unspliced labeled RNA to obtain unspliced RNA, and unspliced labeled RNA and spliced labeled RNA to obtain spliced RNA. Then, we use the dynamic model without relying on steady state assumption in scVelo proposed by Bergen et al. [2] to infer the observation time of cells  $t_{\text{obs},j}$ , switching time  $t_s$ , degradation rate  $\gamma_{s,\text{scv}}$ , and splicing rate  $\beta_{\text{scv}}$ . Despite the problem of scale invariance, the absolute magnitude of these values is not physically meaningful, but can still provide useful information for inferring the absolute magnitude of the parameters, e.g., the value of  $\beta_{\text{scv}}/\gamma_{s,\text{scv}}$  is meaningful, and in addition the cells with  $t_{\text{obs},j}$  less than  $t_s$  are in the on state.

In the second step, we integrate the results from the first step and the labeling information to determine the absolute size of the parameter. We define the notation

$$S_{\text{on}} := \{j | t_{\text{obs},j} < t_s\}$$

to be the set of cells in the on state. In CSP-Splicing, the  $u_{l,j}$  and  $s_{l,j}$  of cells in the on state obey ICSP distribution

$$P_{mn,j}(t) = \frac{(p_j b(t))^m}{m!} e^{-p_j b(t)} \frac{(p_j c(t))^n}{n!} e^{-p_j c(t)}, \quad \forall j \in S_{\text{on}}. \quad (28)$$

We want to maximize the log-likelihood function

$$\ell(\alpha, \beta, \gamma_s) = \sum_{j \in S_{\text{on}}} \log (\mathcal{P}(p_j(t)b(t))|_{u_{l,j}(t)} \cdot \mathcal{P}(p_j(t)c(t))|_{s_{l,j}(t)}), \quad (29)$$

which is equivalent to

$$\begin{aligned} \frac{\sum_{j \in S_{\text{on}}} u_{l,j}}{\sum_{j \in S_{\text{on}}} p_j} &= b(t) = \frac{\alpha}{\beta} (1 - e^{-\beta t}), \\ \frac{\sum_{j \in S_{\text{on}}} s_{l,j}}{\sum_{j \in S_{\text{on}}} p_j} &= c(t) = \frac{\alpha}{\gamma_s} (1 - e^{-\gamma_s t}) + \frac{\alpha}{\gamma_s - \beta} (e^{-\gamma_s t} - e^{-\beta t}). \end{aligned} \quad (30)$$

This problem is not well-defined, but it is after adding the result

$$\frac{\beta}{\gamma_s} = \frac{\beta_{\text{scv}}}{\gamma_{s,\text{scv}}} \quad (31)$$

from the first step. By solving the system of nonlinear equations consisting of equations (30) and (31), we can obtain the absolute magnitudes of  $\alpha$ ,  $\beta$ , and  $\gamma_s$ .

Since modeling assumptions in scVelo are often violated, we selected only well-fitting genes for use in the second stage and downstream analyses. We use  $R^2$  as the goodness-of-fit, which is defined as

$$R^2 = 1 - \frac{\sum_{j=1}^N ((u_j, s_j) - (u(t_{\text{obs},j}), s(t_{\text{obs},j})))^2}{\sum_{j=1}^N ((u_j, s_j) - (\bar{u}, \bar{s}))^2}. \quad (32)$$

Due to the dropout effect, cells with expression close to 0 are not used in the actual calculation of  $R^2$ . More specifically the rule is  $(u_j, s_j) < (\max(u_j)/5, \max(s_j)/5)$ .

## Parameter inference in kinetics experiments

In the kinetics experiments, we observed data  $l_j(t_k)$  (or  $(u_{l,j}(t_k), s_{l,j}(t_k))$ ) for new mRNA (or new unspliced and spliced mRNAs) with different labeling durations. We assumed that there are  $K$  labeling durations  $t_k$  for  $k = 1, 2, \dots, K$ , and the number of cells with labeling duration  $t_k$  is  $n_k$ . We utilized the MLE to infer the unknown parameters in different models without relying on steady-state assumptions.

In CSP-Baseline, we need to maximize the log-likelihood function

$$\ell(\alpha, \gamma_t) = \sum_{k=1}^K \sum_{j=1}^{n_k} \log (\mathcal{P}(p_j(t_k)a(t_k))|_{l_j(t_k)}). \quad (33)$$

It is equivalent to minimizing the following loss function

$$L(\alpha, \gamma_t) = \sum_{k=1}^K \sum_{j=1}^{n_k} -l_j(t_k) \log(p_j(t_k)a(t_k)) + p_j(t_k)a(t_k). \quad (34)$$

The optimum of the loss is achieved when the gradient equals 0. Utilizing the concrete expression of  $a(t)$  in CSP-Baseline, we got  $\partial a(t)/\partial \alpha = (1 - e^{-\gamma_t t})/\gamma t$ . Then  $\partial L(\alpha, \gamma_t)/\partial \alpha = 0$  has a closed form solution

$$\alpha(\gamma_t) = \frac{\langle l_j(t_k) \rangle}{\langle p_j(t_k) \partial a(t_k)/\partial \alpha \rangle}. \quad (35)$$

Another component of the Euler-Lagrange equation  $\partial L/\partial \gamma_t = 0$  has no closed form solution, so we need to solve  $\gamma_t$  by numerical iterations. We took the initial value of  $\gamma_t$  as the solution from Dynamo [26] under the steady-state assumption. Denote it as  $\gamma_{t,0}$ , and correspondingly, we take the initial value of  $\alpha$  as  $\alpha_0 = \alpha(\gamma_{t,0})$ .

In CSP-Splicing, we need to maximize the log-likelihood function

$$\ell(\alpha, \beta, \gamma_s) = \sum_{k=1}^K \sum_{j=1}^{n_k} \log (\mathcal{P}(p_j(t_k)b(t_k))|_{u_{l,j}(t_k)} \cdot \mathcal{P}(p_j(t_k)c(t_k))|_{s_{l,j}(t_k)}), \quad (36)$$

which is equivalent to minimizing the loss function

$$\begin{aligned} L(\alpha, \beta, \gamma_s) &= \sum_{k=1}^K \sum_{j=1}^{n_k} (-u_{l,j}(t_k) \log(p_j(t_k)b(t_k)) + p_j(t_k)b(t_k)) \\ &\quad + (-s_{l,j}(t_k) \log(p_j(t_k)c(t_k)) + p_j(t_k)c(t_k)). \end{aligned} \quad (37)$$

Utilizing (12), we got  $\partial b(t)/\partial \alpha = (1 - e^{-\beta t})/\beta$  and  $\partial c(t)/\partial \alpha = (1 - e^{-\gamma_s t})/\gamma_s + (e^{-\gamma_s t} - e^{-\beta t})/(\gamma_s - \beta)$  when  $\beta \neq \gamma_s$ . The case for  $\beta = \gamma_s$  is similar. So  $\partial L(\alpha, \beta, \gamma_s)/\partial \alpha = 0$  has a closed form solution

$$\alpha(\beta, \gamma_s) = \frac{\langle u_{l,j}(t_k) + s_{l,j}(t_k) \rangle}{\langle p_j(t_k) \left( \frac{\partial b}{\partial \alpha}(t_k) + \frac{\partial c}{\partial \alpha}(t_k) \right) \rangle}. \quad (38)$$

However  $\partial L/\partial \beta = 0$  and  $\partial L/\partial \gamma_s = 0$  have no closed form solution, and we need to solve these equations by iterations. The choice of initial values is similar to the CSP-Baseline case. We took the initial value of  $\beta$  and  $\gamma_s$  as the solution from Dynamo [26] under the steady-state assumption, which we denoted as  $\beta_0, \gamma_{s,0}$ . And then the initial value of  $\alpha$  is taken as  $\alpha_0 = \alpha(\beta_0, \gamma_{s,0})$ .

