

Reviewed Preprint

v1 • December 13, 2024

Not revised

Reviewed Preprint

v2 • August 22, 2025

Revised by authors

Reviewed Preprint

v3 • January 16, 2026

Revised by authors

For correspondence:

jiangbenyuan@gdph.org.cn

qnie@uci.edu

zhijiajun@mail.sysu.edu.cn

† These authors contributed equally to this work.

Competing interests: No competing interests declared

Funding: See [page 26](#)

Reviewing editor: Mariana Gómez-Schiavon, Universidad Nacional Autónoma de México, Mexico

© 2024, Huang et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Deep learning linking mechanistic models to single-cell transcriptomics data reveals transcriptional bursting in response to DNA damage

Zhiwei Huang^{1,2,†}, Songhao Luo^{3,†}, Zihao Wang^{1,2}, Zhenquan Zhang^{1,2}, Benyuan Jiang⁴✉, Qing Nie³✉, Jiajun Zhang^{1,2}✉

¹Guangdong Province Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou, China •

²School of Mathematics, Sun Yat-sen University, Guangzhou, China • ³Department of Mathematics, University of California Irvine, Irvine, United States • ⁴Guangdong Lung Cancer Institute, Guangdong Provincial People's Hospital and Guangdong Academy of Medical Sciences, Guangzhou, China

eLife Assessment

This study presents DeepTX, a **valuable** methodological tool that integrates mechanistic stochastic models with single-cell RNA sequencing data to infer transcriptional burst kinetics at genome scale. The approach is broadly applicable and of interest to subfields such as systems biology, bioinformatics, and gene regulation. The evidence supporting the findings is **solid**, with appropriate validation on synthetic data and thoughtful discussion of limitations related to identifiability and model assumptions.

<https://doi.org/10.7554/eLife.100623.3.sa2>

Abstract

Cells must adopt flexible regulatory strategies to make decisions regarding their fate, including differentiation, apoptosis, or survival in the face of various external stimuli. One key cellular strategy that enables these functions is stochastic gene expression programs. However, understanding how transcriptional bursting, and consequently, cell fate, responds to DNA damage on a genome-wide scale poses a challenge. In this study, we propose an interpretable and scalable inference framework, DeepTX, that leverages deep learning methods to connect mechanistic models and scRNA-seq data, thereby revealing genome-wide transcriptional burst kinetics. This framework enables rapid and accurate solutions to transcription models and the inference of transcriptional burst kinetics from scRNA-seq data. Applying this framework to several scRNA-seq datasets of DNA-damaging drug treatments, we observed that fluctuations in transcriptional bursting induced by different drugs was associated with distinct fate decisions: IdU treatment was associated with differentiation in mouse embryonic stem cells by increasing the burst size of gene expression, while low- and high-dose 5FU treatments in human colon cancer cells were associated with changes in burst frequency that corresponded to apoptosis- and survival-related fate, respectively. Together, these results show that DeepTX enables genome-wide inference of transcriptional bursting from single-cell transcriptomics data and can generate hypotheses about how bursting dynamics relate to cell fate decisions.

Introduction

Cells must employ flexible regulatory strategies to determine their fate, including cell differentiation, survival, or apoptosis, in response to diverse external stimuli (López-Maury L et al., 2008 [↗](#); Stadhouders R et al., 2019 [↗](#)). One key cellular strategy that enables both beneficial and detrimental functions is the stochastic nature of gene expression programs, arising from intricate

biochemical processes and termed “noise” (Raj A & van Oudenaarden A, 2008 [↗](#); Balázsi G et al., 2011 [↗](#); Losick R & Desplan C, 2008 [↗](#)). Noise can arise at many steps—from stochastic transcription factor binding and chromatin remodeling, to variability in RNA transcription and protein translation. Among these diverse sources, transcriptional noise sits at the very start of the central dogma, converting upstream molecular fluctuations into the mRNA outputs that drive all downstream regulatory and functional processes. Transcriptional noise originates from a discontinuous pattern, wherein mRNA production occurs through alternating active and inactive gene states (Eling N et al., 2019 [↗](#); Tunnaccliffe E & Chubb JR, 2020 [↗](#)). This phenomenon, known as “transcriptional bursting”, has been observed in numerous experiments across both prokaryotic and eukaryotic cells (Golding I et al., 2005 [↗](#); Chubb JR et al., 2006 [↗](#); Larson DR et al., 2011 [↗](#)). The burst kinetics, characterized by burst frequency and burst size, describe the stochastic behavior of molecular interactions at the gene expression process, offering valuable insights into how cells encode and leverage variability. However, it remains an unresolved question how cells make fate decisions in response to external signals through genome-wide regulation of transcriptional burst mechanisms (Lammers NC et al., 2020 [↗](#); Rodriguez J & Larson DR, 2020 [↗](#)).

DNA damage is a crucial and prevalent factor contributing to cellular responses, inducing varying degrees of cytotoxicity and thereby influencing diverse cell fate decisions (Krenning L et al., 2019 [↗](#); Su TT, 2006 [↗](#); Hafner A et al., 2019 [↗](#)). For instance, low cytotoxicity limits cell proliferation (van den Berg J et al., 2019 [↗](#); Arora M et al., 2017 [↗](#); Barr AR et al., 2017 [↗](#); Deng Z et al., 2023 [↗](#)), whereas heightened cytotoxicity prompts cell differentiation and senescence (Müllers E et al., 2014 [↗](#); Feringa FM et al., 2018 [↗](#); Toledo LI et al., 2008 [↗](#); Zhao Y et al., 2023 [↗](#)). Furthermore, extreme scenarios of high cytotoxicity can cause cells to engage in the apoptotic process (Yousefzadeh M et al., 2021 [↗](#); Carneiro BA & El-Deiry WS, 2020 [↗](#); Roos WP & Kaina B, 2013 [↗](#); Zheng P et al., 2018 [↗](#)). Experiments have shown that DNA damage can disrupt the gene transcription process (Lans H et al., 2019 [↗](#)). Specifically, DNA damage slows down the movement of RNA Pol II along the DNA strand (Muñoz MJ et al., 2009 [↗](#)). When more severe DNA damage is encountered, RNA Pol II exhibits a sliding pause behavior for DNA repair before resuming transcription (Figure 1A [↗](#)) (Gregersen LH & Svejstrup JQ, 2018 [↗](#); Geijer ME & Marteijn JA, 2018 [↗](#); Giono LE et al., 2016 [↗](#)). These changes in RNA Pol II movement behavior would further reflect the cellular regulation of gene expression, particularly burst kinetics (Friedrich D et al., 2019 [↗](#); Calia GP et al., 2023 [↗](#)). A recent study has shown that cellular responses to DNA damage involve the regulation of burst frequency (BF) on specific genes, rather than a genome-wide conclusion (Friedrich D et al., 2019 [↗](#)). They have demonstrated that cells respond to DNA damage by increasing gene expression noise, but they do not delve into the topic of transcriptional bursts (Calia GP et al., 2023 [↗](#); Desai RV et al., 2021 [↗](#)). Therefore, the question of how DNA damage causes cells to regulate transcriptional burst kinetics on a genome-wide scale remains unresolved.

To understand the dynamics of transcriptional bursts, ideally, one needs to observe the fluctuations of gene expression over a continuous time interval, such as smFISH, scNT-seq, and HT-smFISH (Femino AM et al., 1998 [↗](#); Qiu Q et al., 2020 [↗](#); Safieddine A et al., 2023 [↗](#)). However, most existing techniques can only observe a limited number of genes, making their application to genome-wide studies challenging. Fortunately, single-cell RNA sequencing (scRNA-seq) technology provides excellent opportunities to explore this question, emerging as the leading method for genome-wide mRNA measurements and revealing gene expression noise within individual cells (Tanay A & Regev A, 2017 [↗](#); Picelli S et al., 2013 [↗](#); Zheng GX et al., 2017 [↗](#)). Many studies have employed scRNA-seq to elucidate the diverse sources of gene expression noise and the fundamental principles of transcriptional dynamics on a genome-wide scale (Eling N et al., 2019 [↗](#); Faure AJ et al., 2017 [↗](#); Morgan MD & Marioni JC, 2018 [↗](#); Ochiai H et al., 2020 [↗](#)). However, scRNA-seq data provide only static snapshots of cellular states. To infer the underlying transcriptional burst kinetics, one can apply mathematical models of stochastic gene expression, typically under the assumption that the observed distributions of scRNA-seq data reflect the steady-state outcome of the underlying dynamic processes (Larsson AJM et al., 2019 [↗](#); Luo S et al., 2023 [↗](#); Luo S et al., 2023 [↗](#)).

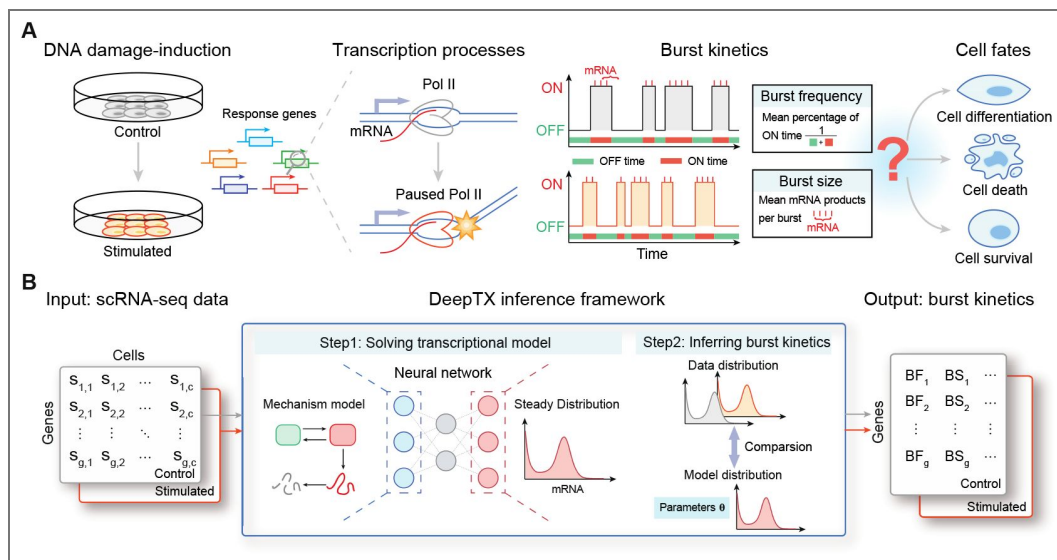


Figure 1. Overview of DeepTX framework.

(A) Cells stimulated by DNA damaging drugs cause damage to the DNA double strands, which will slow down or even stop the movement of RNA Pol II on the DNA double strands. Changes in the state of RNA Pol II movement will lead to changes in the kinetics of gene expression bursts (including burst frequency and burst size), which will affect cell fate decisions such as apoptosis, differentiation, and survival. **(B)** The input to the DeepTX framework is scRNA-seq data, and the output is the burst kinetics corresponding to each gene. The core of the DeepTX inference framework is a hierarchical model, which is a mixture of the stationary distribution of the mechanism model solved by the neural network (Step 1) and the binomial distribution followed by the sequencing process. This hierarchical model is used to infer the out dynamics parameters corresponding to the data distribution (Step 2).

For mathematical modeling, the gene expression model for DNA damage ought to encompass more sophisticated generalized properties compared to the traditional telegraph model, which is a common model describing gene expression burst kinetics by genes switching stochastically between active (ON) and inactive (OFF) states with single-step process (exponentially distributed waiting times) (Stumpf PS et al., 2017 [DOI](#); Voss TC & Hager GL, 2014 [DOI](#)). However, the presence of DNA damage necessitates modeling the transcriptional process as a multi-step process, rather than a single-step process, to capture the additional complexity introduced by the damage (Singh A et al., 2013 [DOI](#); Cavallaro M et al., 2021 [DOI](#)). Many efforts in this field are addressing the challenges of describing this multi-step process of gene expression in an interpretable and tractable model. One common modeling approach is to introduce multiple intermediate states between inactivate and activate states (Zhang J et al., 2012 [DOI](#); Zhang J & Zhou T, 2014 [DOI](#); Zhou T & Zhang J, 2012 [DOI](#)). Although this multi-state model can fit the experimental data better, it is hard to rationalize the state numbers and parameters (Desai RV et al., 2021 [DOI](#); Rodriguez J et al., 2019 [DOI](#); Zoller B et al., 2015 [DOI](#)). Another alternative modeling approach is direct non-Markovian modeling for non-exponential waiting times in gene state, which maps the multiple parameters from multi-state models to a small number of interpretable parameters that are easily observed experimentally (Daigle BJ, Jr. et al., 2015 [DOI](#); Kumar N et al., 2015 [DOI](#); Schwabe A et al., 2012 [DOI](#); Stinchcombe AR et al., 2012 [DOI](#); Zhang J & Zhou T, 2019 [DOI](#); Zhang JJ & Zhou TS, 2019 [DOI](#)). However, the ensuing difficulty is solving the analytic solution of the steady distribution from a non-Markovian model, which may require some numerical simulation methods that are computationally resource-intensive and time-consuming, such as Monte Carlo methods (Sisson SA et al., 2007 [DOI](#)).

Statistical inference, particularly in recovering burst kinetic parameters from genome-wide scRNA-seq data, necessitates efficient and scalable inference algorithms (Gómez-Schiavon M et al., 2017 [DOI](#)). Therefore, the stochastic dynamical system must be quickly solved as parameters are continuously updated throughout the inference process. Deep learning exhibits a broad array of applications as a contemporary method for addressing complex systems (Jiang Q et al., 2021 [DOI](#); Wang S et al., 2019 [DOI](#); Michoski C et al., 2020 [DOI](#)). For example, some studies employed neural networks to establish the mapping relationship between a wide range of parameters in the gene expression model and the corresponding pre-simulated steady distribution (Figure 2B [DOI](#)) (Jiang Q et al., 2021 [DOI](#); Wang S et al., 2019 [DOI](#); Davis CN et al., 2020 [DOI](#); Tang Y et al., 2023 [DOI](#)). This approach, based on neural networks, has been utilized for parameter inference in both deterministic and stochastic models (Jiang Q et al., 2021 [DOI](#); Gaskin T et al., 2023 [DOI](#); Sukys A et al., 2022 [DOI](#); Tang W et al., 2023 [DOI](#)).

In this study, we introduce DeepTX, a deep learning inference framework that integrates mechanistic models and deep learning methods to elucidate the genome-wide regulation of transcriptional bursts in DNA damage response using scRNA-seq data. The DeepTX framework comprises two modules: DeepTXsolver and DeepTXinferer. Specifically, DeepTX initially employs the DeepTXsolver module to efficiently solve complex gene expression dynamics models using neural network architectures. It subsequently utilizes the DeepTXinferer module to accurately infer potential transcriptional burst kinetic parameters using Bayesian methods. DeepTX demonstrates good performance on synthetic datasets. Furthermore, to investigate genome-wide regulation of transcriptional bursts in response to DNA damage, we applied the DeepTX framework to three DNA damage-related scRNA-seq datasets representing cell differentiation, apoptosis, and cell survival. As a result, we observed that the fate decision of mouse embryonic stem cells (mESCs) to undergo cell differentiation in response to DNA damage caused by 5'-iodo-2'-deoxyuridine (IdU) compounds was associated with enhanced burst size (BS enhancer) of genes linked to delayed mitosis phase transition, which may reflect regulatory programs related to reprogramming and differentiation. In colorectal cancer cells, low-dose 5-fluorouracil (5FU) treatment was associated with increased burst frequency (BF enhancer) of genes connected to intracellular oxidative stress, coinciding with apoptosis-related processes. By contrast, high-dose 5FU treatment was associated with BF enhancer genes enriched in telomerase-related pathways, which may mitigate oxidative stress and correlate with cellular resistance. In conclusion, DeepTX is a computational framework with utility and extensibility to infer transcriptional dynamics from scRNA-seq data.

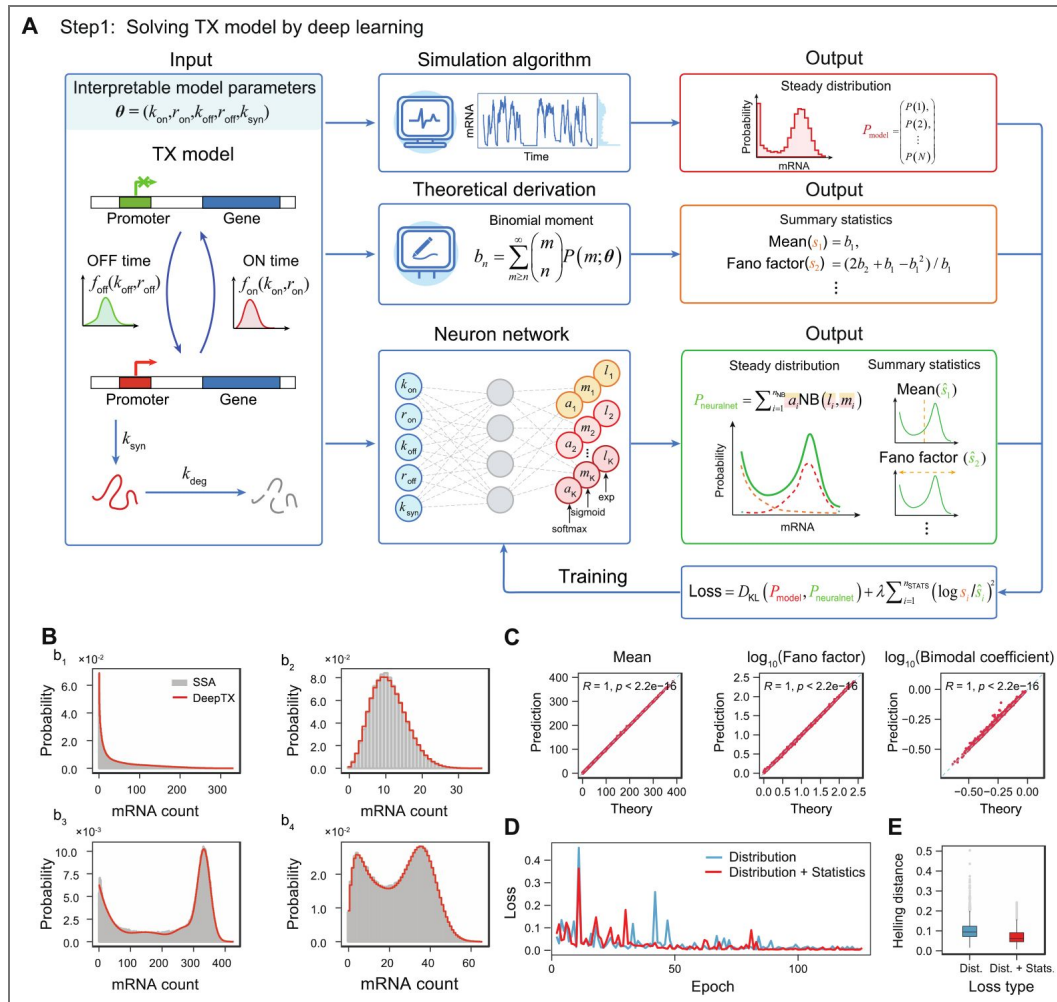


Figure 2. Solving TXmodel using deep learning.

(A) With the TXmodel and corresponding model parameters θ , we obtain a steady distribution using the SSA simulation and the moment statistics employing binomial moment theory. These results are compared with the distribution and moment statistics of the neural network's output to calculate the neural network's loss value. The parameters of the neural network are optimized using gradient descent until convergence is achieved, indicating attainment of optimal parameters. (B) Comparison between SSA results and DeepTX prediction results for four sets of test parameters. The gray bars and the red stepped line correspond to the distributions obtained from SSA simulations and DeepTX predictions, respectively. (C) Verification results of the moment statistics predicted by DeepTX and the true moments on a test set containing 1000 elements. (D) The loss curve during DeepTX training. The blue curve represents that the loss function is composed of KL divergence of distributions, and the red curve represents that the loss function is composed of KL divergence of distributions and statistics. (E) Box plots of the Hellinger distance between the true distribution and the predicted distribution by DeepTX of different loss types on the test set.