In CSP-Switching, we need to maximize the log-likelihood function

$$\begin{aligned} \ell(p_{\text{off}}) &= \sum_{k=1}^K \sum_{j=1}^{n_k} \mathbb{I}_{\{l_j(t_k)=0\}} \log(\text{ZIP}(p_j a(t_k), p_{\text{off}})|_0) \\ &\quad + \mathbb{I}_{\{l_j(t_k)>0\}} \log(\text{ZIP}(p_j a(t_k), p_{\text{off}})|_{l_j(t_k)}), \end{aligned} \quad (39)$$

where  $\text{ZIP}(\lambda, p_{\text{off}})|_n := \text{Prob}(X = n)$  is the probability of  $X = n$  for a ZIP-distributed random

variable  $X$  with parameters  $\lambda$  and  $p_{\text{off}}$ . It is equivalent to minimizing the loss function

$$L(\alpha, \gamma_t, p_{\text{off}}) = \sum_{k=1}^K \sum_{j=1}^{n_k} -\log(\text{ZIP}(p_j a(t_k), p_{\text{off}})|_0) - \mathbb{I}_{\{l_j(t_k) > 0\}} (\log(1 - p_{\text{off}}) + l_j(t_k) \log(p_j(t_k)a(t_k)) - p_j(t_k)a(t_k)). \quad (40)$$

Similar as before, we chose the initial value of  $\gamma_t$ , denoted as  $\gamma_{t,0}$ , based on the steady state assumption, and chose the moment estimator

$$p_{\text{off},0} = 1 - \frac{\langle l_j(t_k) \rangle^2 \left\langle \left( p_j(t_k) \frac{\partial a}{\partial \alpha}(t_k) \right)^2 \right\rangle}{\langle p_j(t_k) \frac{\partial a}{\partial \alpha}(t_k) \rangle^2 (\langle l_j(t_k)^2 \rangle - \langle l_j(t_k) \rangle)} \quad (41)$$

and

$$\alpha_0 = \frac{\langle l_j(t_k) \rangle}{(1 - p_{\text{off},0}) \langle p_j(t_k) \frac{\partial a}{\partial \alpha}(t_k) \rangle} \quad (42)$$

as the initial values of  $p_{\text{off}}$  and  $\alpha$ .

According to the biological meaning of the parameters, we added the constraints  $0 < \alpha < 10\alpha_0$ ,  $0 < \beta < 10\beta_0$ ,  $0 < \gamma_t < 10\gamma_{t,0}$ ,  $0 < \gamma_s < 10\gamma_{s,0}$  and  $0 < p_{\text{off}} < 1$ , and we called the SLSQP optimizer in SciPy to solve the above optimization problem.

## Goodness-of-fit test for the distribution evolution in time

In ordinary least squares (OLS) linear regression, people often use

$$R^2 := 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (43)$$

to define the goodness of fit, where  $y_i$  is the sample observation,  $\hat{y}_i$  is the model prediction, and  $\bar{y}_i$  is the sample mean. For the generalized linear model (GLM), the  $R^2$  can be defined using the deviance  $D$  and null deviance  $D_0$  [29],

$$R_D^2 := 1 - \frac{D}{D_0} = 1 - \frac{-2(\ell(\hat{\beta}) - \ell_s)}{-2(\ell_0 - \ell_s)} = 1 - \frac{\ell(\hat{\beta}) - \ell_s}{\ell_0 - \ell_s}, \quad (44)$$

where  $\ell(\hat{\beta})$ ,  $\ell_0$  and  $\ell_s$  denotes the log-likelihood function of the model with parameter  $\hat{\beta}$ , the null model (that is, fitted with only the intercept), and the saturated model (that is, fitted with one parameter per sample), respectively.  $R_D^2$  can be seen as a generalization of  $R^2$ , which is equal to  $R^2$  when the model is a least squares linear regression [29]. Finally, to overcome the disadvantage of adding more parameters without reducing  $R_D^2$  (similar to  $R^2$ ), we used adjusted  $R_D^2$  (denoted as  $\bar{R}_D^2$ ) as the goodness of fit of different models, which is defined as

$$\bar{R}_D^2 := 1 - \frac{D/d_D}{D_0/d_{D_0}} = 1 - \frac{(\ell(\hat{\beta}) - \ell_s)/d_D}{(\ell_0 - \ell_s)/d_{D_0}}, \quad (45)$$

where  $d_D$  and  $d_{D_0}$  are the degrees of freedom of  $D$  and  $D_0$ , respectively. To more intuitively explain the usefulness of  $R_D^2$ , we named it goodness of fit of model in the main text instead of using adjusted deviance  $R^2$ .

In CSP-Baseline,  $\ell_s$  has the closed form

$$\ell_s = \sum_{k=1}^K \sum_{j=1}^{n_k} l_j(t_k) \log (\mathcal{P}(l_j(t_k))|_{l_j(t_k)}). \quad (46)$$

To calculate  $\ell_0$ , we need to maximize the log-likelihood function

$$\ell(a_0) = \sum_{k=1}^K \sum_{j=1}^{n_k} \log (\mathcal{P}(p_j(t_k)a_0)|_{l_j(t_k)}), \quad (47)$$

where  $a_0$  is the intercept. The problem has a closed form solution  $a_0 = \langle l_j(t_k) \rangle / \langle p_j(t_k) \rangle$ . In addition,  $d_D = N - 2$  and  $d_{D_0} = N - 1$ , where  $N$  is the number of cells.

In CSP-Splicing,  $\ell_s$  has the closed form

$$\ell_s = \sum_{k=1}^K \sum_{j=1}^{n_k} \log (\mathcal{P}(u_{l,j}(t_k))|_{u_{l,j}(t_k)}) + \log (\mathcal{P}(s_{l,j}(t_k))|_{s_{l,j}(t_k)}) \quad (48)$$

To calculate  $\ell_0$ , we need to maximize the log-likelihood function

$$\ell(b_0, c_0) = \log (\mathcal{P}(p_j(t_k)b_0)|_{u_{l,j}(t_k)}) + \log (\mathcal{P}(p_j(t_k)c_0))|_{s_{l,j}(t_k)} \quad (49)$$

where  $b_0$  and  $c_0$  are intercepts and have closed form solutions  $b_0 = \langle u_{l,j}(t_k) \rangle / \langle p_j(t_k) \rangle$  and  $c_0 = \langle s_{l,j}(t_k) \rangle / \langle p_j(t_k) \rangle$ , respectively. In addition,  $d_D = 2N - 3$  and  $d_{D_0} = 2N - 2$ .

In CSP-Switching, to calculate  $\ell_s$ , we need to maximize the log-likelihood function

$$\begin{aligned} \ell(\alpha, \gamma_t, p_{\text{off}}) &= \sum_{k=1}^K \sum_{j=1}^{n_k} \mathbb{I}_{\{l_j(t_k)=0\}} \log (\text{ZIP}(0, p_{\text{off}})|_0) \\ &\quad + \mathbb{I}_{\{l_j(t_k)>0\}} \log (\text{ZIP}(l_j(t_k), p_{\text{off}})|_{l_j(t_k)}) \\ &= \sum_{k=1}^K \sum_{j=1}^{n_k} \mathbb{I}_{\{l_j(t_k)>0\}} \log (\text{ZIP}(l_j(t_k), p_{\text{off}})|_{l_j(t_k)}) \end{aligned} \quad (50)$$

When  $p_{\text{off}}$  is equal to zero, Eq (50) is maximized, and the closed form solution of  $\ell_s$  is

$$\ell_s = \sum_{k=1}^K \sum_{j=1}^{n_k} \mathbb{I}_{\{l_j(t_k)>0\}} (l_j(t_k) \log(l_j(t_k)) - l_j(t_k) - \log(l_j(t_k)!)). \quad (51)$$

To calculate  $\ell_0$ , we need to maximize the log-likelihood function

$$\begin{aligned} \ell(a_0, p_{\text{off}}) &= \sum_{k=1}^K \sum_{j=1}^{n_k} \mathbb{I}_{\{l_j(t_k)=0\}} \log (\text{ZIP}(p_j a_0, p_{\text{off}})|_0) \\ &\quad + \mathbb{I}_{\{l_j(t_k)>0\}} \log (\text{ZIP}(p_j a_0, p_{\text{off}})|_{l_j(t_k)}). \end{aligned} \quad (52)$$

Similar to solving Eq (40),  $p_{\text{off},0}$  and  $a_0$  were initialized using moment estimators with additional constraints  $0 < p_{\text{off}} < 1$  and  $0 < a < 10a_0$ . We then called the SLSQP optimizer in SciPy to solve the problem. In addition,  $d_D = N - 2$  and  $d_{D_0} = N - 1$ . Before projecting high-dimensional RNA velocities to low dimensions for visualization, we first pick genes with higher goodness-of-fit. The default setting is that the top 40% of genes are picked, and this percentage can be specified by the user.

## Post-processing for cell-specific parameters

In our cell-specific modeling of gene expression, we only assumed that  $\gamma_t$  (in CSP-Baseline and CSP-Switching) and  $\gamma_s$  (in CSP-Splicing) are constants over cells and are inferred based on the corresponding stochastic models, while the other parameters are cell-specific and continuously dependent on gene expressions. This relaxed assumption implies that only the degradation rate is common to all cells, and only cells with similar gene expressions have similar other parameters (due to continuous dependence). To realize this assumption, we first constructed the k-nearest neighbor (kNN) graph of cells by a data preprocessing. The cell-specific parameter inference was performed by applying the inference to the kNN graph for each cell with local constant parameter assumption and already inferred degradation rates. In other deterministic model-based methods to infer RNA velocity (either unspliced/spliced-based or new/total-based) [1, 2, 26], they also construct similar kNN graph and perform kNN smoothing on the data based on this graph. This post-processing step of ours can be seen as the generalization of the usual kNN smoothing to our stochastic setting, as we model discrete counts. The inference details for our three models were shown as below.

In CSP-Baseline, we have

$$l_i(t_k) \sim \text{Poisson}(p_i a_j(t_k)), \quad \forall i \in \mathcal{N}_{j,t_k}, \quad (53)$$

where  $\mathcal{N}_{j,t_k}$  denotes the set of top  $k$  (default is 30) cells that have the most similar gene expressions as the  $j$ th cell with labeling duration  $t_k$  (including itself) and  $a_j(t_k) = \alpha_j(t_k)(1 - e^{-\gamma_t t_k})/\gamma_t$ . Assuming that  $\gamma_t$  has been inferred, we can obtain a local estimator

$$\frac{\sum_{i \in \mathcal{N}_{j,t_k}} l_i(t_k)}{\sum_{i \in \mathcal{N}_{j,t_k}} p_i(t_k)} = a_j(t_k) = \frac{\alpha_j(t_k)}{\gamma_t} (1 - e^{-\gamma_t t_k}) \quad (54)$$

by using the MLE. Define  $\hat{l}_j(t_k) = (\sum_{i \in \mathcal{N}_{j,t_k}} l_i(t_k)) / (\sum_{i \in \mathcal{N}_{j,t_k}} p_i(t_k))$ . Then the cell-specific transcription rate  $\alpha_j(t_k)$  has a closed form solution

$$\alpha_j(t_k) = \hat{l}_j(t_k) \gamma_t / (1 - e^{-\gamma_t t_k}). \quad (55)$$

In CSP-Splicing, we have

$$(u_{l,i}(t_k), s_{l,i}(t_k)) \sim \text{independent Poisson}(p_i b_j(t_k), p_i c_j(t_k)), \quad \forall i \in \mathcal{N}_{j,t_k}. \quad (56)$$

Similarly, assuming  $\gamma_s$  has been inferred, and defining the local estimators

$$\hat{u}_{l,j}(t_k) = \frac{\sum_{i \in \mathcal{N}_{j,t_k}} u_{l,i}(t_k)}{\sum_{i \in \mathcal{N}_{j,t_k}} p_i(t_k)}, \quad \hat{s}_{l,j}(t_k) = \frac{\sum_{i \in \mathcal{N}_{j,t_k}} s_{l,i}(t_k)}{\sum_{i \in \mathcal{N}_{j,t_k}} p_i(t_k)}, \quad (57)$$

we have

$$\begin{aligned} \hat{u}_{l,j}(t_k) &= b_j(t_k) = \frac{\alpha_j(t_k)}{\beta_j(t_k)} (1 - e^{-\beta_j(t_k)t_k}), \\ \hat{s}_{l,j}(t_k) &= c_j(t_k) = \frac{\alpha_j(t_k)}{\gamma_s} (1 - e^{-\gamma_s t_k}) + \frac{\alpha_j(t_k)}{\gamma_s - \beta_j(t_k)} (e^{-\gamma_s t_k} - e^{-\beta_j(t_k)t_k}), \end{aligned} \quad (58)$$

which is a nonlinear system. We have

$$\frac{\hat{s}_{l,j}(t_k)}{\hat{u}_{l,j}(t_k)} = \frac{\beta_j(t_k)(1 - e^{-\gamma_s t_k})}{\gamma_s(1 - e^{-\beta_j(t_k)t_k})} + \frac{\beta_j(t_k)(e^{-\gamma_s t_k} - e^{-\beta_j(t_k)t_k})}{(\gamma_s - \beta_j(t_k))(1 - e^{-\beta_j(t_k)t_k})}. \quad (59)$$

To solve  $\beta_j(t_k)$ , we set its initial value as previously inferred  $\beta$  by global constant assumption. We then call the *root* function in SciPy to solve the nonlinear equation (59) to get  $\beta_j(t_k)$ . The  $\alpha_j(t_k)$  has a closed form solution

$$\alpha_j(t_k) = \hat{u}_{l,j}(t_k)\beta_j(t_k)/(1 - e^{-\beta_j(t_k)t_k}). \quad (60)$$

In summary, in CSP-Splicing, we can infer the cell-specific transcription rate  $\alpha_j(t_k)$  and splicing rate  $\beta_j(t_k)$ .