Results

Overview of DeepTX framework

To understand how cell influences underlying transcriptional mechanisms in response to DNA damage, as described in the introduction section (Figure 1A [↗](#)), we propose an effective computational inference framework, referred to as DeepTX (Figure 1B [↗](#)). DeepTX takes as input a set of scRNA-seq datasets and then outputs genome-wide burst kinetics parameters with the benefit from deep learning methods (Figure 1B [↗](#)). The inference process of the DeepTX framework is composed of two crucial modules, DeepTXsolver and DeepTXinferer. The first module, DeepTXsolver, is to solve for the stationary distribution of a gene expression dynamic model (Figure 1B middle [↗](#)). A neural network is constructed in DeepTXsolver to approximate the mapping from model parameters to the corresponding stationary distribution. As in many biological contexts, there is a need to model dynamics that describe more realistic gene expression processes that are difficult to solve. For instance, gene expression under DNA damage is better represented by multi-step transcriptional models, which more faithfully reflect the underlying regulatory complexity. Moreover, sequencing noise should be integrated into the modeling framework (see Methods). The second module, DeepTXinferer, infers the burst kinetics parameter (Figure 1B middle [↗](#)). Using a trained neural network from DeepTXsolver, we can easily and quickly obtain the stationary distribution (called model distribution) of any parameter within a reasonable parameter space. DeepTXinferer compares the model distribution to the stationary distribution of each gene in the scRNA-seq data (called data distribution) as the parameters are iterated until convergence and utilize the Bayesian inference to output the posterior distribution of burst kinetic parameter (Figure 1B right [↗](#)). We will describe these two modules in detail in the following two sections.

DeepTX solves the transcriptional model using deep learning

To enable the recovery of kinetic information from static scRNA-seq data, we first performed rational mechanistic modeling of gene expression processes with DNA damage. The gene transcription burst process is often characterized using the traditional two-state model (Singh A et al., 2013 [↗](#); Cavallaro M et al., 2021 [↗](#)), which assumes that genes switch randomly between OFF and ON states with exponentially distributed waiting times for each state. However, the gene expression process is inherently a multi-step process, which particularly cannot be neglected under conditions of DNA damage. DNA damage can result in slowing or even stopping the RNA pol II movement (Lans H et al., 2019 [↗](#)) and cause many macromolecules to be recruited for damage repair. This process will affect the spatially localized behavior of the promoter (Lans H et al., 2019 [↗](#)), causing the dwell time of promoter inactivation and activation that cannot be approximated by a simple two state. We therefore used a generalized model we developed previously (Luo S et al., 2023 [↗](#)), called TXmodel here, which extended the state waiting time to an arbitrary distribution, i.e., the gene expression system is a non-Markovian system (Figure 2A [↗](#) and see Methods 1.2). More specifically, the waiting times for the transitions between the OFF and ON states are represented by two random variables, assumed to follow the arbitrary distributions $f_{\text{off}}(k_{\text{off}}, r_{\text{off}})$ and $f_{\text{on}}(k_{\text{on}}, r_{\text{on}})$, respectively. Additionally, we assumed that the rates of mRNA synthesis and degradation are constants, i.e., the waiting times for mRNA transcription and degradation follow exponential distributions with rate parameters k_{syn} and k_{deg} , respectively (Figure 2A left [↗](#)). Nevertheless, this non-Markovian gene expression system is difficult to solve stationary distribution analytically, while numerical solution methods are time-consuming.

For that reason, we utilized a deep learning approach, which has been demonstrated to be effective in solving stochastic systems (Gupta A et al., 2021 [↗](#)), to construct a mapping from the parameter of the mechanistic TXmodel to its corresponding stationary distribution. More specifically, we aimed to train a fully connected neural network referred as DeepTXsolver, whose inputs are the parameters of the TXmodel θ and whose output is a stationary distribution parameterized by a mixed negative binomial distribution $P_{\text{neuralnet}}$, which has good performance in approximating the solution of the chemical master equation (Perez-Carrasco R et al., 2020 [↗](#);

Öcal K et al., 2022 [DOI](#)) (Figure 2A [DOI](#), Supplementary Figure S1 and see Methods 1.3). To generate effective training sets for DeepTXsolver, we first generated a large number of parameter sets within a reasonable range of parameter space by using Sobol sampling (see Methods 1.3) (Sobol IM, 1967 [DOI](#)). Subsequently, for each parameter set, we employed a modified stochastic simulation algorithm (SSA) for non-Markovian TXmodels to generate numerous samples (Figure 2A [DOI](#) and see Methods 1.3). The resulting stationary distributions $P_{\text{simulation}}$ served as training labels. Additionally, we demonstrated that the arbitrary order binomial moments of the TXmodel can be analytically solved iteratively within the context of queueing theory (Zhang Z et al., 2021 [DOI](#); Zhang J et al., 2024 [DOI](#)). This enables the straightforward computation of significant summary statistics, including mean and Fano factor (Figure 2A [DOI](#), see Methods 1.2, and Supplementary Text 1.2). Thus, our loss function comprises two primary components: (i) the Kullback-Leibler (KL) divergence between the predicted distribution generated by the neural network $P_{\text{neuralnet}}$ and the labeled distribution generated by the SSA $P_{\text{simulation}}$, which is a widely used metric for quantifying the difference between two probability distributions; (ii) the logarithmic error of the predicted summary statistics computed by the neural network $\hat{s}$ and the labeled summary statistics s computed by theoretical derivation (Figure 2A [DOI](#) and see Methods 1.3). Furthermore, we compared various optimizers and hyperparameters of model architectures to determine the best predicting performance (see Methods 1.3 and Supplementary Figure S2).

Consequently, the loss function of DeepTXsolver converges effectively during the training process. We utilized DeepTXsolver to predict the stationary distribution using parameters from the test set. We observed that it effectively fits all four representative distributions from the TXmodel, encompassing unimodal distribution at zero point, unimodal distribution at non-zero point, bimodal distribution with one peak at zero point, and bimodal distribution with both peaks at non-zero point (Figure 2B [DOI](#)). These distributions have been extensively linked to cell fate decisions in numerous experiments (Gupta PB et al., 2011 [DOI](#); Cohen AA et al., 2008 [DOI](#); Bessarabova M et al., 2010 [DOI](#)). Moreover, we demonstrated that DeepTXsolver accurately predicted crucial distribution properties, including mean, Fano factor, and bimodal coefficient, with a high correlation between theoretical and predicted summary statistics ($r > 0.99$, $p\text{-value} < 2.2 \times 10^{-16}$ of t-test Figure 2C [DOI](#)). It is noteworthy that, to assess the impact of the presence or absence of summary statistics on DeepTXsolver, we performed separate experiments using different loss functions. We demonstrated that incorporating summary statistics into the model led to quicker convergence of DeepTXsolver on the training set (Figure 2D [DOI](#)) and yielded more robust and accurate predictions on the test set (Figure 2E [DOI](#)).

Overall, DeepTXsolver effectively establishes mappings from parameters to stationary distributions. This approach circumvents the high computational resource requirements of classical simulation algorithms and ensures the algorithm's efficiency and scalability in subsequent statistical inference, particularly in genome-wide inference.

DeepTX infers genome-wide transcriptional kinetics from scRNA-seq data

Having obtained the trained neural network model, we can rapidly compute the stationary distributions of any TXmodel parameters. These distributions represent the outcomes of the underlying true gene expression process, referred to as the “true distribution” P_{model} (Figure 3A [DOI](#)). However, during the inference process, the gene expression distributions measured by the scRNA-seq data we utilized, termed the “observed distribution of the data” P_{data} , were subject to noise, including errors stemming from the measurement technique (Sarkar A & Stephens M, 2021 [DOI](#)) (Figure 3A [DOI](#)). Thus, we introduced a mechanistic hierarchical model to bridge the gap between the true and observed distributions. This hierarchical model considers that the observed data results from the interplay between the gene expression process and the measurement process. It combines the two distributions using a convolutional form

$P_{\text{data}}(Y=y) = \int_0^\infty P_{\text{measure}}(y|x)P_{\text{model}}(x)dx$ (see Methods 1.4). $P_{\text{model}}(x)$ represents the mixed negative binomial distribution approximated, whereas for $P_{\text{measure}}(y|x)$, we opted for the binomial distribution. This choice is informed by the fact that it can approximate hypergeometric

distributions representing sequencing sampling without replacement when the sample size is sufficiently large. Consequently, the final distribution can maintain the form of a mixed negative binomial distribution (Supplementary Note 1.4), denoted as the observed distribution of the model P_{data} (Figure 3A [↗](#)).

Leveraging the hierarchical model, we can construct a scalable and interpretable module called DeepTXinferer, facilitating genome-wide inference of transcriptional burst kinetics from scRNA-seq data. First, we input the prior distribution of mechanistic model parameters and a set of scRNA-seq data, and we can get the observed distribution of the model, P_{obsmodel} (derived successively from the trained neural network and the hierarchical model) and the observed distribution of the data P_{obsdata} , respectively (Figure 3A [↗](#)). Second, we used the Hellinger distance $L(\theta) = D(P_{\text{obsmodel}}, P_{\text{obsdata}})$ to measure the difference between P_{obsmodel} and P_{obsdata} (see Methods 1.4). It's worth noting that the loss function remains solvable via gradients, despite the inclusion of complex computations, such as neural networks and hierarchical models. Specifically, to avoid the gradient-based optimization process from getting trapped in local optima, we employ a black-box optimization approach to select the initial point for parameter inference (Das S & Suganthan PN, 2010 [↗](#)). Consequently, we utilize a gradient-based optimizer to update the iterative parameters until convergence. Lastly, we employ a method based on loss potentials, which computes the posterior distribution of the parameters during the optimization of the loss function (Gaskin T et al., 2023 [↗](#)) (Figure 3A [↗](#), see Methods 1.4). Subsequently, this process yields results of the transcriptional burst kinetics derived from the estimated parameters of the mechanistic model.

We assessed the effectiveness of DeepTXinferer by conducting validation on synthetic data (see Methods 1.4). Synthetic samples were generated for each parameter set, and our inference algorithm was then applied to derive the output transcriptional burst kinetics (see Methods 1.4). As a result, the estimation of transcriptional burst kinetics demonstrated high accuracy and strong correlation across most parameter regions ($r > 0.99$, $p\text{-value} < 2.2 \times 10^{-16}$ of t-test Supplementary Figure S3b, c), particularly within the region of fully expressed mRNA abundance (Figure 3B [↗](#)). Moreover, the estimated posterior distribution of each parameter suggests that the peak value closely approximates the true value (Figure 3C-E [↗](#)). Additionally, we verified the robustness of our algorithm concerning the number of samples, observing that higher sample numbers corresponded to increased accuracy in estimation (Supplementary Figure 3d, e). Compared to previous work (Larsson AJM et al., 2019 [↗](#); Luo S et al., 2023 [↗](#); Tang W et al., 2023 [↗](#); Gu J et al., 2025 [↗](#)), DeepTX demonstrates superior performance in both efficiency and accuracy (Supplementary Figure S4). In summary, our inference algorithm exhibited strong performance on synthetic data, affirming its reliability for application to real-world data in inferring transcriptional bursts.

In the subsequent sections, we will employ the DeepTXinferer to explore how cells determine diverse fates in response to DNA damage stimuli through genome-wide regulation of transcriptional bursting. We analyzed three sets of scRNA-seq data representing distinct cell fates: cell differentiation, apoptosis, and survival. The first set of data is a mutually controlled scRNA-seq dataset of mESC cell differentiation induced by IdU drug treatment. This data set contains 12481 genes of transcriptomes from 812 cells and 13780 genes of transcriptomes from 744 cells, respectively. The second set of data is a mutually controlled scRNA-seq dataset of human colon cancer cells treated with low-dose 5FU drug that causes apoptosis. This data set contains 8534 genes of transcriptomes from 1673 cells and 7077 genes of transcriptomes from 632 cells, respectively. The third set of data is a mutually controlled scRNA-seq dataset of human colon cancer cells in which high-dose 5FU drug treatment resulted in cell-resistant survival. This data set contains 8534 genes of transcriptomes from 1673 cells and 6661 genes of transcriptomes from 619 cells, respectively.

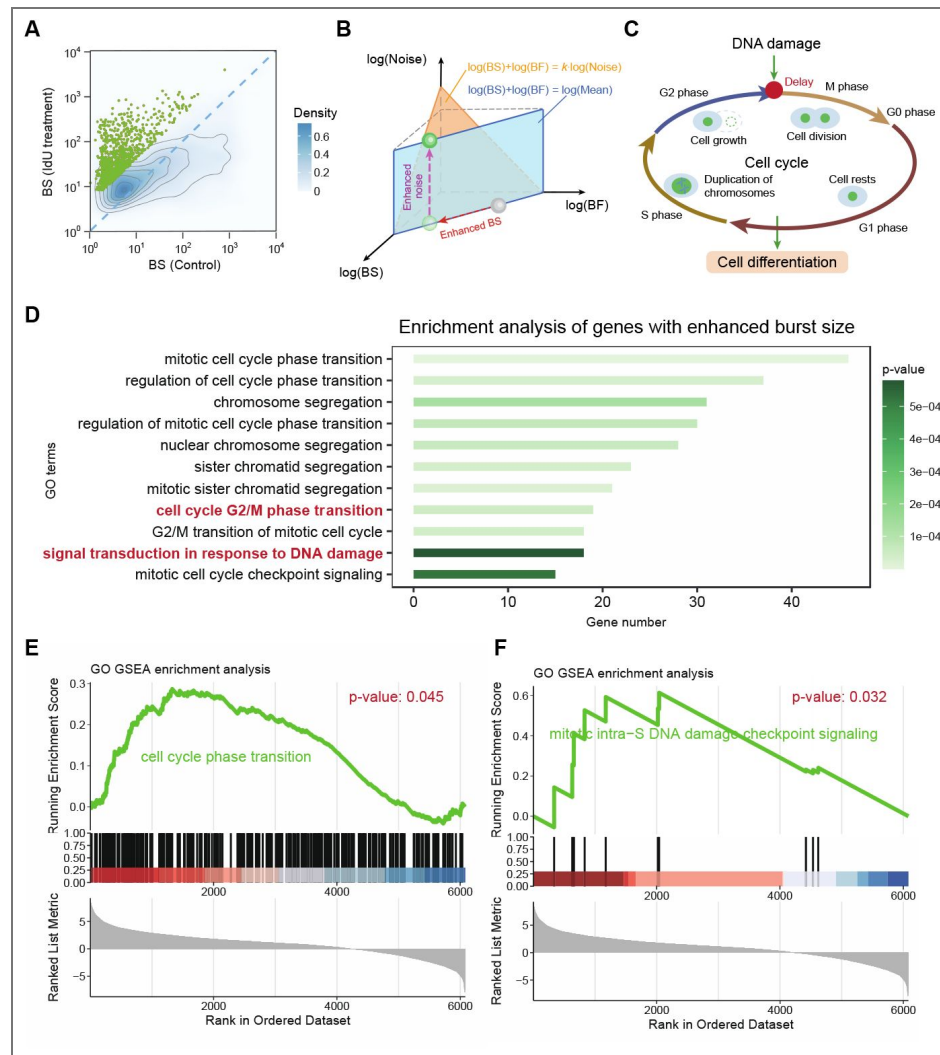


Figure 3. Inferring TXmodel from scRNA-seq data.

(A) The dynamic parameters are sampled from a given prior distribution, and as input to the neural network, the solution of the corresponding dynamic model can be obtained by using this parameter. The solution of the dynamic model is mixed with the binomial distribution to obtain the model observed distribution. Loss values were obtained using the observed distribution of the data obtained from the scRNA-seq data compared to the observed distribution of the model. The loss values are optimized until convergence using gradient descent to obtain the parameters of the mechanism model and the posterior distribution of its parameters. (B) The scatters represent the real BS and BF values of the SSA synthetic set parameters, and the depth of the color represents the error between the inferred BS (BF) and the true BS (BF). (C, D) The blue solid line represents the edge density of the burst kinetics of the model, and the red dashed line represents the true value. (E) The marginal density for the five parameters of the model, where the red dashed line represents the true value. The relative error (RE) quantifies the discrepancy between the true parameter and the peak of the inferred posterior distribution

DeepTX reveals burst size enhancement induced by IdU treatment associated with cell differentiation

The administration of the IdU drug, a thymine analogue, influences the gene expression process within cells, subsequently impacting cell differentiation (Desai RV et al., 2021 [DOI](#); Li C et al., 2024 [DOI](#)). The perturbation in the gene expression process primarily stems from DNA damage, as IdU is randomly integrated into the DNA chain, thereby inducing DNA damage (Li C et al., 2024 [DOI](#)). Although DNA damage caused by IdU treatment has little impact on genome-wide mean gene expression, it increases variance across the genome, thereby affecting cell differentiation (Desai RV et al., 2021 [DOI](#)). This is different from the understanding that increases in variance are usually caused by fluctuations in the mean gene expression (Newman JR et al., 2006 [DOI](#)). Meanwhile, the mean can be deconvolved as the product of BS and BF. This raises the question of the relationship between bursting dynamics and variance on a genome-wide scale, and how changes in bursting dynamics affect cell differentiation.

We used the DeepTX framework to infer a set of preprocessed scRNA-seq data (see Methods 2.1) to obtain underlying burst kinetics. The difference between the statistics and distribution obtained by inference and the statistics and distribution of sequencing data is minimal, ensuring the correctness of the inference results (Supplementary Figure S5a-l). Subsequently, we observed considerable variations in BS and BF across different genes (Figure 4A [DOI](#), Supplementary Figure S6d-f). Notably, genes with significantly increased BS exhibit corresponding decreases in BF, suggesting relative stability in the average gene expression level (Supplementary Figure S6d-f). Additionally, we observed a significant increase in transcriptional BS for most genes in the treatment group (Figure 4A [DOI](#)). This indicates that the rise in variance primarily results from the enhanced BS induced by DNA damage (Figure 4B [DOI](#)). Notably, this conclusion is consistent with the results of theoretical analysis (Supplementary text 2). To further explore the biological characteristics of BS enhancement, we identified genes with upregulated BS using differential analysis method. Further, we perform GO enrichment analysis (Gene Ontology Consortium, 2015 [DOI](#)) on the identified genes. We observed that the genes with enhanced BS revealed significant enrichment in terms related to mitotic cell cycle checkpoint signaling, alongside pathways involved in DNA damage response and cell cycle transitions (Smits VA & Medema RH, 2001 [DOI](#)) (Figure 4D [DOI](#)). Specifically, the mitotic checkpoint serves as a crucial safeguard to ensure accurate chromosome segregation and maintain genomic stability under DNA damage conditions. Activation of the mitotic checkpoint can influence cell fate decisions and differentiation potential. Sustained activation of the spindle assembly checkpoint (SAC) has been reported to induce mitotic slippage and polyploidization, which in turn may enhance the differentiation potential of embryonic stem cells. Subsequently, we performed Gene Set Enrichment Analysis (GSEA) (Subramanian A et al., 2005 [DOI](#)) on the identified differentially expressed genes associated with BS and also observed enrichment in pathways related to cell cycle transitions and mitotic intra-S DNA damage checkpoint signaling (Figure 4E, F [DOI](#)). Similarly, we subjected genes significantly enhanced by BF in the treatment group (Supplementary Figure S6f) to GO analysis (Supplementary Figure S6g). While it is tempting to hypothesize that enhanced BS may contribute to DNA damage-related checkpoint activation and thereby influence cell cycle progression and differentiation, our current results only indicate an association between burst size enhancement and pathways involved in DNA damage response and checkpoint signaling. This is consistent with previous experimental observations showing that IdU treatment induces DNA damage and alters cell fate decisions through effects on the cell cycle (Rosina M et al., 2019 [DOI](#); Riccio F et al., 2022 [DOI](#); Liu Z et al., 2017 [DOI](#)) (Figure 4C [DOI](#)). Additionally, research indicates that enhanced cell differentiation results from delayed cell cycle transitions (Rosina M et al., 2019 [DOI](#)).