In CSP-Switching, we have

$$l_i(t_k) \sim \text{ZIP}(p_i a_j(t_k), p_{\text{off},j}(t_k)), \quad \forall i \in \mathcal{N}_{j,t_k}. \quad (61)$$

When computing RNA velocity, we only need to know  $\alpha_j(t_k)(1 - p_{\text{off},j}(t_k))$  as a whole, and not their respective values (see next subsection). To simplify the computation, we used the moment estimation instead of MLE, and got

$$\hat{l}_j(t_k) = (1 - p_{\text{off},j}(t_k))a_j(t_k) = \frac{(1 - p_{\text{off},j}(t_k))\alpha_j(t_k)}{\gamma_t}(1 - e^{-\gamma_t t_k}). \quad (62)$$

Similarly, assuming  $\gamma_t$  has been inferred,  $\alpha_j(t_k)(1 - p_{\text{off},j}(t_k))$  has a closed form solution

$$\alpha_j(t_k)(1 - p_{\text{off},j}(t_k)) = \hat{l}_j(t_k)\gamma_t/(1 - e^{-\gamma_t t_k}). \quad (63)$$

## Reduction from stochastic to deterministic models for RNA velocity

We used discrete counts data in the proposed parameter inference and goodness-of-fit calculation via stochastic models. However, when we need to compute and visualize the RNA velocity, we should take the reduction from stochastic to deterministic models to get the mean velocity. Below we would show the reduction process and reveal the connection between the stochastic and their corresponding deterministic models.

In CSP-Baseline, let us denote the mean value of  $\tilde{l}(t)$  by  $\langle \tilde{l}(t) \rangle$ , which is defined as  $\langle \tilde{l}(t) \rangle = \sum_{n=1}^{\infty} n \tilde{P}_n(t)$ . From Eq (5) we can obtain the deterministic equation after suitable algebraic manipulations

$$\begin{aligned} \frac{d\langle \tilde{l}(t) \rangle}{dt} &= \sum_{n=1}^{\infty} n \frac{d\tilde{P}_n(t)}{dt} \\ &= \sum_{n=1}^{\infty} n(-(\alpha + n\gamma_t)\tilde{P}_n + \alpha\tilde{P}_{n-1} + \gamma_t(n+1)\tilde{P}_{n+1}) \\ &= \alpha - \gamma_t \langle \tilde{l}(t) \rangle. \end{aligned} \quad (64)$$

Similarly, the mean value of total RNA  $\tilde{r}(t)$  satisfies the equation

$$\frac{d\langle \tilde{r}(t) \rangle}{dt} = \alpha - \gamma_t \langle \tilde{r}(t) \rangle. \quad (65)$$

Since the initial value of  $\tilde{l}(t)$  is zero, we got

$$\langle \tilde{l}(t) \rangle = a(t) = \frac{\alpha}{\gamma_t} (1 - e^{-\gamma_t t}). \quad (66)$$

In CSP-Splicing, the marginal PMFs of  $\tilde{u}_l(t)$  and  $\tilde{s}_l(t)$  are

$$\begin{aligned} \tilde{P}_{m,\cdot}(t) &:= \text{Prob}(\tilde{u}_l(t) = m) = \sum_{n=0}^{\infty} \tilde{P}_{m,n}(t), \\ \tilde{P}_{\cdot,n}(t) &:= \text{Prob}(\tilde{s}_l(t) = n) = \sum_{m=0}^{\infty} \tilde{P}_{m,n}(t), \end{aligned} \quad (67)$$

respectively. The mean values of  $\tilde{u}_l(t)$  and  $\tilde{s}_l(t)$  have the form  $\langle \tilde{u}_l(t) \rangle = \sum_{m=1}^{\infty} m \tilde{P}_{m,\cdot}(t)$  and  $\langle \tilde{s}_l(t) \rangle = \sum_{n=1}^{\infty} n \tilde{P}_{\cdot,n}(t)$ . From the CME (10), we can obtain

$$\begin{aligned} \frac{d\langle \tilde{u}_l(t) \rangle}{dt} &= \sum_{m=1}^{\infty} m \partial_t \tilde{P}_{m,\cdot}(t) = \sum_{m=1}^{\infty} m \sum_{n=0}^{\infty} \partial_t \tilde{P}_{m,n}(t) \\ &= \sum_{m=1}^{\infty} m \sum_{n=0}^{\infty} \alpha (\tilde{P}_{m-1,n} - \tilde{P}_{mn}) + \beta ((m+1) \tilde{P}_{m+1,n-1} - m \tilde{P}_{mn}) \\ &\quad + \gamma_s ((n+1) \tilde{P}_{m,n+1} - n \tilde{P}_{mn}) \\ &= \alpha - \beta \langle \tilde{u}_l(t) \rangle, \end{aligned} \quad (68)$$

and

$$\begin{aligned} \frac{d\langle \tilde{s}_l(t) \rangle}{dt} &= \sum_{n=1}^{\infty} n \partial_t \tilde{P}_{\cdot,n}(t) = \sum_{n=1}^{\infty} n \sum_{m=0}^{\infty} \partial_t \tilde{P}_{m,n}(t) \\ &= \sum_{n=1}^{\infty} n \sum_{m=0}^{\infty} \alpha (\tilde{P}_{m-1,n} - \tilde{P}_{mn}) + \beta ((m+1) \tilde{P}_{m+1,n-1} - m \tilde{P}_{mn}) \\ &\quad + \gamma_s ((n+1) \tilde{P}_{m,n+1} - n \tilde{P}_{mn}) \\ &= \beta \langle \tilde{u}_l(t) \rangle - \gamma_s \langle \tilde{s}_l(t) \rangle. \end{aligned} \quad (69)$$