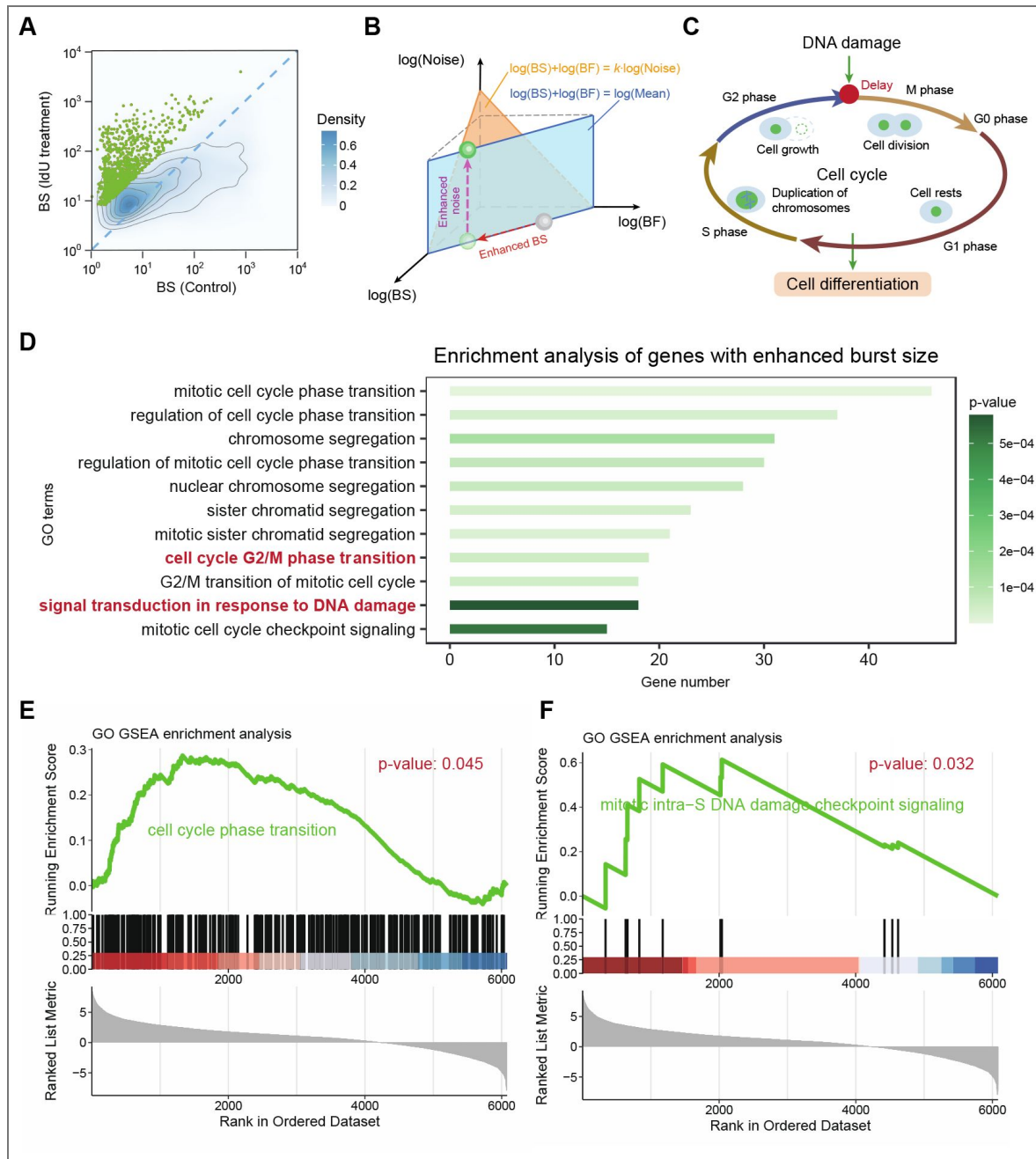


Figure 4. Burst size enhancer delays cell cycle.

(A) Density map of BS of genes treated with IdU drugs compared to BS of genes not treated with drugs. Green dots represent up-regulated BS differential genes. (B) Schematic representation of the simultaneous enhancement of both BS and noise of a gene accompanied by the mean value of the gene remaining unchanged after IdU drug treatment. In the log-log-log 3D space, the mean expression level lies on a blue plane defined by $\log(\text{BS}) + \log(\text{BF}) = \log(\text{Mean})$, indicating that it is determined by the product of burst size and burst frequency. The orange plane captures a scaling relationship between expression noise and burst kinetics, described by $\log(\text{BS}) + \log(\text{BF}) = \kappa \log(\text{Noise})$, where κ is a constant reflecting their covariation. Notably, the trajectory of the green sphere shows that, under a fixed mean expression level (i.e., confined to the blue plane), increased gene expression noise primarily results from an increase in burst size. (C) The schematic diagram illustrates the hypothesis that DNA damage induced by IdU treatment may affect gene transcription, potentially influencing cell cycle transitions during mitosis and regulating cellular differentiation. (D) Gene GO enrichment analysis is performed on the green dots of BS in (A) to obtain enrichment pathway diagram. The bold red pathways highlight the key components involved in the cellular differentiation mechanism. (E, F) Pathways obtained by GSEA enrichment analysis of BS of genes.

DeepTX reveals burst frequency enhancement under low-dose 5FU treatment and its association with apoptosis

The chemotherapeutic compound, 5-fluorouracil (5FU), exerts its cytotoxic effects on colorectal cancer cells by inflicting damage to their DNA (Kuipers EJ et al., 2015 [\[1\]](#); Bunz F et al., 1999 [\[2\]](#); Chang GS et al., 2014 [\[3\]](#)). This damage triggers apoptosis, a process of programmed cell death, thus inhibiting the growth and proliferation of malignant cells (Kuipers EJ et al., 2015 [\[1\]](#); Bunz F et al., 1999 [\[2\]](#); Chang GS et al., 2014 [\[3\]](#)). However, the impact of 5FU-induced DNA damage on the gene expression process and its potential relationship to cell apoptosis remains incompletely understood. In this section, we employ the DeepTX to investigate the effect of 5FU-induced DNA damage on the gene expression process in colorectal cancer cells with scRNA-seq data (Park SR et al., 2020 [\[4\]](#)). First, we inferred the burst kinetics from the scRNA-seq data of cells treated with low doses (10 μ M) of 5FU and controls. After preprocessing the data (see Methods 2.2), we inferred the scRNA-seq data to obtain a distribution that fits the scRNA-seq data (Supplementary Figure S7a-l), as well as the underlying burst kinetics (Figure 5A, B [\[5\]](#)). Further, we found that the mean difference between the two groups of scRNA-seq data was not significant (p-value = 0.5 of t-test), but the variance and BS were significantly different (p-value < 0.05 of t-test Figure 5A [\[5\]](#) and Supplementary Figure S8a-c). Then, we performed a differential analysis of the BS and BF of gene expression between the control cells and the cells treated with 10 μ M 5FU, and then performed GO gene enrichment analysis on differential genes (Figure 5C [\[6\]](#) and Supplementary Figure S8d-f). We found that BS down-regulated differential genes were mainly enriched in apoptosis pathways (Hongmei Z, 2012 [\[7\]](#)), electron exchange pathways, and oxidation-related pathways (Figure 5E [\[8\]](#) and Supplementary Figure S8g). Specifically, the enrichment analysis indicated that although the burst size of genes in 5FU-treated cells was generally larger than that in the control group, the genes primarily influencing cell apoptosis exhibited a down-regulated BS.

These enrichment results are in line with previously reported biological processes (Figure 5D [\[9\]](#)). First, studies in recent years have found that many anticancer drugs trigger apoptosis and necrosis of cancer cells by activating the production of reactive oxygen species (ROS) (Sun Y & Rigas B, 2008 [\[10\]](#)). And experiments have shown that 5FU can induce ROS production in colorectal cancer cells (oxidation-related pathways) (Laha D et al., 2015 [\[11\]](#); Bwatanglang IB et al., 2016 [\[12\]](#); Chenna S et al., 2022 [\[13\]](#)). Meanwhile, mitochondria are the main source of ROS production, and defects in the mitochondrial electron transport chain will increase ROS production (electron exchange pathways) (Adam-Vizi V, 2005 [\[14\]](#)). As a result, excessive production of ROS can cause mitochondrial DNA damage and nuclear DNA damage (Figure 5D [\[9\]](#)). In particular, when mitochondrial DNA is damaged, mitochondrial damage repair will further increase the pressure of ROS (Figure 5D right [\[15\]](#)) (Shokolenko IN et al., 2014 [\[16\]](#)). Overall, our results support an association between 5FU-induced burst kinetics changes and pathways implicated in oxidative stress and apoptosis (Figure 5D [\[9\]](#)) (Handali S et al., 2018 [\[17\]](#); Hwang PM et al., 2001 [\[18\]](#)).

DeepTX reveals burst frequency enhancement under high-dose 5FU treatment and its association with survival pathways

Low-dose 5FU treatment can induce cell apoptosis, while research shows high-dose 5FU treatment can induce cell resistance and evade apoptosis to survive (Kuranaga N et al., 2001 [\[19\]](#)). Hence, this part mainly studies the impact of differences in burst kinetics of gene expression between high-dose (50 μ M) 5FU-treated and control colorectal cancer cells on cell fate decisions.

After data preprocessing (see Methods 2.2), we similarly observed the phenomenon that the average expression value of genes remained stable, while there is an enhanced increase in the variance genome-wide (Supplementary Figure S10a-c). Then, after preprocessing the data (see Methods 2.2), we inferred the scRNA-seq data to obtain the distribution that fits the scRNA-seq data (Supplementary Figure S9a-l), as well as the underlying burst kinetics (Figure 6A [\[20\]](#)). Further, we performed a difference analysis on the BS of the two sets of inferred data, and selected genes with large differences in BS and BF for GO gene enrichment (Figure 6A, B [\[21\]](#) and Supplementary Figure S10d-f). We found that the BS down-regulates differential genes were mainly enriched in

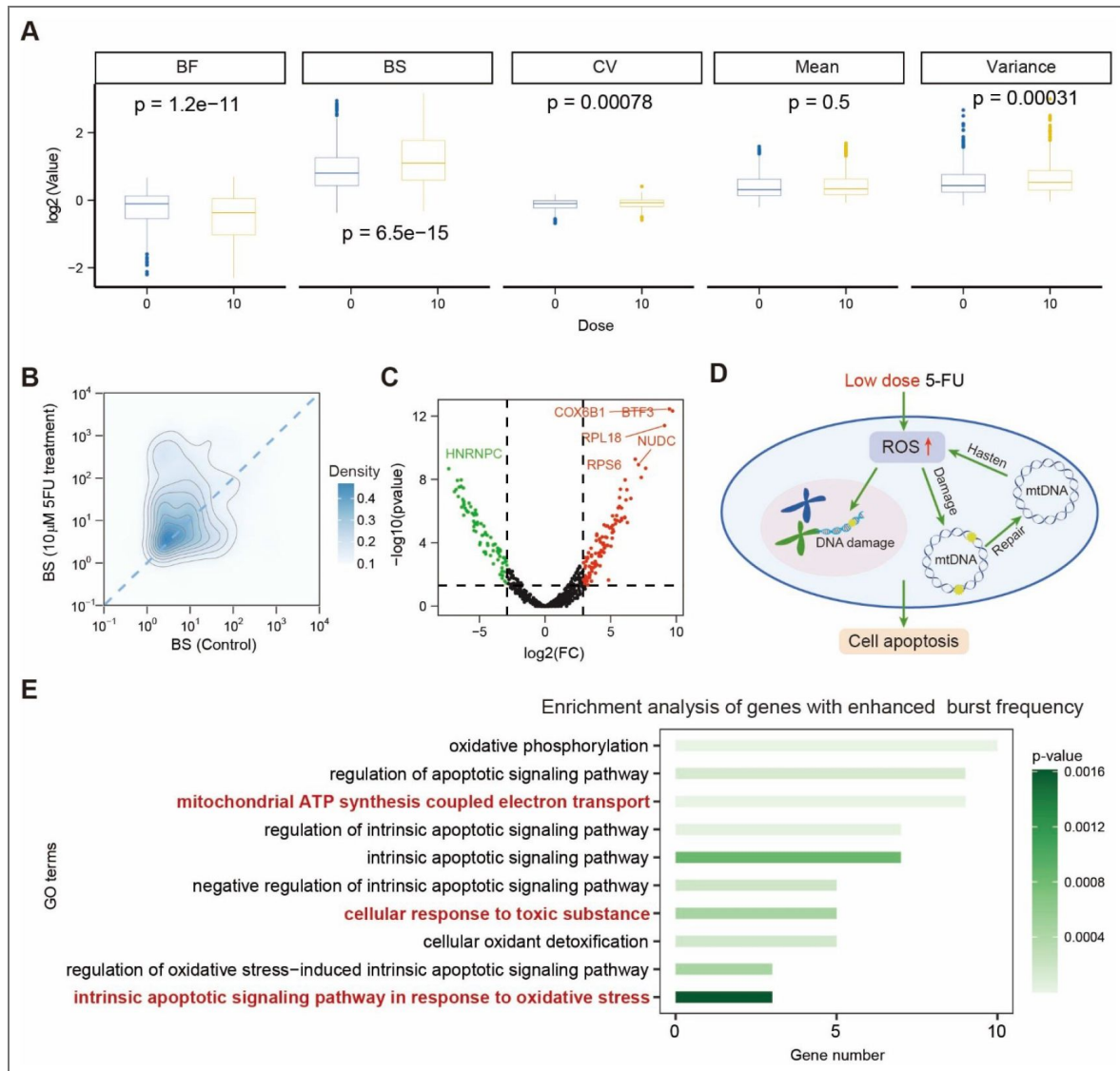


Figure 5. Low dose 5FU treatment induces cell apoptosis by enhancing burst frequency.

(A) Comparison of gene bursting kinetics and statistics between control and 10-dose 5FU treated cells. **(B)** Density map of inferred burst size. **(C)** Volcano map of the fold changes (FC) of inferred burst size between control and 10-dose 5FU treated cells. **(D)** The diagram hypothesizing the potential association between low-dose 5FU treatment and the induction of apoptosis. **(E)** The enrichment analysis results were derived from genes that were downregulated in BS and upregulated in BF in the experimental group. The bold red pathways highlight the key components involved in the cellular apoptotic mechanism.

the Cajal body related pathway, the electron exchange pathway, and telomerase-related pathways (Nabetani A & Ishikawa F, 2011 [\[10\]](#)), but the apoptosis-related pathways were not significant (Figure 6D [\[10\]](#)). In particular, according to experimental studies, telomerase-related pathways can activate telomerase and play an important role in the elongation of telomeres (Cristofari G et al., 2007 [\[11\]](#)). These enrichment results suggest that high-dose 5FU treatment may be associated with processes such as telomerase activation and mitochondrial function maintenance, both of which have been implicated in cell survival and apoptosis evasion in previous experimental studies. For example, prior work indicates that hTERT translocation can activate telomerase pathways to support telomere maintenance and reduce oxidative stress, which is thought to contribute to apoptosis resistance (Kuranaga N et al., 2001 [\[12\]](#)). Overall, the observed transcriptional bursting changes are consistent with these reported survival-associated mechanisms.

Gene enrichment results are consistent with existing ways of cells evading apoptosis (Figure 6C [\[10\]](#)). That is, endogenous ROS production exceeds the cellular antioxidant defense capacity, leading to chemical damage to mitochondrial DNA (mtDNA). Meanwhile, high doses of 5FU causes Human Telomerase Reverse Transcriptase (hTERT) translocation, which on the one hand triggers telomerase activity and maintenance pathways, which enables cells to evade apoptosis (Kuranaga N et al., 2001 [\[12\]](#)). On the other hand, this translocation will enrich hTERT in mitochondrial DNA, thereby reducing oxidative stress and mitochondrial DNA damage to protect mitochondria and prevent cells from apoptosis (Lipinska N et al., 2017 [\[13\]](#)).

Resistance to therapy has also been associated with G0 arrest, a non-proliferative state characteristic of persister cells (Wiecek AJ et al., 2023 [\[14\]](#)). To verify the presence of resistant cells in cells treated with high doses of 5FU, we selected genes related to G0 arrest and performed GSA (Hänzelmann S et al., 2013 [\[15\]](#)) to obtain the G0 arrest quality score (Wiecek AJ et al., 2023 [\[14\]](#)) of each cell and perform UMAP dimensionality reduction visualization (Figure 6E [\[10\]](#)). Through arrest quality score discrimination, we found that the proportion of resistant cells was large among cells treated with high doses of 5FU (Figure 6F [\[10\]](#)). This result was consistent with the results of our enrichment analysis. Specifically, by comparing the BS and mean values of these genes, we found that most of the BS of G0 arrest related genes were up-regulated, but the mean changes were small (Figure 6G [\[10\]](#)).

Discussion

Gene expression is inherently stochastic, and transcriptional bursting contributes to inter-individual variability in gene expression patterns (Dar RD et al., 2012 [\[16\]](#); Zenklusen D et al., 2008 [\[17\]](#); Phillips R et al., 2012 [\[18\]](#)), although it represents only one of multiple sources of cellular heterogeneity. Experimental observations confirm this bursting mechanism as a cellular response to environmental changes, including DNA damage (Friedrich D et al., 2019 [\[19\]](#); Desai RV et al., 2021 [\[20\]](#)), but its precise role in DNA damage response and subsequent cell fate decisions remains an emerging hypothesis. In this study, we introduced a mechanism-based deep learning framework (DeepTX) for performing genome-wide inference on transcriptional burst kinetics of a comprehensive gene expression model using scRNA-seq data. DeepTX provides an efficient computational approach to characterize burst kinetics across genes, enabling the generation of testable hypotheses about how transcriptional bursting may be associated with DNA damage response and potential changes in cellular state.

Utilizing DeepTX, we analyzed three sets of scRNA-seq data comprising control and stimulus pairs driving different cell fate decisions: cell differentiation, apoptosis, and survival. This analysis unveiled significant insights into the underlying mechanisms of transcriptional burst kinetics. Firstly, the IdU drug treatment-induced enhancement of BS in genes appears to be associated with a delayed transition in the cell mitosis phase, which may in turn be related to changes in cell reprogramming and differentiation. Interestingly, while prior studies have suggested that noise enhancement contributes to cell fate decisions, our observations indicate that BS enhancement may reflect a deeper regulatory feature correlated with cell differentiation. Secondly, low-dose 5FU treatment is associated with increased oxidative stress and an accompanying elevation in BF, which may relate to enhanced apoptotic activity. Consistent with previous findings linking

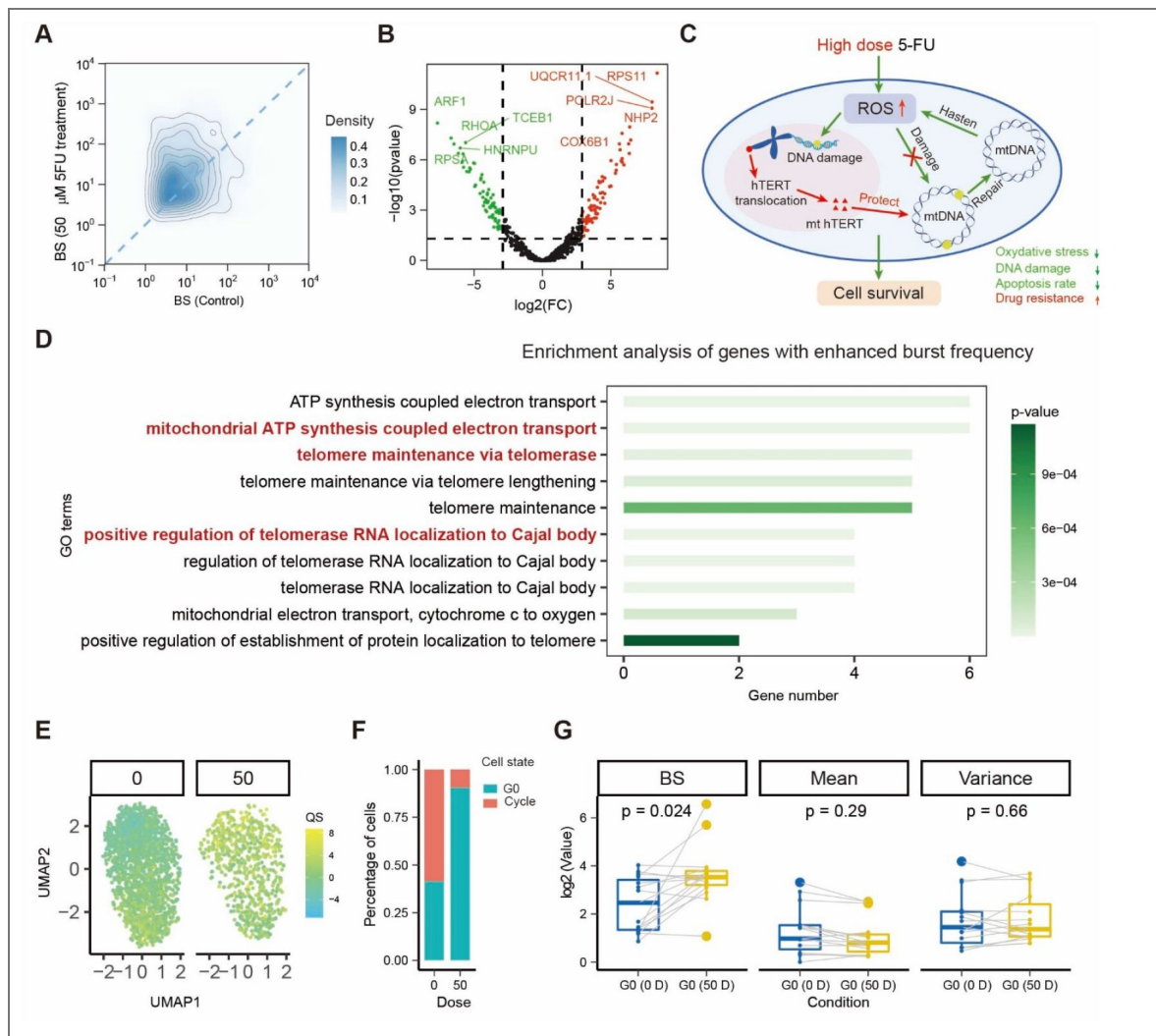


Figure 6. High dose 5FU treatment may promotes cell drug resistance by altering burst frequency.

(A) Density map of inferred burst size. (B) Volcano map of the fold changes (FC) of inferred burst size between control and 50-dose 5FU treated cells. (C) Schematic diagram hypothesizing the mechanism by which 5FU drug treatment may induce cells to acquire antioxidant properties, potentially enabling continued survival. (D) The enrichment analysis results were derived from genes that were downregulated in BS and upregulated in BF in the experimental group. (E) UMAP of G0 arrest quality score (QS) of control and 50-dose 5FU treated cells. (F) Proportion of G0 arrested (drug-resistant) and cycling (non-resistant) cells. (G) Comparison of gene bursting kinetics and statistics of G0 arrested genes between control and 50-dose 5FU treated cells. The light grey lines indicate the trend of changes in the corresponding bursting kinetics values, mean and variance for each gene.

apoptosis to ROS stress, our study suggests that ROS-induced DNA damage is correlated with changes in gene expression burst kinetics, which may contribute to the apoptotic outcomes observed. Thirdly, high-dose 5FU treatment induces telomerase extension, mitigating oxidative stress and fostering drug resistance by upregulating BF. Despite prior evidence of drug-resistant cells, our enrichment analysis underscores telomerase extension's role in resisting oxidative stress, underscoring the impact of BF enhancement.

We highlight several advantages of our DeepTX framework, which differs significantly from previous work that utilizes scRNA-seq data to study diverse noise sources from several aspects (Eling N et al., 2019 [↗](#); Faure AJ et al., 2017 [↗](#); Morgan MD & Marioni JC, 2018 [↗](#); Ochiai H et al., 2020 [↗](#)). Firstly, DeepTX employs a mechanistic hierarchical model that is both interpretable and extensible. To enhance the model's realism in interpreting scRNA-seq data, we developed a novel approach integrating the sequencing process and the underlying gene expression process. Specifically, we employed the TXmodel, a complex non-Markovian stochastic dynamic model that extends the waiting times for gene OFF and ON states into a non-exponential distribution, reflecting the multi-step nature of gene regulation. As a result, TXmodel not only captures gene expression dynamics under DNA damage but also depicts other multi-step regulatory processes, such as chromatin opening (Lomvardas S & Thanos D, 2002 [↗](#)), preinitiation complex formation (Fuda NJ et al., 2009 [↗](#)), transcription factors binding (Nordick B et al., 2022 [↗](#)). Secondly, despite the mathematical complexity of the mechanistic hierarchical model, DeepTX renders it tractable. This is achieved by (i) directly mapping model parameters to corresponding stationary distributions via a neural network trained on extensive simulated single-cell RNA data and theoretical analysis summary statistics for TXmodel; (ii) approximating the stationary distribution output of TXmodel, coupled with the sequencing process, by a mixed negative binomial distribution, maintaining the final distribution format. Thirdly, DeepTX's inference process is scalable, allowing genome-wide bursting kinetics inference from static scRNA-seq data. Despite the computational complexity of parameter inference involving hierarchical models and trained neural networks, DeepTX retains auto-differentiation for efficient and accurate gradient computation.

The DeepTX research paradigm contributes to a growing line of area aiming to link mechanistic models of gene regulation with scRNA-seq data. Maizels provided a comprehensive review of computational strategies for incorporating dynamic mechanisms into single-cell transcriptomics (Maizels RJ, 2024 [↗](#)). In this context, RNA velocity is one of the most important examples as it infers short-term transcriptional trends based on splicing kinetics and deterministic ODEs model. However, such approaches are limited by their deterministic assumptions and cannot fully capture the stochastic nature of gene regulation. DeepTX can be viewed as an extension of this framework to stochastic modelling, explicitly addressing transcriptional bursting kinetics under DNA damage. Similarly, DeepCycle, developed by Sukys and Grima (Sukys A & Grima R, 2025 [↗](#)), investigates transcriptional burst kinetics during the cell cycle, employing a stochastic age-dependent model and a neural network to infer burst parameters while correcting for measurement noise. By contrast, MIGNON integrates genomic variation data and static transcriptomic measurements into a mechanistic pathway model (HiPathia) to infer pathway-level activity changes, rather than gene-level stochastic transcriptional dynamics (Garrido-Rodriguez M et al., 2021 [↗](#)). In this sense, DeepTX and MIGNON are complementary, with DeepTX resolving burst kinetics at the single-gene level and MIGNON emphasizing pathway responses to genomic perturbations, which could inspire future extensions of DeepTX that incorporate sequence-level information.

The DeepTX research paradigm is poised for extension to encompass both dynamic models and biological data. On the one hand, our focus is currently confined to the transcriptional processes influenced by DNA damage, limiting the explanatory scope of the TXmodel to larger-scale biological phenomena such as alternative splicing (Wan Y et al., 2021 [↗](#); Gorin G & Pachter L, 2022 [↗](#)), protein translation (Luo S et al., 2023 [↗](#); Golan-Lavi R et al., 2017 [↗](#)), epigenetic modification (Nicolas D et al., 2018 [↗](#)), and chromatin movement (Liu T et al., 2016 [↗](#); Wang Z et al., 2024 [↗](#); Wang Z et al., 2023 [↗](#)). Moreover, further development is needed to incorporate

additional regulatory factors, including signaling pathways (Garrido-Rodriguez M et al., 2021 [DOI](#)), cell cycle progression, cell size (Sukys A & Grima R, 2025 [DOI](#)), and leaky transcription (Kepler TB & Elston TC, 2001 [DOI](#)). These broader biological processes necessitate more intricate dynamic models, often challenging to solve and thus challenging to apply to realistic biological data.

On the other hand, inference from static scRNA-seq relies on the steady-state assumption, which has some limitations. First, in some scenarios, the cell system may exhibit highly transient transcriptional programs that do not satisfy stationarity, leading to biased or misleading parameter estimates. For example, immediately following a synchronized developmental stimulus such as serum shock induced activation of immediate-early genes, transcription levels increase significantly (Hope B et al., 1992 [DOI](#)). Second, because DeepTX infers the mean burst frequency and size across the population, it cannot recover the underlying time-resolved dynamics or distinguish heterogeneous kinetic subpopulations. Third, lack of time-resolved measurements may affect the accuracy of inferences about dynamic parameters, especially the unidentifiability of parameters inferred from steady-state distributions, i.e., multiple parameters leading to the same steady-state distribution. The unidentifiability of individual parameters is a common and critical problem in systems biology studies. Hence, in DeepTX, we employed a Bayesian approach based on loss potential to infer the posterior distributions of the parameters. Although DeepTX also encounters the issue of unidentifiability for individual parameters (Supplementary Figure S11), the multimodal nature of the posterior distribution suggests that multiple distinct parameter sets can produce similarly good fits to the observed data, highlighting the inherent non-identifiability of the model. Nevertheless, in the multimodal posterior distribution, at least one of the posterior peaks aligns closely with the ground truth, thereby demonstrating the validity of the inferred result. In future work, integrating time-series single-cell measurements with other emerging data types could help overcome these limitations. Examples include increasingly comprehensive transcriptome data with both temporal and spatial resolution (Wen L & Tang F, 2022 [DOI](#)), as well as state-of-the-art multi-omics data (Maizels RJ, 2024 [DOI](#); Miao Z et al., 2021 [DOI](#)). Moreover, another limitation of the current implementation is that DeepTX is only trained on simulated datasets. While providing controlled benchmarks for mechanistic validation, it may not fully recapitulate the heterogeneity and stochastic complexity of real biological systems. In the future work, publicly available annotated scRNA-seq datasets could be used to complement this simulation-based training strategy and enhance generalizability.

In conclusion, DeepTX provides a step toward inferring complex transcriptional kinetics from scRNA-seq data through mechanism-based deep learning inference. It sheds light on the augmented burst kinetics in DNA damage response across various cell fate decisions, thereby offering novel insights.

Methods

1 The DeepTX framework

1.1 Framework overview

The DeepTX Framework aims to infer the underlying bursting dynamics from static single-cell RNA sequencing (scRNA-seq) data to explore the impact of changes in bursting dynamics on cell fate decisions. The DeepTX framework includes two modules, DeepTXsolver and DeepTXinferer. DeepTXsolver can efficiently and accurately solve transcription models using neural network architecture, and DeepTXinferer can infer scRNA-seq data to obtain its underlying transcription burst kinetics using Bayesian methods.

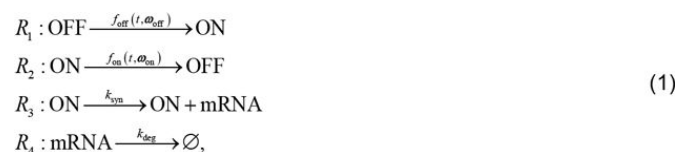
The input of the DeepTXsolver is the parameters of the mechanism model, and the output is the corresponding stationary distribution solution of the model. The input to the DeepTXinferer is the discrete probability distribution of gene expression obtained from scRNA-seq data, and the output of the model is the parameters of the mechanism model.

The construction of the DeepTX framework consists of the following parts: (1) Model the gene expression process. (2) Train a neural network that solves the mechanism model using the data set generated by stochastic simulation algorithm (SSA) simulation. (3) Hierarchical model modeling to infer underlying burst kinetic parameters from scRNA-seq data.

When building the DeepTX Framework, we made the following assumptions: (1) The gene expression of cells was in a stationary distribution during sequencing. This assumption has been widely adopted in scRNA-seq studies, as it effectively captures mRNA expression distributions and enables the inference of underlying dynamic parameters. (Larsson AJM et al., 2019 [\[1\]](#); Luo S et al., 2023 [\[2\]](#); Ramsköld D et al., 2024 [\[3\]](#); Gupta A et al., 2022 [\[4\]](#)) (2) Gene expression counts of the same cell type follow the same distribution. Given that cells of the same type typically share similar transcriptional programs, their gene expression distributions are approximately identical, supporting the validity of this commonly used assumption. (3) The solution of the model can be approximated by a mixed negative binomial distribution. Since the stationary solution of chemical master equation can be expressed as a Poisson mixture, which can be more efficiently approximated using negative binomial mixtures (Gorin G et al., 2024 [\[5\]](#)). (4) The state space sampled from a sufficiently long single simulation is statistically equivalent to that obtained from multiple simulations at steady state in gene expression models (Kuntz J et al., 2021 [\[6\]](#)).

1.2 Modelling stochastic gene expression processes

To recover dynamic transcriptional burst kinetic parameters from static scRNA-seq data, we require the stochastic modeling approach to profile the gene expression process. The traditional telegraph model is often used to model gene transcriptional bursting, which assumes that switching between genes follows an exponential distribution (i.e., switching is a single-step process) (Peccoud J & Ycart B, 1995 [\[7\]](#); Kim JK & Marioni JC, 2013 [\[8\]](#)). However, this assumption is an oversimplification of the gene expression process in the context of DNA damage. Due to the fact that DNA damage slows down and even stops the elongation of RNA pol II across the double-strand DNA (Lans H et al., 2019 [\[9\]](#)), the gene expression process should be regarded as a multi-step process. This property causes the waiting time for the gene switching between active and inactive states to follow a non-exponential distribution (Daigle BJ, Jr. et al., 2015 [\[10\]](#); Zhang J & Zhou T, 2019 [\[11\]](#); Zhang JJ & Zhou TS, 2019 [\[12\]](#)). Therefore, we use TXmodel (Luo S et al., 2023 [\[2\]](#)) to better characterize the gene transcription process in the context of DNA damage (Figure 2 [\[13\]](#)), whose chemical reaction network (CRN) can be written as follows:



where R_1 and R_2 describe the gene-state switching randomly between the ON state and OFF state, and $f_{\text{off}}(t, \omega_{\text{off}})$ and $f_{\text{on}}(t, \omega_{\text{on}})$ are the distribution of dwell time in ON and OFF state, respectively. Specifically, dwell time distribution refers to the probability distribution of the time in which a gene remains in a particular transcriptional state (ON or OFF) before transitioning to the other state. While the gene is in the active state, the synthesis of mRNA molecules is governed by a Poisson process with a constant synthesis rate k_{syn} (R_3). Meanwhile, each mRNA molecule can be stochastically degraded (R_4) with a constant degradation rate k_{deg} . Overall, the TXmodel can be determined by a set of parameters, denoted as $\theta = (\omega_{\text{off}}, \omega_{\text{on}}, k_{\text{syn}}, k_{\text{deg}})$.

The TXmodel as an extension of the traditional telegraph model can capture more realistic transcriptional burst kinetics, with BF and BS being two important parameters. The definition of BF is

$$\text{BF} = \frac{1}{\langle \tau_{\text{on}} \rangle + \langle \tau_{\text{off}} \rangle}, \quad (2)$$

where $\langle \tau_{\text{off}} \rangle = \int_0^\infty t f_{\text{off}}(t, \omega_{\text{off}}) dt$ and $\langle \tau_{\text{on}} \rangle = \int_0^\infty t f_{\text{on}}(t, \omega_{\text{on}}) dt$ are the mean dwell time in OFF and ON state, respectively. BF represents the average time it takes for the promoter to complete one full stochastic cycle between its active and inactive states. This definition is slightly different from the traditional definition $1/(k_{\text{deg}} * \tau_{\text{off}})$, which can be regarded as a simplified version of our definition (Eq. 2), under the assumptions that τ_{on} is negligible and $k_{\text{deg}} = 1$ (i.e., rate parameters are normalized to be dimensionless). Although it is reasonable to neglect activation time τ_{on} , as it is typically much shorter than inactive time under some conditions, we chose a more complete way to define the burst frequency so that it is applicable to more general situations. Additionally, Eq. 2 has been widely used in recent literature (Ramsköld D et al., 2024; Zoller B et al., 2018; Hoppe C et al., 2020). The definition of BS is

$$\text{BS} = k_{\text{syn}} \langle \tau_{\text{on}} \rangle. \quad (3)$$

BS refers to the average number of mRNA transcripts produced during a single transcriptional activation event of a gene. Notably, the mean transcription level of mRNA ($\langle y \rangle$) can be expressed as $\langle y \rangle = k_{\text{syn}} \langle \tau_{\text{on}} \rangle / (\langle \tau_{\text{on}} \rangle + \langle \tau_{\text{off}} \rangle) = \text{BS} \times \text{BF}$.

Without loss of generality, the gamma distribution can be used to characterize non-exponential distributions, i.e., $f_{\text{off}}(t, \omega_{\text{off}}) = r_{\text{off}}^{k_{\text{off}}} t^{k_{\text{off}}-1} e^{-r_{\text{off}} t} / \Gamma(k_{\text{off}})$ such that $\omega_{\text{off}} = (r_{\text{off}}, k_{\text{off}})$ and $f_{\text{on}}(t, \omega_{\text{on}}) = r_{\text{on}}^{k_{\text{on}}} t^{k_{\text{on}}-1} e^{-r_{\text{on}} t} / \Gamma(k_{\text{on}})$ such that $\omega_{\text{on}} = (r_{\text{on}}, k_{\text{on}})$, where $\Gamma(\cdot)$ is Gamma function, r_{off} and r_{on} are the rate-parameters of state switching, k_{off} and k_{on} are the possible number of reaction steps between OFF and ON state. The rationale for employing the gamma distribution lies in the complex, multi-step nature of promoter state transitions. Biologically, both the “ON” and “OFF” states of a promoter encompass multiple underlying sub-states, with transitions governed by a series of molecular events, each characterized by exponentially distributed waiting times. When aggregated, these sequential exponential processes yield non-exponential overall waiting time distributions. The gamma distribution, being a natural analytical extension of such a convolution of multiple exponential distributions, provides a mathematically and biologically appropriate approximation for such processes. Here, in the TX model, the state-switching waiting time follows a non-Markovian distribution (i.e., it is not exponentially distributed), indicating that the system exhibits non-Markovian dynamics.

It is a challenge to solve the stationary distribution of this gene expression system directly from a theoretical approach. But we proved that the binomial moments of this system can be obtained in an iterative form, by rationalizing this gene expression system as a queueing theory model. Thus, we can theoretically compute some important summary statistics $\mathbf{s}(\boldsymbol{\theta})$ given a parameter $\boldsymbol{\theta}$, such as mean, variance, kurtosis, and bimodal coefficient (see details in Supplementary Text 1.2).

1.3 Training neural network to solve gene expression model

The stationary distribution of CRN (1) is difficult to be solved analytically. In general, stochastic simulation algorithms are used to approximate the solution of chemical master equations (Boguñá M et al., 2014; Masuda N & Rocha LE, 2018). However, the inference process involves updating the parameters, and each update requires a simulation to obtain the stationary distribution, which makes statistical inference time-consuming. We note that deep learning methods may have potential applications in solving this challenge because it is widely used for approximating the solution of partial differential equations, chemical master equations, and non-Markovian chemical master equations (Jiang Q et al., 2021; Wang S et al., 2019; Davis CN et al., 2020; Gupta A et al., 2021; Bar-Sinai Y et al., 2019).

Therefore, in this part, we used a neural network to approximately solve CRN (1). This enables CRN (1) to be solved efficiently and accurately. Details will be introduced in the following part.

Generating synthetic training dataset

To construct the mapping from the parameters of TXmodel to the stationary distribution, we first need to generate the neural network's training set, i.e., a large-scale range of parameter sets as features and corresponding probability distributions as labels. For the features, we used the Sobol

algorithm (Sobol IM, 1967) to conduct large-scale sampling to generate the neural network data set, which can cover the entire parameter space more evenly than random sampling.

For the labels, we perform the SSA of the TXmodel to obtain the stationary distribution given a set of preset parameters. For clarity, the synthetic training dataset is denoted by $\{(\theta_i, P_{\text{simulation},i})\}$, where θ_i presents the model parameters of TXmodel, and $P_{\text{simulation},i}$ represents the stationary distribution solution obtained by SSA (Figure 2A). Specifically, the value range of each parameter component of θ_i is $k_{\text{off}} \in (1,15)$, $r_{\text{off}} \in (0.1,10)$, $k_{\text{on}} \in (1,15)$, $r_{\text{on}} \in (0.01,10)$, $k_{\text{syn}} \in (0.1, 400)$. The experimentally measured values of the mean duration of the OFF state, burst duration, transcriptional burst sizes, and the number of intermediate steps in transcriptional state transitions fall within the specified parameter ranges. (Tunnacliffe E & Chubb JR, 2020; Gupta A et al., 2022).

Deep learning model architecture

We build the basic deep learning model architecture in DeepTX via fully connected neural networks, which has been effectively used to solve approximate solutions from biochemical reaction networks (Sukys A et al., 2022; Gupta A et al., 2021). The deep learning model architecture consists of an input layer, hidden layers, and an output layer.

Firstly, the input layer contains 5 neurons corresponding to the parameters of the mechanism model $\theta_i = (k_{\text{off},i}, r_{\text{off},i}, k_{\text{on},i}, r_{\text{on},i}, k_{\text{syn},i})$. Secondly, the input layer is followed by n_{hidden} hidden layers with each layer containing $n_{\text{neurons}} = (n_{\text{neurons},1}, \dots, n_{\text{neurons},n_{\text{hidden}}})$ neurons. For the neurons in each hidden layer, we implement a nonlinear mapping by the commonly used ReLU activation function (Supplementary Figure S1):

$$\begin{aligned} h_j^{(1)} &= \text{ReLU}(w_{1j}^{(1)}k_{\text{off},i} + w_{2j}^{(1)}r_{\text{off},i} + w_{3j}^{(1)}k_{\text{on},i} + w_{4j}^{(1)}r_{\text{on},i} + w_{5j}^{(1)}k_{\text{syn},i}), (j=1,2,\dots,n_{\text{neurons},1}), \\ h_j^{(2)} &= \text{ReLU}(w_{1j}^{(2)}h_{n_{\text{neurons},1}}^{(1)} + w_{2j}^{(2)}h_{n_{\text{neurons},2}}^{(1)} + \dots + w_{n_{\text{neurons},j}}^{(2)}h_{n_{\text{neurons},j-1}}^{(1)}), (j=1,2,\dots,n_{\text{neurons},2}), \\ &\vdots \\ h_j^{(n_{\text{hidden}})} &= \text{ReLU}(w_{1j}^{(n_{\text{hidden}})}h_{n_{\text{neurons},1}}^{(n_{\text{hidden}}-1)} + w_{2j}^{(n_{\text{hidden}})}h_{n_{\text{neurons},2}}^{(n_{\text{hidden}}-1)} + \dots + w_{n_{\text{neurons},j}}^{(n_{\text{hidden}})}h_{n_{\text{neurons},j-1}}^{(n_{\text{hidden}}-1)}), (j=1,2,\dots,n_{\text{neurons},n_{\text{hidden}}}), \end{aligned} \quad (4)$$

where $h_j^{(i)}$ represents the value of the j -th neuron of the n_{hidden} -th hidden layer of the i -th group of parameters, $w_{kj}^{(i)}$ represents the neural network weight connecting the k -th neuron of the n_{hidden} -th hidden layer and the j -th neuron of the $(n_{\text{hidden}} - 1)$ -th hidden layer.