Similarly, we can derive the equations for the mean values of total unspliced and spliced mRNA ( $\tilde{u}(t), \tilde{s}(t)$ ):

$$\begin{aligned} \frac{d\langle \tilde{u}(t) \rangle}{dt} &= \alpha - \beta \langle \tilde{u}(t) \rangle, \\ \frac{d\langle \tilde{s}(t) \rangle}{dt} &= \beta \langle \tilde{u}(t) \rangle - \gamma_s \langle \tilde{s}(t) \rangle. \end{aligned} \quad (70)$$

Since the initial value of  $(\tilde{u}_l(t), \tilde{s}_l(t))$  is  $(0, 0)$ , we got

$$\langle \tilde{u}_l(t) \rangle = b(t) = \frac{\alpha}{\beta} (1 - e^{-\beta t}) \quad (71)$$

and

$$\langle \tilde{s}_l(t) \rangle = c(t) = \begin{cases} \frac{\alpha}{\gamma_s} (1 - e^{-\gamma_s t}) + \frac{\alpha}{\gamma_s - \beta} (e^{-\gamma_s t} - e^{-\beta t}), & \beta \neq \gamma_s, \\ \frac{\alpha}{\beta} (1 - e^{-\beta t}) - \alpha t e^{-\beta t}, & \beta = \gamma_s. \end{cases} \quad (72)$$

Similar to CSP-Baseline, in CSP-Switching,  $d\langle \tilde{l}(t) \rangle / dt$  and  $d\langle \tilde{r}(t) \rangle / dt$  satisfy the equations

$$\begin{aligned} \frac{d\langle \tilde{l}(t) \rangle}{dt} &= (1 - p_{\text{off}})\alpha - \gamma_t \langle \tilde{l}(t) \rangle, \\ \frac{d\langle \tilde{r}(t) \rangle}{dt} &= (1 - p_{\text{off}})\alpha - \gamma_t \langle \tilde{r}(t) \rangle. \end{aligned} \quad (73)$$

Since the initial value of  $\langle \tilde{l}(t) \rangle$  is zero, we got

$$\langle \tilde{l}(t) \rangle = \frac{(1 - p_{\text{off}})\alpha}{\gamma_t} (1 - e^{-\gamma_t t}). \quad (74)$$

## Computation of RNA velocity

To ease the notation, we denoted the new mRNA after data preprocessing by  $\bar{l}(t)$ , defined as

$$\bar{l}_j(t_k) = \frac{1}{|\mathcal{N}_{j,t_k}|} \sum_{i \in \mathcal{N}_{j,t_k}} \frac{l_i(t_k)}{p_i(t_k)},$$

which is different from the true expression  $\tilde{l}(t)$ , the discrete counts data  $l(t)$ , and the notation  $\hat{l}(t)$  in the post-processing subsection. We would also use the notation  $\bar{u}(t)$ ,  $\bar{s}(t)$  and  $\bar{r}(t)$  with similar definition.

In CSP-Baseline, only the total RNA velocity can be obtained due to the lack of the splicing stage. From Eq (65), we have

$$v_{\text{total}, r_j(t_k)} = \alpha_j(t_k) - \gamma_t \bar{r}_j(t_k), \quad (75)$$

where  $\bar{r}_j(t_k)$  is the number of total mRNA molecules of the  $j$ th cell labeled with length  $t_k$  after data preprocessing.

In CSP-Splicing, we add the two equations in Eq (70) to obtain

$$\frac{d\langle \tilde{r}(t) \rangle}{dt} = \frac{d\langle \tilde{u}(t) \rangle}{dt} + \frac{d\langle \tilde{s}(t) \rangle}{dt} = \alpha - \gamma_s \langle \tilde{s}(t) \rangle, \quad (76)$$

and thus get the equation for total RNA velocity

$$v_{\text{total}, r_j(t_k)} = \alpha_j(t_k) - \gamma_s \bar{s}_j(t_k). \quad (77)$$

In addition, in CSP-Splicing, we can also calculate the spliced RNA velocity by the following equation

$$v_{\text{spliced}, s_j(t_k)} = \beta_j(t_k) \bar{u}_j(t_k) - \gamma_s \bar{s}_j(t_k). \quad (78)$$

Similar to CSP-Baseline, the total RNA velocity in CSP-Switching can be obtained by the equation

$$\nu_{\text{total},r_j(t_k)} = (1 - p_{\text{off},j}(t_k))\alpha_j(t_k) - \gamma_t \bar{r}_j(t_k). \quad (79)$$

### Relationship between $\gamma_t$ and $\gamma_s$ and its implications

The difference between Eqs (65) and (76) implies the difference between the total mRNA degradation rate  $\gamma_t$  and spliced mRNA degradation rate  $\gamma_s$ . After suitable manipulations, we had the relation between  $\gamma_t$  and  $\gamma_s$  as below

$$\frac{\gamma_s}{\gamma_t} = \frac{\langle \tilde{r}(t) \rangle}{\langle \tilde{s}(t) \rangle}. \quad (80)$$

Therefore, we naturally got a method to infer  $\gamma_t$  when  $\gamma_s$  is known. Specifically, we first performed a zero-intercept linear regression

$$\bar{r}_j(t_k) = k \bar{s}_j(t_k) \quad (81)$$

to get the slope  $k$ . Then we computed  $\gamma_t$  by  $\gamma_t = \gamma_s/k$ . Therefore, we can also infer  $\gamma_t$  and compute the total RNA velocity by Eq (75) in CSP-Splicing.

We would also like to point out that CSP-Baseline and CSP-Splicing are incompatible upon assuming that  $\gamma_t$  and  $\gamma_s$  are both constants. These two assumptions usually do not hold simultaneously. Otherwise, from Eq (80) we knew that  $\langle \tilde{s}(t) \rangle / \langle \tilde{r}(t) \rangle$  is a constant, which is equivalent to that  $\langle \tilde{u}(t) \rangle / \langle \tilde{r}(t) \rangle$  is a constant, i.e.,  $\gamma_t(1 - e^{-\beta t}) / (\beta(1 - e^{-\gamma_t t}))$  is a constant. But this is only true when  $\beta$  and  $\gamma_t$  are equal.

### Robust analysis of the parameter inference in the CSP-Baseline

When  $\gamma_t t$  is small,  $1 - e^{-\gamma_t t} \sim \gamma_t t$  holds, then

$$l(t) = \frac{\alpha(1 - e^{-\gamma_t t})}{\gamma_t} \approx \alpha t, \quad (82)$$

which implies that from the mean perspective the nonlinear fitting of  $\alpha$  and  $\gamma_t$  degenerated into a linear fitting of  $\alpha$  at this point. For a more precise analysis, let  $\partial a(t)/\partial \alpha = (1 - e^{-\gamma_t t})/\gamma_t$ , we have  $\partial \ell(\alpha, \gamma_t)/\partial \alpha = 0$  is equivalent to

$$\alpha(\gamma_t) = \frac{\sum_{k=1}^K \sum_{j=1}^{n_k} l_j(t_k)}{\sum_{k=1}^K \sum_{j=1}^{n_k} p_j(t_k) \partial a(t_k)/\partial \alpha}. \quad (83)$$

But when  $1 - e^{-\gamma_t t} \sim \gamma_t t$  holds,  $\partial a(t)/\partial \alpha \approx t$ , then we have

$$\alpha \approx \frac{\sum_{k=1}^K \sum_{j=1}^{n_k} l_j(t_k)}{\sum_{k=1}^K \sum_{j=1}^{n_k} p_j(t_k) t_k} \quad (84)$$

is a constant, which we denoted by  $\alpha_{\text{cons}}$ . In addition, to quantitatively measure the robustness of inference on  $\gamma_t$ , since the optimal parameter is always located where the gradient is zero, we defined the  $l_1$ -norm of the derivative of the loss function with respect to  $\gamma_t$  restricted to  $\partial \ell / \partial \alpha$

= 0 (i.e. black line),

$$\left\| \frac{d\ell}{d\gamma_t} \Big|_{\frac{\partial \ell}{\partial x}=0}(\gamma_t) \right\|_{l_1} = \int_0^{\gamma_{t,\max}} \left| \frac{d\ell}{d\gamma_t} \right| d\gamma_t, \quad (85)$$

as a measure of robustness. Since the half-life of the total mRNA molecules is usually not less than half an hour, we took  $\gamma_{t,\max} = 1.5$ .

### Definition of correctness and consistency of velocity

The correctness of cell velocities is defined as follows: Consider the cell  $i$  with position  $x_i$  and velocity  $v_i$ . Define its one-step extrapolated position as  $x_i + v_i$ . We say that  $v_i$  is correct (correctness index = 1) if the cell  $j$  closest to the extrapolated position  $x_i + v_i$  ranks after  $i$  in the temporal ordering. Otherwise the correctness does not hold and we set the correctness index to be 0. Thus the average correctness refers to the percentage of correct velocities. Because the boundaries of the cell cycle being estimated are not clear and sharp, we did not use the RNA velocity benchmark metric cross-boundary direction (CBDir) proposed by Qiao et al. [12] and widely used for comparison of RNA velocity methods.

The consistency means the extent to which the velocity of one cell is consistent with the velocities of its neighboring cells, and we use the average correlation coefficient proposed in scVelo [2] to measure this consistency.

### Calculation of the duration of each cell cycle phase

After the total RNA velocities are obtained, we can evaluate the time of each phase of a cell cycle based on them. Specifically, we first pick  $k$  cells  $x_i^0$  ( $i = 1, 2, \dots, k$ ) whose relative positions are closest to 0 as a cell group, calculate their average expression  $\bar{x}^0$  and velocity  $\bar{v}^0$  as the initial expression  $x^0$  and velocity  $v^0$ , and extrapolate the state of the cell group with a short time step  $dt$ , that is,  $x^1 = x^0 + v^0 dt$ . We then search for another  $k$  cells  $x_i^1$  ( $i = 1, 2, \dots, k$ ) which are closest to the extrapolated state  $x^1$ , set the majority of the phase of these  $k$  cells to the phase of  $x^1$ , and set their average velocity  $\bar{v}^1$  as  $v^1$  for the second cell group. Next, the extrapolation and local  $k$ -cells group identification step can be repeated until a given threshold of the relative position is exceeded. In the actual calculation, we set  $k = 300$ ,  $dt = 0.01$ , and the threshold of the relative position to be 88% quantile of all relative positions. The above approach for processing the cell groups instead of cells themselves is to reduce the data noise by local averaging.

### Generation of simulation data

In this paper we generate two simulation datasets, one of which is bifurcated one-shot dataset following VeloSim's [31] flow, and the other is a non-steady-state kinetics dataset following scVelo's [2] model.

The bifurcated one-shot dataset is generated as described below. First we set the maximum observation time to  $T$  and the labeling time  $t_l$ . The observation times for half of the cells were generated randomly with a uniform distribution  $[t_l, T/2]$ , and the other half of the cells were divided into two equal parts, which were generated randomly with a uniform distribution  $[T/2, T]$ . We then followed SymSim [30] and VeloSim [31] to generate cell extrinsic variability factor(EVF)s and gene effect vector and used them to generate the cell-gene-wise transcription rate  $\alpha$ . Splicing rate  $\beta$  and degradation rate  $\gamma_s$  are gene-wise and were generated from uniform distributions of  $[0, 0.5]$  and  $[0, 5]$ , respectively. After all kinetic parameters were generated, we used the Gillespie algorithm to generate raw uu, ul, su and sl RNA counts data.

We did not consider technical noise when generating the simulated data and therefore set the size factor of all cells to 1.

The non-steady-state kinetics dataset is based on the model in scVelo [2], but ignores the splicing process. First we set the maximum observation time  $T$  and the  $K$  labeling times  $t_1, t_2 \dots t_K$ . The observation times for cells are randomly generated with a uniform distribution  $[t_K, T]$ , and the labeling times of the cells were randomly selected with equal probability from  $t_1$  to  $t_K$ . Transcription rate  $\alpha$  and degradation rate  $\gamma_t$ , both of which are gene-wise, were randomly generated with a uniform distribution  $[0.5, 1]$  and  $[0, 0.5]$ , respectively. To generate non-steady-state data, we set the switching time  $t_s = 0.5\rho/\gamma_t$ , where  $\rho$  is a random number generated with a uniform distribution  $[0, 1]$ . Finally similarly we use Gillespie algorithm to generate raw new and total RNA counts data and do not consider technical noise.

## Supporting information

**S1 Fig. Stochastic model combined with steady-state assumptions for one-shot experiments, realated to Fig 2.** Storm in this figure refers to the inference strategy of CSP-Baseline model combined with the steady state assumption. **A.** Cell quiver plot in the PCA space of the scSLAM-seq dataset [17]. **B.** Degradation rates  $\gamma_t$  estimated with steady-based method in Storm compared to that of the Dynamo method in the scSLAM-seq dataset [17]. **C.** Same as **B**, but for the datasets from the sci-fate [19]. **D.** Same as **B**, but for the datasets from the PerturbSci-Kinetics [22].

(PNG)

**S2 Fig. Storm analyzes one-shot data with both splicing and labeling without steady-state assumption, realated to Fig 3.** **A.** Streamline projected in the PCA space plots of one-shot bifurcation simulation data of cellDancer. **B.** Streamline plot in the UMAP space of the murine intestinal organoid system dataset from scEU-seq [21] of cellDancer. **C.** Heat map of absolute error between estimated and true gene-cell-wise transcription rates  $\alpha$  of one-shot bifurcation simulation data of Dynamo.