Thirdly, In the output layer, we aim to allow neurons to approximate synthetic scRNA-seq data generated by TXmodel. Negative binomial distributions are a class of distributions that are widely used in modeling scRNA-seq data and have been shown to account for both the “drop-out” phenomenon and the fact that the Fano factor is greater than 1 in scRNA-seq data (Risso D et al., 2018; Love MI et al., 2014). Also, the negative binomial distribution has been employed as an approximate stationary solution to the TXmodel for many gene expression systems (Sukys A et al., 2022; Öcal K et al., 2022). We assign weights and parameters (a_{ij}, m_{ij}, l_{ij}) to the output layer neurons (Supplementary Figure S1):

$$\begin{aligned} h_j^{(\text{output})} &= w_{1j}^{(\text{output})}h_{n_{\text{neurons},1}}^{(n_{\text{hidden}})} + w_{2j}^{(\text{output})}h_{n_{\text{neurons},2}}^{(n_{\text{hidden}})} + \dots + w_{n_{\text{neurons},j}}^{(n_{\text{hidden}})}h_{n_{\text{neurons},j-1}}^{(n_{\text{hidden}})}, (j=1,2,\dots,n_{\text{output}}), \\ (a_{i1}, a_{i2}, \dots, a_{i n_{\text{NB}}}) &= \text{Softmax}(h_{i1}^{(\text{output})}, \dots, h_{i,3^{*}j+1}^{(\text{output})}, \dots, h_{i,3^{*}n_{\text{NB}}-2}^{(\text{output})}), \\ m_{ij} &= \text{Exp}(h_{i,3^{*}n_{\text{NB}}-1}^{(\text{output})}), j=1,\dots,n_{\text{NB}}, \\ l_{ij} &= \text{Sigmoid}(h_{i,3^{*}n_{\text{NB}}}^{(\text{output})}), j=1,\dots,n_{\text{NB}}, \end{aligned} \quad (5)$$

where $h_j^{(\text{output})}$ represents the value of the j -th neuron of the output layer of the i -th set of parameters, a_{ij} is the weight coefficient of the negative binomial distribution, satisfying $\sum_{i=1}^{n_{\text{NB}}} a_{i1} = 1$, l_{ij} , m_{ij} represents the i -th parameter of the j -th negative binomial distribution of the mixed negative binomial distribution, n_{NB} is the number of negative binomial distribution. And the mixed negative binomial distributions can be expressed as (Supplementary Figure S1):

$$\begin{aligned} P_{\text{neuralnet},i}(n_{\text{counts}}; \mathbf{w}(\theta_i)) &= \sum_{j=1}^{n_{\text{NB}}} a_{ij}(\mathbf{w}(\theta_i)) \text{NB}(n_{\text{counts}}; m_{ij}(\mathbf{w}(\theta_i)), l_{ij}(\mathbf{w}(\theta_i))), \\ \text{where } \text{NB}(N=n_{\text{counts}}; l, m) &= \binom{n_{\text{counts}} + m - 1}{n_{\text{counts}}} l^{n_{\text{counts}}} (1-l)^m. \end{aligned} \quad (6)$$

where $NB(N = n_{\text{counts}}, l, m) = \binom{n_{\text{counts}} + m - 1}{n_{\text{counts}}} l^{n_{\text{counts}}} (1-l)^m$.

Loss function of DeepTXsolver

Our loss function includes two parts: one part is the KL divergence between the mixed negative binomial distribution generated by neural network $P_{\text{neuralnet},i}$ and label distribution $P_{\text{simulation},i}$ generated by SSA, and the another part is the log difference of the moment statistics between theoretical solution of the TXmodel s_{ij} and the output of the neural network $\hat{s}_{ij}$:

$$\text{Loss} = \sum_{i=1}^{n_{\text{batch}}} \left(\text{KL}(P_{\text{simulation},i}, P_{\text{neuralnet},i}) + \lambda \sum_{j=1}^{n_{\text{stats}}} \left(\log \left(\frac{s_{ij}}{\hat{s}_{ij}} \right) \right)^2 \right), \quad (7)$$

where $\text{KL}(P_{\text{simulation},i}, P_{\text{neuralnet},i}) = \sum_{n_{\text{counts}}} P_{\text{simulation},i} \log(P_{\text{neuralnet},i} / P_{\text{simulation},i})$, and λ is the hyperparameter, represents the weight coefficient of the statistical loss, and $\lambda = 0.1$, n_{batch} represents the number of samples, where each sample is characterized by $(\theta_i, P_{\text{simulation},i})$ in each batch, and n_{stats} is the number of summary statistics. For the latter part of loss function, we can still obtain the moment statistic by theoretical derivation, although the analytic distribution is difficult to solve. Adding moment statistics to the loss function can make the neural network more robust (Figure 2E). Specifically, we use four statistics here: s_{i1} indicates the mean, s_{i2} denotes the Fano factor, s_{i3} denotes kurtosis, s_{i4} represents the bimodal coefficient. Specifically, the Fano factor, which normalizes variance by the mean, provides a robust measure of transcriptional noise across genes or conditions with varying expression levels. It's worth mentioning that the statistics here is scalable and we can add more information inside the loss function.

Training details of DeepTXsolver

Effective model training necessitates the selection of an appropriate optimizer to minimize the loss function. In our case, we employed the Adam optimizer (Kingma DP & Ba J, 2014), a widely adopted choice in deep learning due to its robust performance. And each time, a small batch of samples is selected to update the neural network parameters $\mathbf{w}$ (Goodfellow I et al., 2016), and the iteration of all samples is called an epoch. Furthermore, we utilized the Glorot Uniform method to establish an optimal initial point for our neural network parameter (Glorot X & Bengio Y, 2010).

Batch size and learning rate are key optimization hyperparameters affecting model convergence. A smaller batch size enhances model generalizability but lengthens training time, hence a batch size of 64 is chosen for balance (Keskar NS et al., 2016). In gradient descent process, the learning rate significantly influences the step size, where large rates may induce instability or divergence, and conversely, small rates may precipitate premature convergence to local optima. To counteract this, a decaying learning rate is used, allowing larger early updates for faster convergence and preventing instability in later stages. Further, training of the model stops if it has been trained for over 200 epochs or if the learning rate has been reduced five times. These criteria are adaptive and efficient as further training does not significantly enhance the model's performance.

Hyperparameter tuning of DeepTXsolver

To obtain a neuron network model with accurate prediction and generalization, we compared the model architectures in terms of the number of neurons per layer, the number of hidden layers, and the number of neurons in the output layer, and compared the performance of the model with different dataset sizes. And we applied the Hellinger distance between the true distribution and the predicted distribution by the neuron network as a judgment criterion.

First, we compared the performance of architectures with a single hidden layer with (16, 32, 64, 128, 256, 512) neurons, and found that the performance of the model barely increased after increasing to 128 neurons (Supplementary Figure S2a). Secondly, after comparing the performance of architectures with different numbers of hidden layers, we found that a single hidden layer has the best performance (Supplementary Figure S2b). Thirdly, comparing the performance of $3i$ ($i = 1, 2, \dots, 10$) neurons in the output layer, we found that after 12 neurons (4 negative binomial distribution mixtures), the performance of the model was not improved

(Supplementary Figure S2c). Particularly, four negative binomial distribution mixtures are able to fit the bimodal nature of the distribution as well as ensure good training efficiency (Öcal K et al., 2022 [DOI](#)).

Finally, to obtain a dataset that contributes to convergence as well as model generalization while ensuring training efficiency, we compared the model performance for dataset sizes of (500, 1000, 5000, 10000, 15000) and found that increasing the dataset size further does not significantly improve the model performance after to 15000 samples (Supplementary Figure S2d). Therefore, we identified a single hidden layer of 128 neurons and an output layer of 12 neurons as the appropriate model architecture, and 15,000 samples as the appropriate training set.

1.4 Inferring burst kinetics from scRNA-seq data with DeepTXinferer

This section describes how to recover dynamic burst kinetics from static scRNA-seq data. In principle, once we have constructed the mapping of the parameters to the corresponding stationary distributions through the pre-trained neural network in section 1.3, we can quickly access the corresponding likelihood functions when updating the parameters during the statistical inference process. However, true scRNA-seq data is the result of a combination of gene expression processes and sequencing processes, and the current pre-trained neural network can only yield a description of the gene expression process modeled by TXmodel. Therefore, we require to extent it into a hierarchical model to depict the more realistic generation mechanism behind the scRNA-seq data (Sarkar A & Stephens M, 2021 [DOI](#)).

Modeling observed scRNA-seq data with hierarchical model

The hierarchical model aims to couple two different processes, specifically, the true gene expression mechanism process and the sequencing process that immediately follows. In terms of a formula, the probability distribution of observations can be represented as a convolution:

$$P(Y = y) = \int_0^\infty P_{\text{measure}}(y|x) P_{\text{model}}(x) dx, \quad (8)$$

Where P_{measure} is the measure model and P_{model} is the expression model.

First, we model the sequencing process. Each cell can be regarded as a pool of mRNA, and $X_{cg}^{(\text{true})}$ represents the true expression amount of the g -th gene in the c -th cell. $Y_{cg}^{(\text{observed})}$ refers to the observed value obtained by scRNA-seq from the given true expression value $X_{cg}^{(\text{true})}$, satisfying the following conditional probability distribution

$$Y_{cg}^{(\text{observed})} | X_{cg}^{(\text{true})} \sim P_{\text{measure}}(y|x), \quad (9)$$

Where $P_{\text{measure}}(y|x)$ characterizes all noise in the sequencing process, and is generally assumed to be Binomial distribution or Poisson distribution, which has been proven experimentally and theoretically (Sarkar A & Stephens M, 2021 [DOI](#); Wang J et al., 2018 [DOI](#)). By incorporating an additional sampling probability, denoted as α_{cg} , into the sequencing process, we are able to characterize capture efficiency. We make the assumption that intercellular molecules are independent of one another, and that only proportional products are captured and sequenced, following a Binomial distribution

$$Y_{cg}^{(\text{observed})} | X_{cg}^{(\text{true})} \sim P_{\text{measure}}(y|x, \alpha_{cg}), \quad (10)$$

Where α_{cg} is the sampling probability. Although the experimental capture efficiency ranges from 0.06 to 0.32 (Zheng GX et al., 2017 [DOI](#); Macosko EZ et al., 2015 [DOI](#)), we fixed the parameter $\alpha_{cg} = 0.5$ to minimize complexity and unidentifiability issues. Further, we demonstrated that inference across different efficiencies (0.5, 0.3, 0.2) consistently yielded strong correlations between inferred burst size and burst frequency (Supplementary Figure S12).

Second, we model the true gene expression process, assuming that it satisfies the following probability distribution

$$X_{cg}^{(\text{true})} \sim P_{\text{model}}(x). \quad (11)$$

The selection of the true gene expression probability distribution is very critical. It needs to meet the following conditions: (i) It can well characterize the transcription burst kinetics. (ii) The inference derived from this distribution necessitates both efficiency and precision.

In section 1.2, we have constructed a model of the gene expression process with DNA damage, enabling the model's stationary distribution to encapsulate the bursty dynamics inherent in gene expression. The solution to the model is approximated by a mixed negative binomial distribution, which is obtained from a trained neural network. Importantly, this mixed negative binomial distribution enables efficient inference. Hence,

$$P_{\text{model}}(x) = P_{\text{neuralnet}}(x, \theta). \quad (12)$$

Further, the following hierarchical model can be obtained

$$\begin{aligned} Y_{\text{cg}}^{(\text{observed})} | X_{\text{cg}}^{(\text{true})} &\sim \text{Binomial}(Y_{\text{cg}} | X_{\text{cg}}, \alpha_{\text{cg}}), \\ X_{\text{cg}}^{(\text{true})} &\sim \sum_{j=1}^{n_{\text{NB}}} a_j \text{NB}_j(X_{\text{cg}}, m_j(\theta), l_j(\theta)). \end{aligned} \quad (13)$$

Where $X_{\text{cg}}^{(\text{true})}$ represents the random variable of the true expression value, $Y_{\text{cg}}^{(\text{observed})}$ represents the random variable of the observed expression value, Binomial represents the binomial distribution, NB represents negative binomial distribution and α_{cg} represents the sampling probability. Further, we can obtain

$$Y_{\text{cg}}^{(\text{observed})} \sim \sum_{n_{\text{NB}}=1}^{n_{\text{count}}} \left(\text{Binomial}(Y_{\text{cg}} | X_{\text{cg}}, \alpha_{\text{cg}}) \sum_{j=1}^{n_{\text{NB}}} a_j \text{NB}_j(X_{\text{cg}}, m_j(\theta), l_j(\theta)) \right). \quad (14)$$

It is proved that $Y_{\text{cg}}^{(\text{observed})}$ also follows a mixed negative binomial distribution (Supplementary text 1.3), denoted as P_{obsmodel} .

Estimation and optimization process of DeepTXinferer

We can use the DeepTX framework to infer scRNA-seq data to obtain the corresponding gene expression burst kinetics through optimization methods. Given the expression counts of gene g in all cells, its corresponding probability distribution P_{obsdata} can be obtained. We use the Hellinger distance to compare the error between the true probability distribution P_{obsdata} of genes and the model probability distribution $P_{\text{modeldata}}$. The following optimization problem can be obtained

$$\hat{\theta} = \arg \min_{\theta} \sum_{m=0}^{n_{\text{count}}} \left(\sqrt{P_{\text{obsmodel}}(Y_{\text{cg}} = m; \theta)} - \sqrt{P_{\text{obsdata}}(x = m)} \right)^2 \quad (15)$$

where n_{count} is the maximum value of mRNA expression, P_{obsdata} is the density probability for each gene g . Specifically, P_{obsdata} is a function that is differentiable with respect to parameter θ_g . We can use the gradient descent method to obtain the optimal parameters of the TXmodel and the corresponding distribution of the TXmodel.

Posterior distribution

The optimization method in the previous section can obtain the parameters of the mechanism model that satisfies the distribution of scRNA-seq data. In this section, we give a method to solve the confidence interval of each optimized parameter (Gaskin T et al., 2023 [\[15\]](#)).

Each iteration of the optimization process can obtain the estimated value $\hat{\theta}^{(k)}$ ($k = 1, \dots, n_{\text{iteration}}$) and loss value $J(\hat{\theta})$ corresponding to the estimated value. In particular, according to Equation (15) [\[15\]](#), the loss is calculated as follows.

$$J(\hat{\theta}^{(k)}, \theta) = J^{(k)}(\hat{P}_{\text{obsmodel}}(\hat{\theta}^{(k)}), P_{\text{obsdata}}) = \sum_{m=0}^{n_{\text{count}}} \left(\sqrt{\hat{P}_{\text{obsmodel}}(x = m; \hat{\theta}^{(k)})} - \sqrt{P_{\text{obsdata}}(x = m; \theta)} \right)^2, \quad (16)$$

The following posterior probability can be obtained

$$\pi(\hat{\theta}^{(k)} | P_{\text{obsdata}}) \sim \frac{n^{(k)} \exp(-J^{(k)}(\hat{P}_{\text{obsmodel}}, P_{\text{obsdata}}))}{\sum_{k=1}^{n_{\text{iteration}}} \exp(-J^{(k)}(\hat{P}_{\text{obsmodel}}, P_{\text{obsdata}}))} \pi^{(0)}(\hat{\theta}^{(k)}), \quad (17)$$

where $\pi^{(0)}$ is the prior, we take it as uniform density, $n^{(k)}$ represents the number of values $\hat{\theta}^{(k)}$ takes during the iteration process, and $\exp(-J(\hat{P}_{\text{mixture}}, P_{\text{simulation}}))$ represents the loss potential of each iteration. Let the posterior probability be $P_{\text{posterior}}(\hat{\theta}^{(k)})$, and the marginal distribution corresponding to each component of parameter $\hat{\theta}^{(k)}$ can be obtained

$$p(\theta_j^{(k)}) \sim \int P_{\text{posterior}}(\hat{\theta}^{(k)}) d\hat{\theta}_{-j}^{(k)}. \quad (18)$$

The subscript $-j$ means that we integrate the components of $\theta^{(k)}$ except $\theta_j^{(k)}$. In particular, to allow the iterative parameters to traverse the entire parameter space, multiple optimizations can be performed with different initial values to improve the performance of the posterior distribution.

Validation on synthetic data

We conducted an evaluation of the inference performance of the DeepTX framework on synthetic scRNA-seq data, focusing on three key aspects. (i) Model accuracy: This refers to the minimal discrepancy between the distribution corresponding to the inferred parameters and the input distribution. (ii) Parameter identifiability: This denotes the proximity of the values of the inferred parameters to the true parameters, suggesting that the model can accurately identify the true parameters. (iii) Result robustness: This implies that the outcomes of each inference do not exhibit significant variations.

We created a synthetic dataset $\{(\theta_i^{\text{syn}}, P_i^{\text{syn}})\} (i = 1, 2, \dots, 500)$ using Sobol sampling and SSA under preset parameters. The synthesized distribution P_i^{syn} is used as input to the DeepTX framework, and upon optimization, we obtained estimated parameters and their corresponding distribution $\{(\theta_i^{\text{est}}, P_i^{\text{est}})\}$. We showed that the Hellinger distance between the estimated and synthetic dataset's distribution was minimal, indicating accurate estimations (Supplementary Figure S3a). Additionally, the burst kinetics of both the synthetic data and the estimated data can be computed utilizing their respective parameters, and θ_i^{est} . We showed that a strong correlation was observed between the synthetic burst kinetics $\gamma^{\text{syn}} = (\text{BF}^{\text{syn}}, \text{BS}^{\text{syn}})$ and estimated burst kinetics $\gamma^{\text{est}} = (\text{BF}^{\text{est}}, \text{BS}^{\text{est}})$ (Supplementary Figure S3b-c), suggesting that burst kinetics are identifiable.

Subsequently, we conducted a robustness verification of the framework's ability to infer burst kinetics. The error was computed as follows:

$$\text{Error}(\gamma^{\text{syn}}, \gamma^{\text{est}}) = (\log(\text{BF}^{\text{syn}}) - \log(\text{BF}^{\text{est}}))^2 + (\log(\text{BS}^{\text{syn}}) - \log(\text{BS}^{\text{est}}))^2. \quad (19)$$

Given a set of parameters θ_i^{syn} , we drew samples of sizes (100, 500, 1000, 1500, 2000) from the probability distribution P_i^{syn} corresponding to θ_i^{syn} . For each sample size, we performed 50 repetitions, and then DeepTX framework is used to infer the distribution obtained from each sampling, setting different initial values each time. We found that the inferred results are robust, and the robustness increases with the increase in the number of samples (Supplementary Figure S3d-e).

2 Dataset

2.1 IdU and DMSO dataset

To investigate the effects of 5'-iodo-2'-deoxyuridine (IdU) on genome-wide gene expression, Desai et al. conducted scRNA-seq data on mouse embryonic stem cells, both with and without IdU treatment. Consequently, they obtained scRNA-seq data for 12481 genes of transcriptomes from 812 cells and 13780 genes of transcriptomes from 744 cells, respectively. The authors found a phenomenon that IdU drug treatment increased transcriptional noise genome-wide but did not affect transcriptional mean, further affecting cell fate decisions.

We used the DeepTX framework to infer this set of scRNA-seq data to obtain potential transcription burst kinetics, and further study the impact of changes in transcription burst kinetics on cell fate decisions. To eliminate the impact of technical noise on data inference, we

preprocess the data as follows. Firstly, We normalized the data y_{cg} using Seurat, where y_{cg} represents the expression level of the gene g in the cell c . The normalization process is that we multiplied each y by a scaling factor $S = y_{\text{normalized}} / \sum_{g=1}^{n_{\text{genes}}} y_{cg}$, where $y_{\text{normalized}} = 50000$. Particularly, the scaling factor S is related to the total expression of genes in each cell. Secondly, after normalizing the expression, we divided it into bins with a unit of 1 to obtain the bin where each expression value is located. Finally, to eliminate the impact of low-expression genes on the inference process, we filter out genes whose average gene expression is less than 1. After data processing, there are 6082 genes from 812 cells left in our IdU and 6082 genes from 744 cells left in DMSO data sets.

2.2 5FU Dataset

To study the impact of 5-fluorouracil (5FU) drug-induced DNA damage on cell fate decisions, Park SR et al. performed scRNA-seq on colon cancer cells treated with 0 μm , 10 μm , and 50 μm 5FU drug (Park SR et al., 2020 [\[10\]](#)) and obtained 8534 genes from 1673 cells, 7077 genes from 632 cells, and 6661 genes from 619 cells, respectively. The authors found that Colon cancer cells treated with different doses of 5FU exhibit different transcriptional phenotypes, and these different phenotypes correspond to different cell fate decisions.

We used the DeepTX framework to infer these three sets of scRNA-seq data to obtain potential transcriptional burst kinetics, thereby studying the impact of changes in transcriptional burst kinetics caused by different doses of 5FU treatment on cell fate decisions. To eliminate the impact of technical noise on data inference, we preprocess the data in the following steps.

First, we normalize y_{cg} by multiplying a scale factor S , where $S = y_{\text{normalized}} / \sum_{g=1}^{n_{\text{genes}}} y_{cg}$, $y_{\text{normalized}} = 25000$. Second, we divide the normalized data into intervals of unit 1 to round the data. Finally, to eliminate the impact of low-expression genes on the inference process, we filter out genes whose average gene expression is less than 1. After data processing, there are 730 genes from 593 cells left in control data set, 174 genes from 593 cells left in our 10 μm 5FU treatment data set and 187 genes from 564 cells left in 50 μm 5FU treatment data set.

Data availability

All the scRNA-seq data can be obtained from the public database. The scRNA-seq data of mouse embryonic stem cells (mESC) + DMSO and mESC + IdU data reported here is available at the Gene Expression Omnibus (GEO) database under accession number: *GSE176044* [\[11\]](#). The scRNA-seq data of RKO cell with different 5-fluorouracil treatment is available at the Gene Expression Omnibus (GEO) database under accession number: *GSE149224* [\[12\]](#). The code for reproducing the presented analysis results is available at GitHub repository (<https://github.com/cellfateTX/DeepTX> [\[13\]](#)).

Additional information

Funding

This work was supported by National Key R&D Program of China grant 2021YFA1302500 by Natural Science Foundation of P.R. China grants 12171494, 11931019, 11775314, 62373384, and 12301646 by Key-Area Research and Development Program of Guangzhou, P.R. China, grants 202007030004 and 2019B110233002 by Guangdong Basic and Applied Basic Research Foundation grants 2022A1515011540 and 2023A1515011982 by Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University grant 2020B1212060032 by the Fundamental Research Funds for the Central Universities, Sun Yat sen University grants 23qnp48 and 23pnp49 and by China Postdoctoral Science Foundation grant 2023M734061.

Author contributions

J.Z., Q.N., and B.J. conceived the project. J.Z., Z.H., and S.L. designed the algorithm. Z.H. and S.L. implemented the code, performed the simulations, analyzed the data, and interpreted the results. Z.W. and Z.Z. analyzed the data and interpreted the results. Q.N. and B.J. supported the data

interpretation. Z.H., S.L., and J.Z. wrote the paper. All authors revised and approved the paper. J.Z. supervised the project.

Funding

Funder	Grant reference number	Author
MOST National Key Research and Development Program of China (NKPs)	2021YFA1302500	Jiajun Zhang
Natural Science Foundation of P.R. China	12171494	Jiajun Zhang

Author ORCID iDs

Jiajun Zhang: <https://orcid.org/0000-0001-7107-4814>

Additional files

[Supplemental File](#) 

References

- López-Maury L., Marguerat S., Bähler J. (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics* **9**:583-593 <https://doi.org/10.1038/nrg2398> | [PubMed](#)
- Stadhouders R., Filion G.J., Graf T. (2019) Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**:345-354 <https://doi.org/10.1038/s41586-019-1182-7> | [PubMed](#)
- Raj A., van Oudenaarden A. (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**:216-226 <https://doi.org/10.1016/j.cell.2008.09.050> | [PubMed](#)
- Balázsi G., van Oudenaarden A., Collins J.J. (2011) Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**:910-925 <https://doi.org/10.1016/j.cell.2011.01.030> | [PubMed](#)
- Losick R., Desplan C. (2008) Stochasticity and cell fate. *Science* **320**:65-68 <https://doi.org/10.1126/science.1147888> | [PubMed](#)
- Eling N., Morgan M.D., Marioni J.C. (2019) Challenges in measuring and understanding biological noise. *Nature Reviews Genetics* **20**:536-548 <https://doi.org/10.1038/s41576-019-0130-6> | [PubMed](#)
- Tunnacliffe E., Chubb J.R. (2020) What is a transcriptional burst?. *Trends in Genetics* **36**:288-297 <https://doi.org/10.1016/j.tig.2020.01.003> | [PubMed](#)
- Golding I., Paulsson J., Zawilski S.M., Cox E.C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* **123**:1025-1036 <https://doi.org/10.1016/j.cell.2005.09.031> | [PubMed](#)
- Chubb J.R., Trcek T., Shenoy S.M., Singer R.H. (2006) Transcriptional pulsing of a developmental gene. *Current Biology* **16**:1018-1025 <https://doi.org/10.1016/j.cub.2006.03.092> | [PubMed](#)
- Larson D.R., Zenklusen D., Wu B., Chao J.A., Singer R.H. (2011) Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332**:475-478 <https://doi.org/10.1126/science.1202142> | [PubMed](#)
- Lammers N.C., Kim Y.J., Zhao J., Garcia H.G. (2020) A matter of time: using dynamics and theory to uncover mechanisms of transcriptional bursting. *Current Opinion in Cell Biology* **67**:147-157 <https://doi.org/10.1016/j.ceb.2020.08.001> | [PubMed](#)
- Rodriguez J., Larson D.R. (2020) Transcription in living Cells: molecular mechanisms of bursting. *Annual Review of Biochemistry* **89**:189-212 <https://doi.org/10.1146/annurev-biochem-011520-105250> | [PubMed](#)
- Krenning L., van den Berg J., Medema R.H. (2019) Life or death after a break: what determines the choice?. *Molecular Cell* **76**:346-358 <https://doi.org/10.1016/j.molcel.2019.08.023> | [PubMed](#)

- Su T.T. (2006) Cellular responses to DNA damage: one signal, multiple choices. *Annual Review of Genetics* **40**:187-208 <https://doi.org/10.1146/annurev.genet.40.110405.090428> | PubMed
- Hafner A., Bulyk M.L., Jambhekar A., Lahav G. (2019) The multiple mechanisms that regulate p53 activity and cell fate. *Nature Reviews Molecular Cell Biology* **20**:199-210 <https://doi.org/10.1038/s41580-019-0110-x> | PubMed
- van den Berg J., Joosten S.E., Kim Y., Manjón A.G., Krenning L., Koob L., Feringa F.M., Klomp maker R., van den Broek B., Jalink K. (2019) DNA end-resection in highly accessible chromatin produces a toxic break. *arXiv* <https://doi.org/10.1101/691857>
- Arora M., Moser J., Phadke H., Basha A.A., Spencer S.L. (2017) Endogenous replication stress in mother cells leads to 1uiescence of daughter cells. *Cell Reports* **19**:1351-1364 <https://doi.org/10.1016/j.celrep.2017.04.055> | PubMed
- Barr A.R., Cooper S., Heldt F.S., Butera F., Stoy H., Mansfeld J., Novák B., Bakal C. (2017) DNA damage during S-phase mediates the proliferation-quiescence decision in the subsequent G1 via p21 expression. *Nature Communications* **8**:14728 <https://doi.org/10.1038/ncomms14728> | PubMed
- Deng Z., Shen D., Yu M., Zhou F., Shan D., Fang Y., Jin W., Qian K., Li S., Wang G., *et al.* (2023) Pectolarigenin inhibits bladder urothelial carcinoma cell proliferation by regulating DNA damage/autophagy pathways. *Cell Death Discovery* **9**:214 <https://doi.org/10.1038/s41420-023-01508-9> | PubMed
- Müllers E., Silva Cascales H., Jaiswal H., Saurin A.T., Lindqvist A. (2014) Nuclear translocation of Cyclin B1 marks the restriction point for terminal cell cycle exit in G2 phase. *Cell Cycle* **13**:2733-2743 <https://doi.org/10.4161/15384101.2015.945831> | PubMed
- Feringa F.M., Raaijmakers J.A., Hadders M.A., Vaarting C., Macurek L., Heitink L., Krenning L., Medema R.H. (2018) Persistent repair intermediates induce senescence. *Nature Communications* **9**:3923 <https://doi.org/10.1038/s41467-018-06308-9> | PubMed
- Toledo L.I., Murga M., Gutierrez-Martinez P., Soria R., Fernandez-Capetillo O. (2008) ATR signaling can drive cells into senescence in the absence of DNA breaks. *Genes & Development* **22**:297-302 <https://doi.org/10.1101/gad.452308> | PubMed
- Zhao Y., Simon M., Seluanov A., Gorbunova V. (2023) DNA damage and repair in age-related inflammation. *Nature Reviews Immunology* **23**:75-89 <https://doi.org/10.1038/s41577-022-00751-y> | PubMed
- Yousefzadeh M., Henpita C., Vyas R., Soto-Palma C., Robbins P., Niedernhofer L. (2021) DNA damage-how and why we age?. *eLife* **10**:e62852 <https://doi.org/10.7554/eLife.62852> | PubMed
- Carneiro B.A., El-Deiry W.S. (2020) Targeting apoptosis in cancer therapy. *Nature Reviews Clinical Oncology* **17**:395-417 <https://doi.org/10.1038/s41571-020-0341-y> | PubMed
- Roos W.P., Kaina B. (2013) DNA damage-induced cell death: from specific DNA lesions to the DNA damage response and apoptosis. *Cancer Letters* **332**:237-248 <https://doi.org/10.1016/j.canlet.2012.01.007> | PubMed
- Zheng P., Chen Q., Tian X., Qian N., Chai P., Liu B., Hu J., Blackstone C., Zhu D., Teng J., *et al.* (2018) DNA damage triggers tubular endoplasmic reticulum extension to promote apoptosis by facilitating ER-mitochondria signaling. *Cell Research* **28**:833-854 <https://doi.org/10.1038/s41422-018-0065-z> | PubMed
- Lans H., Hoeijmakers J.H.J., Vermeulen W., Marteijn J.A. (2019) The DNA damage response to transcription stress. *Nature Reviews Molecular Cell Biology* **20**:766-784 <https://doi.org/10.1038/s41580-019-0169-4> | PubMed
- Muñoz M.J., Pérez Santangelo M.S., Paronetto M.P., de la Mata M., Pelisch F., Boireau S., Glover-Cutter K., Ben-Dov C., Blaustein M., Lozano J.J., *et al.* (2009) DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* **137**:708-720 <https://doi.org/10.1016/j.cell.2009.03.010> | PubMed

- Gregersen L.H., Svejstrup J.Q. (2018) The cellular response to transcription-blocking DNA damage. *Trends in Biochemical Sciences* **43**:327-341 <https://doi.org/10.1016/j.tibs.2018.02.010> | PubMed
- Geijer M.E., Marteijn J.A. (2018) What happens at the lesion does not stay at the lesion: transcription-coupled nucleotide excision repair and the effects of DNA damage on transcription in cis and trans. *DNA Repair* **71**:56-68 <https://doi.org/10.1016/j.dnarep.2018.08.007> | PubMed
- Giono L.E., Nieto Moreno N., Cambindo Botto A.E., Dujardin G., Muñoz M.J., Kornblihtt A.R. (2016) The RNA response to DNA damage. *Journal of Molecular Biology* **428**:2636-2651 <https://doi.org/10.1016/j.jmb.2016.03.004> | PubMed
- Friedrich D., Friedel L., Finzel A., Herrmann A., Preibisch S., Loewer A. (2019) Stochastic transcription in the p53-mediated response to DNA damage is modulated by burst frequency. *Molecular Systems Biology* **15**:e9068 <https://doi.org/10.15252/msb.20199068> | PubMed
- Calia G.P., Chen X., Zuckerman B., Weinberger L. (2023) Comparative analysis between RNA-seq and single-molecule RNA FISH indicates that the pyrimidine nucleobase idoxuridine (IdU) globally amplifies transcriptional noise. *bioRxiv* <https://doi.org/10.1101/2023.03.14.532632> | PubMed
- Desai R.V., Chen X., Martin B., Chaturvedi S., Hwang D.W., Li W., Yu C., Ding S., Thomson M., Singer R.H., et al. (2021) A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. *Science* **373** <https://doi.org/10.1126/science.abc6506> | PubMed
- Femino A.M., Fay F.S., Fogarty K., Singer R.H. (1998) Visualization of single RNA transcripts in situ. *Science* **280**:585-590 <https://doi.org/10.1126/science.280.5363.585> | PubMed
- Qiu Q., Hu P., Qiu X., Govek K.W., Cámara P.G., Wu H. (2020) Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. *Nature Methods* **17**:991-1001 <https://doi.org/10.1038/s41592-020-0935-4> | PubMed
- Safieddine A., Coleno E., Lionneton F., Traboulsi A.M., Salloum S., Lecellier C.H., Gostan T., Georget V., Hassen-Khodja C., Imbert A., et al. (2023) HT-smFISH: a cost-effective and flexible workflow for high-throughput single-molecule RNA imaging. *Nature Protocols* **18**:157-187 <https://doi.org/10.1038/s41596-022-00750-2> | PubMed
- Tanay A., Regev A. (2017) Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**:331-338 <https://doi.org/10.1038/nature21350> | PubMed
- Picelli S., Björklund Å K., Faridani O.R., Sagasser S., Winberg G., Sandberg R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods* **10**:1096-1098 <https://doi.org/10.1038/nmeth.2639> | PubMed
- Zheng G.X., Terry J.M., Belgrader P., Ryvkin P., Bent Z.W., Wilson R., Ziraldo S.B., Wheeler T.D., McDermott G.P., Zhu J., et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**:14049 <https://doi.org/10.1038/ncomms14049> | PubMed
- Faure A.J., Schmiedel J.M., Lehner B. (2017) Systematic analysis of the determinants of gene expression noise in embryonic stem cells. *Cell Systems* **5**:471-484.e474. <https://doi.org/10.1016/j.cels.2017.10.003> | PubMed
- Morgan M.D., Marioni J.C. (2018) CpG island composition differences are a source of gene expression noise indicative of promoter responsiveness. *Genome Biology* **19**:81 <https://doi.org/10.1186/s13059-018-1461-x> | PubMed
- Ochiai H., Hayashi T., Umeda M., Yoshimura M., Harada A., Shimizu Y., Nakano K., Saitoh N., Liu Z., Yamamoto T., et al. (2020) Genome-wide kinetic properties of transcriptional bursting in mouse embryonic stem cells. *Science Advances* **6**:eaaz6699 <https://doi.org/10.1126/sciadv.aaz6699> | PubMed
- Larsson A.J.M., Johnsson P., Hagemann-Jensen M., Hartmanis L., Faridani O.R., Reinius B., Segerstolpe A., Rivera C.M., Ren B., Sandberg R. (2019) Genomic encoding of transcriptional burst kinetics. *Nature* **565**:251-254 <https://doi.org/10.1038/s41586-018-0836-1> | PubMed

- Luo S., Wang Z., Zhang Z., Zhou T., Zhang J. (2023) Genome-wide inference reveals that feedback regulations constrain promoter-dependent transcriptional burst kinetics. *Nucleic Acids Research* **51**:68-83 <https://doi.org/10.1093/nar/gkac1204> | PubMed
- Luo S., Zhang Z., Wang Z., Yang X., Chen X., Zhou T., Zhang J. (2023) Inferring transcriptional bursting kinetics from single-cell snapshot data using a generalized telegraph model. *Royal Society Open Science* **10**:221057 <https://doi.org/10.1098/rsos.221057> | PubMed
- Stumpf P.S., Smith R.C., Lenz M., Schuppert A., Müller F.-J., Babbie A., Chan T.E., Stumpf M.P., Please C.P., Howison S.D. (2017) Stem cell differentiation as a non-Markov stochastic process. *Cell Systems* **5**:268-282.e7 <https://doi.org/10.1016/j.cels.2017.08.009> | PubMed
- Voss T.C., Hager G.L. (2014) Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Reviews Genetics* **15**:69-81 <https://doi.org/10.1038/nrg3623> | PubMed
- Singh A., Vargas C.A., Karmakar R. (2013) Stochastic analysis and inference of a two-state genetic promoter model. In: Proceedings of the American Control Conference. pp. 4563-4568
- Cavallaro M., Walsh M.D., Jones M., Teahan J., Tiberi S., Finkenstadt B., Hebenstreit D. (2021) 3' (')-5' (') crosstalk contributes to transcriptional bursting. *Genome Biology* **22**:56 <https://doi.org/10.1186/s13059-020-02227-5> | PubMed
- Zhang J., Chen L., Zhou T. (2012) Analytical distribution and tunability of noise in a model of promoter progress. *Biophysical Journal* **102**:1247-1257 <https://doi.org/10.1016/j.bpj.2012.02.001> | PubMed
- Zhang J., Zhou T. (2014) Promoter-mediated transcriptional dynamics. *Biophysical Journal* **106**:479-488 <https://doi.org/10.1016/j.bpj.2013.12.011> | PubMed
- Zhou T., Zhang J. (2012) Analytical results for a multistate gene model. *SIAM Journal On Applied Mathematics* **72**:789-818 <https://doi.org/10.1137/110852887>
- Rodriguez J., Ren G., Day C.R., Zhao K., Chow C.C., Larson D.R. (2019) Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell* **176**:213-226.e218. <https://doi.org/10.1016/j.cell.2018.11.026> | PubMed
- Zoller B., Nicolas D., Molina N., Naef F. (2015) Structure of silent transcription intervals and noise characteristics of mammalian genes. *Molecular Systems Biology* **11**:823 <https://doi.org/10.15252/msb.20156257> | PubMed
- Daigle B.J., Soltani M., Petzold L.R., Singh A. (2015) Inferring single-cell gene expression mechanisms using stochastic simulation. *Bioinformatics* **31**:1428-1435 <https://doi.org/10.1093/bioinformatics/btv007> | PubMed
- Kumar N., Singh A., Kulkarni R.V. (2015) Transcriptional bursting in gene expression: analytical results for general stochastic models. *PLoS Computational Biology* **11**:e1004292 <https://doi.org/10.1371/journal.pcbi.1004292> | PubMed
- Schwabe A., Rybakova K.N., Bruggeman F.J. (2012) Transcription stochasticity of complex gene regulation models. *Biophysical Journal* **103**:1152-1161 <https://doi.org/10.1016/j.bpj.2012.07.011> | PubMed
- Stinchcombe A.R., Peskin C.S., Tranchina D. (2012) Population density approach for discrete mRNA distributions in generalized switching models for stochastic gene expression. *Physical Review E: Statistical Nonlinear and Soft Matter Physics* **85**:061919 <https://doi.org/10.1103/PhysRevE.85.061919> | PubMed
- Zhang J., Zhou T. (2019) Markovian approaches to modeling intracellular reaction processes with molecular memory. *Proceedings of the National Academy of Sciences* **116**:23542-23550 <https://doi.org/10.1073/pnas.1913926116> | PubMed
- Zhang J.J., Zhou T.S. (2019) Stationary moments, distribution conjugation and phenotypic regions in stochastic gene transcription. *Mathematical Biosciences and Engineering* **16**:6134-6166 <https://doi.org/10.3934/mbe.2019307> | PubMed
- Sisson S.A., Fan Y., Tanaka M.M. (2007) Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences* **104**:1760-1765 <https://doi.org/10.1073/pnas.0607208104> | PubMed