(PNG)

**S3 Fig. GO (gene ontology) pathway enrichment results of genes with high  $\alpha$  and  $p_{off}$  (top 50%) and low  $\gamma_t, \beta, \alpha$  and  $p_{off}$  (bottom 50%) in well-fit genes (top 40% of goodness of fit), related to Fig 5 in main text.**

(PNG)

**S4 Fig. RNA velocity analysis of the cell cycle dataset, related to Fig 6.** The inference strategy involved in this figure is for kinetics/pulse data. **A.** Comparison of spliced RNA velocity streamline visualizations between CSP-Splicing method and Dynamo. **B.** Comparison of the average correctness of spliced velocity in gene expression space RFP\_GFP space. The p-values are given by the one-sided Wilcoxon test. **C.** Similar to **B**, but for velocity consistency. **D.** Total RNA velocity streamlines calculated using gene-wise parameters (instead of using gene-cell-wise parameters except for the degradation rate). **Left:** ICSP. **Right:** CSZIP **E.** Comparison of total RNA velocity in DCBLD2 between CSP-Splicing and CSP-Switching. **F.** Phase portraits of new-total RNA planes of DCBLD2 of CSP-Splicing and CSP-Switching. Quivers correspond to the total (x-component) or new (y-component) RNA velocity calculated by the different methods. **G.** Similar to **E**, but for gene HIPK2 of three stochastic methods and Dynamo. **H.** Similar to **F**, but for gene HIPK2 of three stochastic methods. **I.** The smoothed expression pattern of HIPK2 across cells.

(PNG)

**S5 Fig. RNA velocity analysis of the simulated pulse dataset, related to Fig 6.** Storm in this figure refers to the inference strategy of CSP-Baseline model for pulse data. **A.** Comparison of total RNA velocity streamline visualizations between Storm and Dynamo in simulated pulse dataset. **B.** Comparison of the estimated degradation rate with the true degradation rate in simulated pulse data. **C.** Distribution plot of the difference between the estimated degradation rate and the true value, including Storm, well-fitted genes in Storm, Dynamo and well-fitted genes in Dynamo.  
(PNG)

**S6 Fig. PAGA analysis of different datasets and different methods.** **A.** Comparison of PAGA velocity graph on the neuronal activity under KCl polarization datasets from scNT-seq. Left: Storm; Right: Dynamo. **B.** Comparison of PAGA velocity graph on the cellcycle dataset from scEU-seq. From left to right, from top to bottom, Storm's CSP-Baseline stochastic model without steady-state assumption, CSP-Baseline stochastic model with steady-state assumption, Dynamo's deterministic model with steady-state assumption and random velocity. Type annotations were derived from an equal number division of cells into 8 classes based on the relative positions of cells provided by the original scEU-seq study, and the first and last classes were combined into 1 class.  
(PNG)

## Acknowledgments

We thank Prof. Fang Yao for helpful discussions.

## Author Contributions

**Conceptualization:** Xiaojie Qiu, Tiejun Li.

**Data curation:** Qiangwei Peng, Xiaojie Qiu, Tiejun Li.

**Investigation:** Qiangwei Peng, Xiaojie Qiu, Tiejun Li.

**Methodology:** Qiangwei Peng, Xiaojie Qiu, Tiejun Li.

**Project administration:** Xiaojie Qiu, Tiejun Li.

**Resources:** Xiaojie Qiu, Tiejun Li.

**Software:** Qiangwei Peng, Xiaojie Qiu.

**Supervision:** Xiaojie Qiu, Tiejun Li.

**Validation:** Qiangwei Peng, Xiaojie Qiu, Tiejun Li.

**Visualization:** Qiangwei Peng, Xiaojie Qiu, Tiejun Li.

**Writing – original draft:** Qiangwei Peng, Xiaojie Qiu, Tiejun Li.

**Writing – review & editing:** Qiangwei Peng, Xiaojie Qiu, Tiejun Li.

## References

1. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature*. 2018; 560(7719):494–498. <https://doi.org/10.1038/s41586-018-0414-6> PMID: 30089906
2. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature biotechnology*. 2020; 38(12):1408–1414. <https://doi.org/10.1038/s41587-020-0591-3> PMID: 32747759
3. Li T, Shi J, Wu Y, Zhou P. On the Mathematics of RNA Velocity I: Theoretical Analysis. *CSIAM Transactions on Applied Mathematics*. 2021; 2(1):1–55. <https://doi.org/10.4208/csiam-am.SO-2020-0001>

4. Li T, Wang Y, Guoguo Y, Zhou P. On the mathematics of RNA velocity II: algorithmic aspects. bioRxiv. 2023; p. 2023–06.
5. Jahnke T, Huiszinga W. Solving the chemical master equation for monomolecular reaction systems analytically. Journal of mathematical biology. 2007; 54:1–26. <https://doi.org/10.1007/s00285-006-0034-x> PMID: 16953443
6. Gorin G, Fang M, Chari T, Pachter L. RNA velocity unraveled. PLoS Computational Biology. 2022; 18(9):e1010492. <https://doi.org/10.1371/journal.pcbi.1010492> PMID: 36094956
7. Li C, Virgilio MC, Collins KL, Welch JD. Multi-omic single-cell velocity models epigenome–transcriptome interactions and improves cell fate prediction. Nature Biotechnology. 2022; p. 1–12. <https://doi.org/10.1038/s41587-022-01476-y> PMID: 36229609
8. Gorin G, Svensson V, Pachter L. Protein velocity and acceleration from single-cell multiomics experiments. Genome biology. 2020; 21(1):1–6. <https://doi.org/10.1186/s13059-020-1945-3> PMID: 32070398
9. Gao M, Qiao C, Huang Y. UniTVelo: temporally unified RNA velocity reinforces single-cell trajectory inference. Nature Communications. 2022; 13(1):6586. <https://doi.org/10.1038/s41467-022-34188-7> PMID: 36329018
10. Cui H, Maan H, Wang B. DeepVelo: Deep Learning extends RNA velocity to multi-lineage systems with cell-specific kinetics. bioRxiv. 2022;.
11. Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. RNA velocity—current challenges and future perspectives. Molecular systems biology. 2021; 17(8):e10282. <https://doi.org/10.1525/msb.202110282> PMID: 34435732
12. Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. Representation learning of RNA velocity reveals robust cell transitions. Proceedings of the National Academy of Sciences. 2021; 118(49):e2105859118. <https://doi.org/10.1073/pnas.2105859118>
13. Gayoso A, Weiler P, Lotfollahi M, Klein D, Hong J, Streets AM, et al. Deep generative modeling of transcriptional dynamics for RNA velocity analysis in single cells. bioRxiv. 2022;.
14. Gu Y, Blaauw D, Welch JD. Bayesian inference of rna velocity from multi-lineage single-cell data. bioRxiv. 2022; p. 2022–07.
15. Farrell S, Mani M, Goyal S. Inferring Single-Cell Transcriptomic Dynamics with Structured Latent Gene Expression Dynamics. Available at SSRN 4330809;.
16. Li S, Zhang P, Chen W, Ye L, Brannan KW, Le NT, et al. A relay velocity model infers cell-dependent RNA velocity. Nature Biotechnology. 2023; p. 1–10. <https://doi.org/10.1038/s41587-023-01728-5> PMID: 37012448
17. Erhard F, Baptista MA, Krammer T, Hennig T, Lange M, Arampatzis P, et al. scSLAM-seq reveals core features of transcription dynamics in single cells. Nature. 2019; 571(7765):419–423. <https://doi.org/10.1038/s41586-019-1369-y> PMID: 31292545
18. Qiu Q, Hu P, Qiu X, Govek KW, Cámera PG, Wu H. Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. Nature methods. 2020; 17(10):991–1001. <https://doi.org/10.1038/s41592-020-0935-4> PMID: 32868927
19. Cao J, Zhou W, Steemers F, Trapnell C, Shendure J. Sci-fate characterizes the dynamics of gene expression in single cells. Nature biotechnology. 2020; 38(8):980–988. <https://doi.org/10.1038/s41587-020-0480-9> PMID: 32284584
20. Hendriks GJ, Jung LA, Larsson AJ, Lidschreiber M, Andersson Forsman O, Lidschreiber K, et al. NASC-seq monitors RNA synthesis in single cells. Nature communications. 2019; 10(1):1–9. <https://doi.org/10.1038/s41467-019-11028-9> PMID: 31316066
21. Battich N, Beumer J, de Barbanson B, Krenning L, Baron CS, Tanenbaum ME, et al. Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. Science. 2020; 367(6482):1151–1156. <https://doi.org/10.1126/science.aax3072> PMID: 32139547
22. Xu Z, Sziraki A, Lee J, Zhou W, Cao J. Dissecting key regulators of transcriptome kinetics through scalable single-cell RNA profiling of pooled CRISPR screens. Nature Biotechnology. 2023; p. 1–6.
23. Holler K, Neuschulz A, Drewe-Böß P, Mintcheva J, Spanjaard B, Arsiè R, et al. Spatio-temporal mRNA tracking in the early zebrafish embryo. Nature Communications. 2021; 12(1):3358. <https://doi.org/10.1038/s41467-021-23834-1> PMID: 34099733
24. Lin S, Yin K, Zhang Y, Lin F, Chen X, Zeng X, et al. Well-TEMP-seq as a microwell-based strategy for massively parallel profiling of single-cell temporal RNA dynamics. Nature Communications. 2023; 14(1):1272. <https://doi.org/10.1038/s41467-023-36902-5> PMID: 36882403
25. Ren J, Zhou H, Zeng H, Wang CK, Huang J, Qiu X, et al. Spatiotemporally resolved transcriptomics reveals the subcellular RNA kinetic landscape. Nature Methods. 2023; p. 1–11. <https://doi.org/10.1038/s41592-023-01829-8> PMID: 37038000

26. Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, et al. Mapping transcriptomic vector fields of single cells. *Cell*. 2022; 185(4):690–711. <https://doi.org/10.1016/j.cell.2021.12.045> PMID: 35108499
27. Weiler P, Lange M, Klein M, Pe'er D, Theis F. Unified fate mapping in multiview single-cell data. *bioRxiv*. 2023; p. 2023–07.
28. Koehler KJ, Larntz K. An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal of the American Statistical Association*. 1980; 75(370):336–344. <https://doi.org/10.1080/01621459.1980.10477473>
29. Menard S. Coefficients of determination for multiple logistic regression analysis. *The American Statistician*. 2000; 54(1):17–24. <https://doi.org/10.1080/00031305.2000.10474502>
30. Zhang X, Xu C, Yosef N. Simulating multiple faceted variability in single cell RNA sequencing. *Nature communications*. 2019; 10(1):2611. <https://doi.org/10.1038/s41467-019-10500-w> PMID: 31197158
31. Zhang Z, Zhang X. VeloSim: Simulating single cell gene-expression and RNA velocity. *BioRxiv*. 2021; p. 2021–01.
32. Chong S, Chen C, Ge H, Xie XS. Mechanism of transcriptional bursting in bacteria. *Cell*. 2014; 158(2):314–326. <https://doi.org/10.1016/j.cell.2014.05.038> PMID: 25036631
33. Jürges C, Döhlen L, Lazarević D, Erhard F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics*. 2018; 34(13):i218–i226. <https://doi.org/10.1093/bioinformatics/bty256> PMID: 29949974
34. Grün D, Kester L, Van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nature methods*. 2014; 11(6):637–640. <https://doi.org/10.1038/nmeth.2930> PMID: 24747814
35. Pearson K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1900; 50(302):157–175. <https://doi.org/10.1080/14786440009463897>
36. Benhamou E, Melot V. Seven proofs of the Pearson Chi-squared independence test and its graphical interpretation. *arXiv preprint arXiv:180809171*. 2018;
37. Chao HX, Fakhreddin RI, Shimerov HK, Kedziora KM, Kumar RJ, Perez J, et al. Evidence that the human cell cycle is a series of uncoupled, memoryless phases. *Molecular systems biology*. 2019; 15(3):e8604. <https://doi.org/10.1525/msb.20188604> PMID: 30886052
38. Vallender S. Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*. 1974; 18(4):784–786. <https://doi.org/10.1137/1118101>
39. Zhang J, Zhong W, Ma P. A review on modern computational optimal transport methods with applications in biomedical research. *Modern Statistical Methods for Health Research*. 2021; p. 279–300.
40. Tedesco M, Giannese F, Lazarević D, Giansanti V, Rosano D, Monzani S, et al. Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. *Nature Biotechnology*. 2022; 40(2):235–244. <https://doi.org/10.1038/s41587-021-01031-1> PMID: 34635836
41. Chari T, Pachter L. The specious art of single-cell genomics. *PLoS Computational Biology*. 2023; 19(8):e1011288. <https://doi.org/10.1371/journal.pcbi.1011288> PMID: 37590228
42. Marot-Lassauzaie V, Bouman BJ, Donaghy FD, Demerdash Y, Essers MAG, et al. Towards reliable quantification of cell state velocities. *PLoS Computational Biology*. 2022; 18(9):e1010031. <https://doi.org/10.1371/journal.pcbi.1010031> PMID: 36170235
43. Zheng SC, Stein-O'Brien G, Boukas L, Goff LA, Hansen KD. Pumping the brakes on RNA velocity by understanding and interpreting RNA velocity estimates. *Genome biology*. 2023; 24(1):246. <https://doi.org/10.1186/s13059-023-03065-x> PMID: 37885016
44. Zhang Y, Qiu XJ, Weissman JS, Bahar I, Xing JH. Graph-Dynamo: Learning stochastic cellular state transition dynamics from single cell data. *BioRxiv*. 2023; p. 2023–09.
45. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin S, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology*. 2019; 20:1–9. <https://doi.org/10.1186/s13059-019-1663-x> PMID: 30890159
46. Riba A, Oravecz A, Durik M, Jiménez S, Alunni V, Cerciat M, et al. Cell cycle gene regulation dynamics revealed by RNA velocity and deep-learning. *Nature communications*. 2022; 13(1):2865. <https://doi.org/10.1038/s41467-022-30545-8> PMID: 35606383
47. Lederer AR, Leonardi M, Talamanca L, Herrera A, Droin C, Khven I, et al. Statistical inference with a manifold-constrained RNA velocity model uncovers cell cycle speed modulations. *BioRxiv*. 2024; p. 2024–01.