- Gómez-Schiavon M., Chen L.F., West A.E., Buchler N.E. (2017) BayFish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells. *Genome Biology* **18**:164 <https://doi.org/10.1186/s13059-017-1297-9> | PubMed
- Jiang Q., Fu X., Yan S., Li R., Du W., Cao Z., Qian F., Grima R. (2021) Neural network aided approximation and parameter inference of non-Markovian models of gene expression. *Nature Communications* **12**:2618 <https://doi.org/10.1038/s41467-021-22919-1> | PubMed
- Wang S., Fan K., Luo N., Cao Y., Wu F., Zhang C., Heller K.A., You L. (2019) Massive computational acceleration by using neural networks to emulate mechanism-based biological models. *Nature Communications* **10**:4354 <https://doi.org/10.1038/s41467-019-12342-y> | PubMed
- Michoski C., Milosavljević M., Oliver T., Hatch D.R. (2020) Solving differential equations using deep neural networks. *Neurocomputing* **399**:193-212 <https://doi.org/10.1016/j.neucom.2020.02.015>
- Davis C.N., Hollingsworth T.D., Caudron Q., Irvine M.A. (2020) The use of mixture density networks in the emulation of complex epidemiological individual-based models. *PLoS Computational Biology* **16**:e1006869 <https://doi.org/10.1371/journal.pcbi.1006869>
- Tang Y., Weng J., Zhang P. (2023) Neural-network solutions to stochastic reaction networks. *Nature Machine Intelligence* **5**:376-385 <https://doi.org/10.1038/s42256-023-00632-6>
- Gaskin T., Pavliotis G.A., Girolami M. (2023) Neural parameter calibration for large-scale multiagent models. *Proceedings of the National Academy of Sciences* **120**:e2216415120 <https://doi.org/10.1073/pnas.2216415120> | PubMed
- Sukys A., Öcal K., Grima R. (2022) Approximating solutions of the chemical master equation using neural networks. *iScience* **25**:105010 <https://doi.org/10.1016/j.isci.2022.105010> | PubMed
- Tang W., Jørgensen A.C.S., Marguerat S., Thomas P., Shahrezaei V. (2023) Modelling capture efficiency of single cell RNA-sequencing data improves inference of transcriptome-wide burst kinetics. *bioRxiv* <https://doi.org/10.1101/2023.03.06.531327>
- Gupta A., Schwab C., Khammash M. (2021) DeepCME: a deep learning framework for computing solution statistics of the chemical master equation. *PLoS Computational Biology* **17**:e1009623 <https://doi.org/10.1371/journal.pcbi.1009623> | PubMed
- Perez-Carrasco R., Beentjes C., Grima R. (2020) Effects of cell cycle variability on lineage and population measurements of messenger RNA abundance. *Journal of the Royal Society Interface* **17**:20200360 <https://doi.org/10.1098/rsif.2020.0360> | PubMed
- Öcal K., Gutmann M.U., Sanguinetti G., Grima R. (2022) Inference and uncertainty quantification of stochastic gene expression via synthetic models. *Journal of the Royal Society Interface* **19**:20220153 <https://doi.org/10.1098/rsif.2022.0153> | PubMed
- Sobol I.M. (1967) On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput Math Math Phys* **7**:784-802 [https://doi.org/10.1016/0041-5553\(67\)90144-9](https://doi.org/10.1016/0041-5553(67)90144-9)
- Zhang Z., Deng Q., Wang Z., Chen Y., Zhou T. (2021) Exact results for queuing models of stochastic transcription with memory and crosstalk. *Physical Review E* **103**:062414 <https://doi.org/10.1103/PhysRevE.103.062414> | PubMed
- Zhang J., Chen A., Qiu H., Zhang J., Tian T., Zhou T. (2024) Exact results for gene-expression models with general waiting-time distributions. *Physical Review E* **109**:024119 <https://doi.org/10.1103/PhysRevE.109.024119> | PubMed
- Gupta P.B., Fillmore C.M., Jiang G., Shapira S.D., Tao K., Kuperwasser C., Lander E.S. (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**:633-644 <https://doi.org/10.1016/j.cell.2011.07.026> | PubMed
- Cohen A.A., Geva-Zatorsky N., Eden E., Frenkel-Morgenstern M., Issaeva I., Sigal A., Milo R., Cohen-Saidon C., Liron Y., Kam Z., et al. (2008) Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322**:1511-1516 <https://doi.org/10.1126/science.1160165> | PubMed
- Bessarabova M., Kirillov E., Shi W., Bugrim A., Nikolsky Y., Nikolskaya T. (2010) Bimodal gene expression patterns in breast cancer. *BMC Genomics* **11** <https://doi.org/10.1186/1471-2164-11-s1-s8> | PubMed

- Sarkar A., Stephens M. (2021) Separating measurement and expression models clarifies confusion in single-cell RNA sequencing analysis. *Nature Genetics* **53**:770-777 <https://doi.org/10.1038/s41588-021-00873-4> | PubMed
- Das S., Suganthan P.N. (2010) Differential evolution: A survey of the state-of-the-art. *IEEE transactions on evolutionary computation* **15**:4-31 <https://doi.org/10.1109/TEVC.2010.2059031>
- Tang W., Jørgensen A.C.S., Marguerat S., Thomas P., Shahrezaei V. (2023) Modelling capture efficiency of single-cell RNA-sequencing data improves inference of transcriptome-wide burst kinetics. *Bioinformatics* **39** <https://doi.org/10.1093/bioinformatics/btad395> | PubMed
- Gu J., Laszik N., Miles C.E., Allard J., Downing T.L., Read E.L. (2025) Scalable inference and identifiability of kinetic parameters for transcriptional bursting from single cell data. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btaf581> | PubMed
- Li C., Xu X., Chen S., Xu A., Guan T., Wu H., Pei D., Liu J. (2024) Epigenetic reshaping through damage: promoting cell fate transition by BrdU and IdU incorporation. *Cell and Bioscience* **14**:9 <https://doi.org/10.1186/s13578-024-01192-x> | PubMed
- Newman J.R., Ghaemmaghami S., Ihmels J., Breslow D.K., Noble M., DeRisi J.L., Weissman J.S. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**:840-846 <https://doi.org/10.1038/nature04785> | PubMed
- Gene Ontology Consortium (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Research* **43**:D1049-1056 <https://doi.org/10.1093/nar/gku1179> | PubMed
- Smits V.A., Medema R.H. (2001) Checking out the G2/M transition. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* **1519**:1-12
- Subramanian A., Tamayo P., Mootha V.K., Mukherjee S., Ebert B.L., Gillette M.A., Paulovich A., Pomeroy S.L., Golub T.R., Lander E.S., et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**:15545-15550 <https://doi.org/10.1073/pnas.0506580102> | PubMed
- Rosina M., Langone F., Giuliani G., Cerquone Perpetuini A., Reggio A., Calderone A., Fuoco C., Castagnoli L., Gargioli C., Cesareni G. (2019) Osteogenic differentiation of skeletal muscle progenitor cells is activated by the DNA damage response. *Scientific Reports* **9**:5447 <https://doi.org/10.1038/s41598-019-41926-3> | PubMed
- Riccio F., Micarelli E., Secci R., Giuliani G., Vumbaca S., Massacci G., Castagnoli L., Fuoco C., Cesareni G. (2022) Transcription Factor Activation Profiles (TFAP) identify compounds promoting differentiation of Acute Myeloid Leukemia cell lines. *Cell Death Discovery* **8**:16 <https://doi.org/10.1038/s41420-021-00811-7> | PubMed
- Liu Z., Wang L., Welch J.D., Ma H., Zhou Y., Vaseghi H.R., Yu S., Wall J.B., Alimohamadi S., Zheng M., et al. (2017) Single-cell transcriptomics reconstructs fate conversion from fibroblast to cardiomyocyte. *Nature* **551**:100-104 <https://doi.org/10.1038/nature24454> | PubMed
- Kuipers E.J., Grady W.M., Lieberman D., Seufferlein T., Sung J.J., Boelens P.G., van de Velde C.J., Watanabe T. (2015) Colorectal cancer. *Nature Reviews Disease Primers* **1**:15065 <https://doi.org/10.1038/nrdp.2015.65> | PubMed
- Bunz F., Hwang P.M., Torrance C., Waldman T., Zhang Y., Dillehay L., Williams J., Lengauer C., Kinzler K.W., Vogelstein B. (1999) Disruption of p53 in human cancer cells alters the responses to therapeutic agents. *Journal of Clinical Investigation* **104**:263-269 <https://doi.org/10.1172/jci6863> | PubMed
- Chang G.S., Chen X.A., Park B., Rhee H.S., Li P., Han K.H., Mishra T., Chan-Salis K.Y., Li Y., Hardison R.C., et al. (2014) A comprehensive and high-resolution genome-wide response of p53 to stress. *Cell Reports* **8**:514-527 <https://doi.org/10.1016/j.celrep.2014.06.030> | PubMed
- Park S.R., Namkoong S., Friesen L., Cho C.S., Zhang Z.Z., Chen Y.C., Yoon E., Kim C.H., Kwak H., Kang H.M., et al. (2020) Single-cell transcriptome analysis of colon cancer cell response to 5-fluorouracil-induced DNA damage. *Cell Reports* **32**:108077 <https://doi.org/10.1016/j.celrep.2020.108077> | PubMed

- Hongmei Z. (2012) Extrinsic and Intrinsic Apoptosis Signal Pathway Review. In: *Apoptosis and medicine* InTechOpen.
- Sun Y., Rigas B. (2008) The thioredoxin system mediates redox-induced cell death in human colon cancer cells: implications for the mechanism of action of anticancer agents. *Cancer Research* **68**:8269-8277 <https://doi.org/10.1158/0008-5472.Can-08-2010> | PubMed
- Laha D., Pramanik A., Chattopadhyay S., kumar Dash S., Roy S., Pramanik P., Karmakar P. (2015) Folic acid modified copper oxide nanoparticles for targeted delivery in in vitro and in vivo systems. *RSC Advances* **5**:68169-68178 <https://doi.org/10.1039/C5RA08110F>
- Bwatanglang I.B., Mohammad F., Yusof N.A., Abdullah J., Alitheen N.B., Hussein M.Z., Abu N., Mohammed N.E., Nordin N., Zambari N.R., et al. (2016) In vivo tumor targeting and anti-tumor effects of 5-fluorouracil loaded, folic acid targeted quantum dot system. *Journal of Colloid and Interface Science* **480**:146-158 <https://doi.org/10.1016/j.jcis.2016.07.011> | PubMed
- Chenna S., Koopman W.J.H., Prehn J.H.M., Connolly N.M.C. (2022) Mechanisms and mathematical modeling of ROS production by the mitochondrial electron transport chain. *American Journal of Physiology: Cell Physiology* **323**:C69-c83 <https://doi.org/10.1152/ajpcell.00455.2021> | PubMed
- Adam-Vizi V. (2005) Production of reactive oxygen species in brain mitochondria: contribution by electron transport chain and non-electron transport chain sources. *Antioxidants & Redox Signaling* **7**:1140-1149 <https://doi.org/10.1089/ars.2005.7.1140> | PubMed
- Shokolenko I.N., Wilson G.L., Alexeyev M.F. (2014) Aging: a mitochondrial DNA perspective, critical analysis and an update. *World Journal of Experimental Medicine* **4**:46-57 <https://doi.org/10.5493/wjem.v4.i4.46> | PubMed
- Handali S., Moghimipour E., Rezaei M., Ramezani Z., Kouchak M., Amini M., Angali K.A., Saremy S., Dorkoosh F.A. (2018) A novel 5-Fluorouracil targeted delivery to colon cancer using folic acid conjugated liposomes. *Biomedicine & Pharmacotherapy* **108**:1259-1273 <https://doi.org/10.1016/j.biopha.2018.09.128> | PubMed
- Hwang P.M., Bunz F., Yu J., Rago C., Chan T.A., Murphy M.P., Kelso G.F., Smith R.A., Kinzler K.W., Vogelstein B. (2001) Ferredoxin reductase affects p53-dependent, 5-fluorouracil-induced apoptosis in colorectal cancer cells. *Nature Medicine* **7**:1111-1117 <https://doi.org/10.1038/nm1001-1111> | PubMed
- Kuranaga N., Shinomiya N., Mochizuki H. (2001) Long-term cultivation of colorectal carcinoma cells with anticancer drugs induces drug resistance and telomere elongation: an in vitro study. *BMC Cancer* **1**:10 <https://doi.org/10.1186/1471-2407-1-10> | PubMed
- Nabetani A., Ishikawa F. (2011) Alternative lengthening of telomeres pathway: recombination-mediated telomere maintenance mechanism in human cells. *J Biochem* **149**:5-14 <https://doi.org/10.1093/jb/mvq119> | PubMed
- Cristofari G., Adolf E., Reichenbach P., Sikora K., Terns R.M., Terns M.P., Lingner J. (2007) Human telomerase RNA accumulation in Cajal bodies facilitates telomerase recruitment to telomeres and telomere elongation. *Molecular Cell* **27**:882-889 <https://doi.org/10.1016/j.molcel.2007.07.020> | PubMed
- Lipinska N., Romaniuk A., Paszel-Jaworska A., Toton E., Kopczynski P., Rubis B. (2017) Telomerase and drug resistance in cancer. *Cellular and Molecular Life Sciences* **74**:4121-4132 <https://doi.org/10.1007/s00018-017-2573-2> | PubMed
- Wiecek A.J., Cutty S.J., Kornai D., Parreno-Centeno M., Gourmet L.E., Tagliazucchi G.M., Jacobson D.H., Zhang P., Xiong L., Bond G.L., et al. (2023) Genomic hallmarks and therapeutic implications of G0 cell cycle arrest in cancer. *Genome Biology* **24**:128 <https://doi.org/10.1186/s13059-023-02963-4> | PubMed
- Hänzelmann S., Castelo R., Guinney J. (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**:7 <https://doi.org/10.1186/1471-2105-14-7> | PubMed
- Dar R.D., Razooki B.S., Singh A., Trimeloni T.V., McCollum J.M., Cox C.D., Simpson M.L., Weinberger L.S. (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences* **109**:17454-17459

<https://doi.org/10.1073/pnas.1213530109> | PubMed

Zenklusen D., Larson D.R., Singer R.H. (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature Structural & Molecular Biology* **15**:1263-1271

<https://doi.org/10.1038/nsmb.1514> | PubMed

Phillips R., Kondev J., Theriot J., Garcia H. (2012) *Physical Biology of the Cell* Garland Science.

Lomvardas S., Thanos D. (2002) Opening chromatin. *Molecular Cell* **9**:209-211

[https://doi.org/10.1016/s1097-2765\(02\)00463-x](https://doi.org/10.1016/s1097-2765(02)00463-x) | PubMed

Fuda N.J., Ardehali M.B., Lis J.T. (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**:186-192 <https://doi.org/10.1038/nature08449> | PubMed

Nordick B., Yu P.Y., Liao G., Hong T. (2022) Nonmodular oscillator and switch based on RNA decay drive regeneration of multimodal gene expression. *Nucleic Acids Research* **50**:3693-3708

<https://doi.org/10.1093/nar/gkac217> | PubMed

Maizels R.J. (2024) A dynamical perspective: moving towards mechanism in single-cell transcriptomics. *Philos Trans R Soc Lond B Biol Sci* **379**:20230049

<https://doi.org/10.1098/rstb.2023.0049> | PubMed

Sukys A., Grima R. (2025) Cell-cycle dependence of bursty gene expression: insights from fitting mechanistic models to single-cell RNA-seq data. *Nucleic Acids Research* **53**

<https://doi.org/10.1093/nar/gkaf295> | PubMed

Garrido-Rodriguez M., Lopez-Lopez D., Ortuno F.M., Peña-Chilet M., Muñoz E., Calzado M.A., Dopazo J. (2021) A versatile workflow to integrate RNA-seq genomic and transcriptomic data into mechanistic models of signaling pathways. *PLoS Computational Biology* **17**:e1008748

<https://doi.org/10.1371/journal.pcbi.1008748> | PubMed

Wan Y., Anastasakis D.G., Rodriguez J., Palangat M., Gudla P., Zaki G., Tandon M., Pegoraro G., Chow C.C., Hafner M., et al. (2021) Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell* **184**:2878-2895.e2820.

<https://doi.org/10.1016/j.cell.2021.04.012> | PubMed

Gorin G., Pachter L. (2022) Modeling bursty transcription and splicing with the chemical master equation. *Biophysical Journal* **121**:1056-1069 <https://doi.org/10.1016/j.bpj.2022.02.004> | PubMed

Golan-Lavi R., Giacomelli C., Fuks G., Zeisel A., Sonntag J., Sinha S., Köstler W., Wiemann S., Korf U., Yarden Y., et al. (2017) Coordinated pulses of mRNA and of protein translation or degradation produce EGF-Induced protein bursts. *Cell Reports* **18**:3129-3142

<https://doi.org/10.1016/j.celrep.2017.03.014> | PubMed

Nicolas D., Zoller B., Suter D.M., Naef F. (2018) Modulation of transcriptional burst frequency by histone acetylation. *Proceedings of the National Academy of Sciences* **115**:7153-7158

<https://doi.org/10.1073/pnas.1722330115> | PubMed

Liu T., Zhang J., Zhou T. (2016) Effect of Interaction between chromatin loops on cell-to-cell variability in gene expression. *PLoS Computational Biology* **12**:e1004917

<https://doi.org/10.1371/journal.pcbi.1004917> | PubMed

Wang Z., Zhang Z., Luo S., Zhou T., Zhang J. (2024) Power-law behavior of transcriptional bursting regulated by enhancer-promoter communication. *Genome Research* **34**:106-118

<https://doi.org/10.1101/gr.278631.123> | PubMed

Wang Z., Luo S., Zhang Z., Zhou T., Zhang J. (2023) 4D nucleome equation predicts gene expression controlled by long-range enhancer-promoter interaction. *PLoS Computational Biology* **19**:e1011722

<https://doi.org/10.1371/journal.pcbi.1011722> | PubMed

Kepler T.B., Elston T.C. (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophysical Journal* **81**:3116-3136 [https://doi.org/10.1016/s0006-3495\(01\)75949-8](https://doi.org/10.1016/s0006-3495(01)75949-8) | PubMed

- Hope B., Kosofsky B., Hyman S.E., Nestler E.J. (1992) Regulation of immediate early gene expression and AP-1 binding in the rat nucleus accumbens by chronic cocaine. *Proc Natl Acad Sci U S A* **89**:5764-5768 <https://doi.org/10.1073/pnas.89.13.5764> | PubMed
- Wen L., Tang F. (2022) Recent advances in single-cell sequencing technologies. *Precision Clinical Medicine* **5**:pbac002 <https://doi.org/10.1093/pcmedi/pbac002> | PubMed
- Miao Z., Humphreys B.D., McMahon A.P., Kim J. (2021) Multi-omics integration in the age of million singlecell data. *Nature Reviews Neurology* **17**:710-724 <https://doi.org/10.1038/s41581-021-00463-x> | PubMed
- Ramsköld D., Hendriks G.J., Larsson A.J.M., Mayr J.V., Ziegenhain C., Hagemann-Jensen M., Hartmanis L., Sandberg R. (2024) Single-cell new RNA sequencing reveals principles of transcription at the resolution of individual bursts. *Nature Cell Biology* **26**:1725-1733 <https://doi.org/10.1038/s41556-024-01486-9> | PubMed
- Gupta A., Martin-Rufino J.D., Jones T.R., Subramanian V., Qiu X., Grody E.I., Bloemendal A., Weng C., Niu S.Y., Min K.H., et al. (2022) Inferring gene regulation from stochastic transcriptional variation across single cells at steady state. *Proceedings of the National Academy of Sciences* **119**:e2207392119 <https://doi.org/10.1073/pnas.2207392119> | PubMed
- Gorin G., Carilli M., Chari T., Pachter L. (2024) Spectral neural approximations for models of transcriptional dynamics. *Biophysical Journal* **123**:2892-2901 <https://doi.org/10.1016/j.bpj.2024.04.034> | PubMed
- Kuntz J., Thomas P., Stan G.-B., Barahona M. (2021) Stationary distributions of continuous-time Markov chains: a review of theory and truncation-based approximations. *SIAM Review* **63**:3-64
- Peccoud J., Ycart B. (1995) Markovian modeling of gene-product synthesis. *Theoretical Population Biology* **48**:222-234 <https://doi.org/10.1006/tpbi.1995.1027>
- Kim J.K., Marioni J.C. (2013) Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biology* **14**:R7 <https://doi.org/10.1186/gb-2013-14-1-r7> | PubMed
- Zoller B., Little S.C., Gregor T. (2018) Diverse spatial expression patterns emerge from unified kinetics of transcriptional bursting. *Cell* **175**:835-847.e825. <https://doi.org/10.1016/j.cell.2018.09.056> | PubMed
- Hoppe C., Bowles J.R., Minchington T.G., Sutcliffe C., Upadhyai P., Rattray M., Ashe H.L. (2020) Modulation of the promoter activation rate dictates the transcriptional response to graded BMP signaling levels in the drosophila embryo. *Dev Cell* **54**:727-741.e727. <https://doi.org/10.1016/j.devcel.2020.07.007> | PubMed
- Boguñá M., Lafuerza L.F., Toral R., Serrano M. (2014) Simulating non-Markovian stochastic processes. *Physical Review E: Statistical Nonlinear and Soft Matter Physics* **90**:042108 <https://doi.org/10.1103/PhysRevE.90.042108> | PubMed
- Masuda N., Rocha L.E. (2018) A Gillespie algorithm for non-Markovian stochastic processes. *SIAM Review* **60**:95-115 <https://doi.org/10.1137/16M1055876>
- Bar-Sinai Y., Hoyer S., Hickey J., Brenner M.P. (2019) Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences* **116**:15344-15349 <https://doi.org/10.1073/pnas.1814058116> | PubMed
- Risso D., Perraudeau F., Gribkova S., Dudoit S., Vert J.P. (2018) A general and flexible method for signal extraction from single-cell RNA-seq data. *Nature Communications* **9**:284 <https://doi.org/10.1038/s41467-017-02554-5> | PubMed
- Love M.I., Huber W., Anders S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550 <https://doi.org/10.1186/s13059-014-0550-8> | PubMed
- Kingma D.P., Ba J. (2014) Adam: a method for stochastic optimization. *arXiv* <https://doi.org/10.48550/arXiv.1412.6980>
- Goodfellow I., Bengio Y., Courville A. (2016) *Deep Learning* MIT press.

Glorot X., Bengio Y. (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249-256

Keskar N.S., Mudigere D., Nocedal J., Smelyanskiy M., Tang P.T.P. (2016) On large-batch training for deep learning: generalization gap and sharp minima. *arXiv* <https://arxiv.org/abs/1609.04836>

Wang J., Huang M., Torre E., Dueck H., Shaffer S., Murray J., Raj A., Li M., Zhang N.R. (2018) Gene expression distribution deconvolution in single-cell RNA sequencing. *Proceedings of the National Academy of Sciences* **115**:E6437-E6446 <https://doi.org/10.1073/pnas.1721085115> | PubMed

Macosko E.Z., Basu A., Satija R., Nemesh J., Shekhar K., Goldman M., Tirosh I., Bialas A.R., Kamitaki N., Martersteck E.M., *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**:1202-1214 <https://doi.org/10.1016/j.cell.2015.05.002> | PubMed

Peer reviews

Joint Public Review:

In this work, the authors present DeepTX, a computational tool for studying transcriptional bursting using single-cell RNA sequencing (scRNA-seq) data and deep learning. The method aims to infer transcriptional burst dynamics-including key model parameters and the associated steady-state distributions-directly from noisy single-cell data. The authors apply DeepTX to datasets from DNA damage experiments, revealing distinct regulatory patterns: IdU treatment in mouse stem cells increases burst size, promoting differentiation, while 5FU alters burst frequency in human cancer cells, driving apoptosis or survival depending on dose. These findings underscore the role of burst regulation in mediating cell fate responses to DNA damage.

The main strength of this study lies in its methodological contribution. DeepTX integrates a non-Markovian mechanistic model with deep learning to approximate steady-state mRNA distributions as mixtures of negative binomial distributions, enabling genome-scale parameter inference with reduced computational cost. The authors provide a clear discussion of the framework's assumptions, including reliance on steady-state data and the inherent unidentifiability of parameter sets, and they outline how the model could be extended to other regulatory processes.

The revised manuscript addresses the original concerns raised by the reviewers, particularly those related to sample size requirements, distributional assumptions, and the biological interpretation of the inferred parameters. The authors have also included an extensive discussion of the limitations of the methodological framework, including the constraints associated with relying on snapshot data, as well as a broader contextualisation of DeepTX within the landscape of existing tools that link mechanistic modelling and single-cell transcriptomics.

Overall, this work represents a valuable contribution to the integration of mechanistic models with high-dimensional single-cell data. It will be of interest to researchers in systems biology, bioinformatics, and computational modelling.

Comments on revisions:

We thank the authors for their thorough revision and for carefully addressing the points raised in the previous review. At this stage, the reviewers have no further concerns.

<https://doi.org/10.7554/eLife.100623.3.sa1>

Author response:

The following is the authors' response to the previous reviews.

Joint Public Review:

In this work, the authors present DeepTX, a computational tool for studying transcriptional bursting using single-cell RNA sequencing (scRNA-seq) data and deep learning. The method aims to infer transcriptional burst dynamics-including key model parameters and the associated steady-state distributions-directly from noisy single-cell data. The authors apply DeepTX to datasets from DNA damage experiments, revealing distinct regulatory patterns: IdU treatment in mouse stem cells increases burst size, promoting differentiation, while 5FU alters burst frequency in human cancer cells, driving apoptosis or survival depending on dose. These findings underscore the role of burst regulation in mediating cell fate responses to DNA damage.

The main strength of this study lies in its methodological contribution. DeepTX integrates a non-Markovian mechanistic model with deep learning to approximate steady-state mRNA distributions as mixtures of negative binomial distributions, enabling genome-scale parameter inference with reduced computational cost. The authors provide a clear discussion of the framework's assumptions, including reliance on steady-state data and the inherent unidentifiability of parameter sets, and they outline how the model could be extended to other regulatory processes.

The revised manuscript addresses many of the original concerns, particularly regarding sample size requirements, distributional assumptions, and the biological interpretation of inferred parameters. However, the framework remains limited by the constraints of snapshot data and cannot yet resolve dynamic heterogeneity or causality. The manuscript would also benefit from a broader contextualisation of DeepTX within the landscape of existing tools linking mechanistic modelling and single-cell transcriptomics. Finally, the interpretation of pathway enrichment analyses still warrants clarification.

Overall, this work represents a valuable contribution to the integration of mechanistic models with highdimensional single-cell data. It will be of interest to researchers in systems biology, bioinformatics, and computational modelling.

Recommendations for the authors:

We thank the authors for their thorough revision and for addressing many of the points raised during the initial review. The revised manuscript presents an improved and clearer account of the methodology and its implications. However, several aspects would benefit from further clarification and refinement to strengthen the presentation and avoid overstatement.

(1) Contextualization within the existing literature

The manuscript would benefit from placing DeepTX more clearly in the context of other computational tools developed to connect mechanistic modelling and single-cell RNA sequencing data. This is an active area of research with notable recent contributions, including Sukys and Grima (bioRxiv, 2024), Garrido-Rodriguez et al. (PLOS Comp Biol, 2021), and Maizels (2024). Positioning DeepTX in relation to these and other relevant efforts would help readers appreciate its specific advances and contributions.

We sincerely thank you for this valuable suggestion. We agree that situating DeepTX within the broader landscape of computational approaches linking mechanistic modeling and single-cell RNA sequencing data will clarify its contributions and advances. In this revised

version, we have explicitly discussed the comparison and relation of DeepTX in the context of this active area using an individual paragraph in the Discussion section.

Specifically, we mentioned that the DeepTX research paradigm contributes to a growing line of area aiming to link mechanistic models of gene regulation with scRNA-seq data. Maizels provided a comprehensive review of computational strategies for incorporating dynamic mechanisms into single-cell transcriptomics (Maizels RJ, 2024). In this context, RNA velocity is one of the most important examples as it infers short-term transcriptional trends based on splicing kinetics and deterministic ODEs model. However, such approaches are limited by their deterministic assumptions and cannot fully capture the stochastic nature of gene regulation. DeepTX can be viewed as an extension of this framework to stochastic modelling, explicitly addressing transcriptional bursting kinetics under DNA damage. Similarly, DeepCycle, developed by Sukys and Grima (Sukys A & Grima R, 2025), investigates transcriptional burst kinetics during the cell cycle, employing a stochastic age-dependent model and a neural network to infer burst parameters while correcting for measurement noise. By contrast, MIGNON integrates genomic variation data and static transcriptomic measurements into a mechanistic pathway model (HiPathia) to infer pathway-level activity changes, rather than gene-level stochastic transcriptional dynamics (Garrido-Rodriguez M et al., 2021). In this sense, DeepTX and MIGNON are complementary, with DeepTX resolving burst kinetics at the single-gene level and MIGNON emphasizing pathway responses to genomic perturbations, which could inspire future extensions of DeepTX that incorporate sequence-level information.

(2) Interpretation of GO analysis

The interpretation of the GO enrichment results in Figure 4D should be revised. While the text currently associates the enriched terms with signal transduction and cell cycle G2/M phase transition, the most significant terms relate to mitotic cell cycle checkpoint signaling. This distinction should be made clear in the main text, and the conclusions drawn from the GO analysis should be aligned more closely with the statistical results.

We sincerely appreciate you for the insightful comment. We have carefully re-examined the GO enrichment results shown in Figure 4D and agree that the most significantly enriched terms correspond to mitotic cell cycle checkpoint signaling and signal transduction in response to DNA damage, rather than general G2/M phase transition processes. Accordingly, we have revised the main text to highlight the biological significance of mitotic cell cycle checkpoint signaling.

Specifically, we now emphasize two key points: DNA damage and mitotic checkpoint activation are closely interconnected. (1) The mitotic checkpoint serves as a crucial safeguard to ensure accurate chromosome segregation and maintain genomic stability under DNA damage conditions. Activation of the mitotic checkpoint can influence cell fate decisions and differentiation potential (Kim EM & Burke DJ, 2008; Lawrence KS et al., 2015). (2) Sustained activation of the spindle assembly checkpoint (SAC) has been reported to induce mitotic slippage and polyploidization, which in turn may enhance the differentiation potential of embryonic stem cells (Mantel C et al., 2007). These revisions ensure that our interpretation is consistent with the statistical enrichment results and better reflect the underlying biological processes implicated by the data.

(3) Justification for training on simulated data

The decision to train the model on simulated data should be clearly justified. While the advantage of having access to ground-truth parameters is understood, the manuscript would benefit from a discussion of the limitations of this approach, particularly in terms of generalizability to real datasets. Moreover, it is worth noting that many annotated

scRNA-seq datasets are publicly available and could, in principle, be used to complement the training strategy.

We thank you for this insightful comment. We chose to train DeepTXsolver on simulated data because no experimental dataset currently provides genome-wide transcriptional burst kinetics with known ground truth, which is essential for supervised learning. Simulation enables us to (i) generate large, fully annotated datasets spanning the biologically relevant parameter space, (ii) expose the solver to diverse bursting regimes (e.g., low/high burst frequency, small/large burst size, unimodal/bimodal distributions), and (iii) quantitatively benchmark model accuracy, parameter identifiability, and robustness prior to deployment on real scRNA-seq data.

We acknowledge, however, that simulation-based training has inherent limitations in terms of generalizability. Real biological systems may deviate from the idealized bursting model, exhibit more complex noise structures, or display parameter distributions that differ from those in simulations. Moreover, the lack of ground-truth parameters in experimental scRNA-seq datasets prevents an absolute evaluation of inference accuracy. In the future work, publicly available annotated scRNA-seq datasets could be used to complement this simulation-based training strategy and enhance generalizability. We have revised the manuscript to explicitly discuss both the rationale for using simulated data and the potential limitations of this approach.

(4) Benchmarking against external methods

The performance of DeepTX is primarily compared to a prior method from the same group. To strengthen the methodological claims, it would be preferable to include benchmarking against additional established tools from the broader literature. This would offer a more objective evaluation of the performance gains attributed to DeepTX.

We thank you for this constructive suggestion. We fully agree that benchmarking DeepTX against additional established tools from the broader literatures would provide a more comprehensive and objective evaluation of DeepTX. In the revised manuscript, we have included comparative analyses with other widely used methods, including nnRNA (From Shahrezaei group (Tang W et al., 2023)), txABC (from our group (Luo S et al., 2023)), txBurst (from Sandberg group (Larsson AJM et al., 2019)), txInfer (from Junhao group (Gu J et al., 2025)) (Supplementary Figure S4). The comparative results indicate that our method demonstrates superior performance in both efficiency and accuracy.

(5) Interpretation of Figures 4-6

The revised figures are clear and informative; however, the associated interpretations in the main text remain too strong relative to the type of analysis performed. For instance, in Figure 4, it is suggested that changes in burst size are linked to DNA damage-induced signalling cascades that affect cell cycle progression and fate decisions. While this is a plausible hypothesis, GO and GSEA analyses are correlative by nature and not sufficient to support such a mechanistic claim on their own. These analyses should be presented as exploratory, and the strength of the conclusions drawn should be tempered accordingly. Similar caution should be applied to the interpretations of Figures 5 and 6.

We thank you for this important comment. In the revised manuscript, we have carefully moderated the interpretation of the GO and GSEA results in Figures 4, 5, and 6. Specifically, we now present these analyses as exploratory and emphasize their correlative nature, avoiding causal claims that go beyond the scope of the data. The text has been rephrased to highlight the observed associations rather than implying direct causal relationships.

For Figure 4, we emphasize that while it is tempting to hypothesize that enhanced burst size may contribute to DNA damage-related checkpoint activation and thereby influence cell cycle

progression and differentiation, our current results only indicate an association between burst size enhancement and pathways involved in DNA damage response and checkpoint signaling.

For Figure 5, we emphasize that although our GO analysis cannot establish causality, the results are consistent with an association between 5-FU-induced changes in burst kinetics and pathways related to oxidative stress and apoptosis. Based on this, we propose a model outlining a potential process through which DNA damage may ultimately lead to cellular apoptosis.

For Figure 6, we emphasize that these enrichment results suggest that high-dose 5FU treatment may be associated with processes such as telomerase activation and mitochondrial function maintenance, both of which have been implicated in cell survival and apoptosis evasion in previous experimental studies. For example, prior work indicates that hTERT translocation can activate telomerase pathways to support telomere maintenance and reduce oxidative stress, which is thought to contribute to apoptosis resistance. While our enrichment analysis cannot establish causality, the observed transcriptional bursting changes are consistent with these reported survival-associated mechanisms.

(6) Discussion section framing

The initial paragraphs of the discussion section make broad biological claims about the role of transcriptional bursting in cellular decision-making. While transcriptional bursting is undoubtedly relevant, the manuscript would benefit from a more cautious framing. It would be more appropriate to foreground the methodological contributions of DeepTX, and to present the biological insights as hypotheses or observations that may guide future experimental investigation, rather than as established conclusions.

We thank you for this insightful comment. We have revised the discussion to clarify and appropriately temper our claims regarding transcriptional bursting. First, we now explicitly recognize that transcriptional bursting is one of multiple contributors to cellular variability, rather than the sole or dominant factor driving cellular decision-making. Second, we have restructured the opening of the discussion to prioritize the methodological contributions of DeepTX, highlighting its strength as a framework for inferring genomewide burst kinetics from scRNA-seq data. Finally, the biological insights derived from our analysis are now presented as correlative observations and potential hypotheses, which may inform and guide future experimental investigations, rather than as definitive mechanistic conclusions.

Small Comments

(1) Presentation of discrete distributions: In several figures (e.g., Figure 2B and Supplementary Figures S4, S6, and S8), the comparisons between empirical mRNA distributions and DeepTX-inferred distributions are visually represented using connecting lines, which may give the impression that continuous distributions are being compared to discrete ones. Given the focus on transcriptional bursting, a process inherently tied to discrete stochastic events, this representation could be misleading. The figure captions and visual style should be revised to clarify that all distributions are discrete and to avoid potential confusion. In general, it is recommended to avoid connecting points in discrete distributions with lines, as this can suggest interpolation or comparison with continuous distributions. This applies to Figures 2A and 2B in particular.

We thank you for this valuable suggestion. To prevent any potential misinterpretation of discrete distributions as continuous ones, we have revised the visual representation of the empirical and DeepTX-inferred mRNA distributions in Figures 2B, and Supplementary Figures S4, S6, and S8. Specifically, we have replaced the line plots with step plots, which more accurately capture the discrete nature of transcriptional bursting. Additionally, we have updated the figure captions to clearly state that all distributions are discrete.

(2) Transcription is always a multi-step process. While the manuscript aims to model additional complexity introduced by DNA damage, the current phrasing (e.g., on page 5) could be read as implying that transcription becomes multi-step only under damage conditions. This should be clarified.

We thank you for this helpful observation. We agree that transcription is inherently a multi-step process under all conditions. To avoid any possible misunderstanding, we have revised the text to clarify this point.

Specifically, we now explain that many previous studies have employed simplified two-state models to approximate transcriptional dynamics, however, the gene expression process is inherently a multi-step process, which particularly cannot be neglected under conditions of DNA damage. DNA damage can result in slowing or even stopping the RNA pol II movement and cause many macromolecules to be recruited for damage repair. This process will affect the spatially localized behavior of the promoter, causing the dwell time of promoter inactivation and activation that cannot be approximated by a simple two state. Our work adopts a multi-step model because it is more appropriate for capturing the additional complexity introduced by DNA damage.

(3) The first sentence of the discussion section overstates the importance of transcriptional bursting. While it is a key source of variability, it is not the only nor always the dominant one. Furthermore, its role in DNA damage response remains an emerging hypothesis rather than a general principle. The claims in this section should be moderated accordingly.

We thank you for this valuable feedback. In the revised discussion, we have moderated the statements in the opening paragraph to better reflect the current understanding. Specifically, we now acknowledge that transcriptional bursting represents one of multiple sources of variability and is not always the dominant contributor. In addition, we have reframed the role of transcriptional bursting in DNA damage response as an emerging hypothesis, rather than a general principle. To further address this concern, we replaced conclusion-like statements with more cautious, hypothesis-oriented phrasing, presenting our observations as potential directions for future experimental validation.

References

- Maizels, R.J. 2024. A dynamical perspective: moving towards mechanism in single-cell transcriptomics. *Philos Trans R Soc Lond B Biol Sci* 379: 20230049. DOI: <https://dx.doi.org/10.1098/rstb.2023.0049> , PMID: 38432314
- Sukys, A., Grima, R. 2025. Cell-cycle dependence of bursty gene expression: insights from fitting mechanistic models to single-cell RNA-seq data. *Nucleic Acids Research* 53. DOI: <https://dx.doi.org/10.1093/nar/gkaf295> , PMID: 40240003
- Garrido-Rodriguez, M., Lopez-Lopez, D., Ortuno, F.M., Peña-Chilet, M., Muñoz, E., Calzado, M.A., Dopazo, J. 2021. A versatile workflow to integrate RNA-seq genomic and transcriptomic data into mechanistic models of signaling pathways. *PLoS Computational Biology* 17: e1008748. DOI: <https://dx.doi.org/10.1371/journal.pcbi.1008748> , PMID: 33571195
- Kim, E.M., Burke, D.J. 2008. DNA damage activates the SAC in an ATM/ATR-dependent manner, independently of the kinetochore. *PLoS Genet* 4: e1000015. DOI: <https://dx.doi.org/10.1371/journal.pgen.1000015> , PMID: 18454191
- Lawrence, K.S., Chau, T., Engebrecht, J. 2015. DNA damage response and spindle assembly checkpoint function throughout the cell cycle to ensure genomic integrity. *PLoS Genet* 11: e1005150. DOI: <https://dx.doi.org/10.1371/journal.pgen.1005150> , PMID: 25898113

Mantel, C., Guo, Y., Lee, M.R., Kim, M.K., Han, M.K., Shibayama, H., Fukuda, S., Yoder, M.C., Pelus, L.M., Kim, K.S., Broxmeyer, H.E. 2007. Checkpoint-apoptosis uncoupling in human and mouse embryonic stem cells: a source of karyotypic instability. *Blood* 109: 4518-4527. DOI: <https://dx.doi.org/10.1182/blood-2006-10-054247>, PMID: 17289813

Tang, W., Jørgensen, A.C.S., Marguerat, S., Thomas, P., Shahrezaei, V. 2023. Modelling capture efficiency of single-cell RNA-sequencing data improves inference of transcriptome-wide burst kinetics. *Bioinformatics* 39. DOI: <https://dx.doi.org/10.1093/bioinformatics/btad395>, PMID: 37354494

Luo, S., Zhang, Z., Wang, Z., Yang, X., Chen, X., Zhou, T., Zhang, J. 2023. Inferring transcriptional bursting kinetics from single-cell snapshot data using a generalized telegraph model. *Royal Society Open Science* 10: 221057. DOI: <https://dx.doi.org/10.1098/rsos.221057>, PMID: 37035293

Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, A., Rivera, C.M., Ren, B., Sandberg, R. 2019. Genomic encoding of transcriptional burst kinetics. *Nature* 565: 251-254. DOI: <https://dx.doi.org/10.1038/s41586-018-0836-1>, PMID: 30602787

Gu, J., Laszik, N., Miles, C.E., Allard, J., Downing, T.L., Read, E.L. 2025. Scalable inference and identifiability of kinetic parameters for transcriptional bursting from single cell data. *Bioinformatics*. DOI: <https://dx.doi.org/10.1093/bioinformatics/btaf581>, PMID: 41131798.
<https://doi.org/10.7554/eLife.100623.3.sa0>