Analytical distribution of nascent RNA in a queuing model with applications

Changhong Shi[1], Xiyan Yang[2], Tianshou Zhou[3,*] , and Jiajun Zhang[3,*]

1. School of Public Health, Guangzhou Medical University, Guangzhou, China
2. School of Financial Mathematics and Statistics, Guangdong University of Finance, Guangzhou, China
3. School of Mathematics and Computational Science and Guangdong Province Key Laboratory of Computational Science, Sun Yat-Sen University, Guangzhou, China

[*]The correspondence author: mcszhtsh@mail.sysu.edu.cn, zhjiajun@mail.sysu.edu.cn

Abstract: The stochastic kinetics of transcription can be accurately inferred from the distribution of nascent RNA . However, the inferences using conventional gene-expression models with exponential waiting times between promoter active and inactive states would be inaccurate since these waiting times may follow non-exponential distributions. For instance, the refractory period of the prolactin gene in mammalian cells obeys a bimodal distribution. In this paper, we have developed a general stochastic model for nascent RNA that takes into account the complexity of promoter structure and encompasses various traditional models of gene transcription as its special cases. Importantly, we have derive the exact distribution and moments of nascent RNA using techniques in queuing theory, which provides insight into the effect of molecular memory on transcription dynamics. Based on the derived analytical results, we further developed a method to infer non-exponential waiting-time distributions between promoter states from stationary nascent RNA distributions. Data analysis of a realistic example verifies the validity of this inference method.

Introduction

Transcription is a complex biochemical process, inevitably leading to stochastic fluctuation in mRNA and subsequent protein production. The experimental distribution of mature mRNA numbers can often be fitted using an exact solution of a two-state model of gene expression, which has successfully revealed the mechanism of bursting and the non-Poisson nature of transcription [1-7]. However, the fluctuations of mature mRNA in a cell are strongly influenced by various possible post-transcriptional processes such as splicing, nuclear export, and miRNA mediation [8-10]. Therefore, conventional gene-expression models may not be suitable when used for data-based inferences and more elaborate models need to be developed.

An alternative approach for quantifying transcription products (mRNA) is to count the number of nascent RNAs or the RNA polymerase molecules (RNAP) at the single-cell level. Consequently, several methods have been developed to quantify single-cell nascent RNA [11-13], such as the molecular fluorescence in situ hybridization

(smFISH) method, which provides a steady-state distribution of nascent RNA. On the other hand, fitting the experimental data of nascent RNA using stochastic models of nascent RNA allows for a more precise understanding of regulatory steps involved in achieving the desired mRNA expression level [14, 15]. As a result, numerous stochastic models have been recently developed to simulate nascent RNA kinetics [16-23]. For instance, Choubey, et al., [22] proposed a two-state model for stochastic kinetics of nascent RNA that assumes gene switching between an active and an inactive state with transcription initiation occurs only in the active state. Subsequently, polymerase moves from one base pair to the next until reaching the end of the gene. Additionally, the authors used the queuing theory to derive the analytical expressions of the transient and steady-state expectations as well as the variance for the number of nascent RNA when elongation time is deterministic. Furthermore, Heng, et al., [14] introduced a coarse-grained version of this model by neglecting RNAP movement between base pairs and derived an analytical expression of steady-state nascent RNA distribution.

It is worth noting that the two-state model assumes that the time spent in the active or inactive state follows an exponential distribution. While this assumption may be reasonable for bacteria [24], recent experimental studies have revealed that the inactive periods of many genes in mammalian cells follow non-exponential peaked distributions [25]. This suggests that the transition from an inactive to an active state is a multi-step process containing multiple inactive states. More generally, multiple transcriptional active and inactive states may exist in both prokaryotic and eukaryotic cells. For instance, in *E. coli* , the PRM promoter of phage lambda is regulated by two different transcription factors (TFs) binding to two sets of operators, which can be brought together through DNA looping. Consequently, the number of regulatory states of the PRM promoter is up to 128 [26]. In contrast, eukaryotic promoters would be more complex since they involve nucleosomes competing with or being removed by TFs [27]. In addition to the conventional regulation by TFs, the eukaryotic promoters can also undergo epigenetically regulation through histone modifications [28-30], which may lead to complex promoter structure. We point out that a multistep process has the potential to create memory between molecular events, implying that waiting-time distributions between these events are non-exponential and lead to non-Markovian dynamics of the underlying system.

In recent years, gene expression models with complex multi-state promoter switching mechanisms have been widely studied from a theoretical perspective [31-38]. However, most of these studies have focused on the distribution of mature RNA and protein in steady or transient state, and the effect of promoter structure on the stationary distribution of nascent RNA is remains incompletely understood. Recently, Juraj et al. [18] developed a general framework, which allows one to find nascent RNA distribution of models with stochastic and multistep initiation, under the assumption of deterministic elongation and termination. However, this framework is based on the assumption that the transcription initiation is a renewal process, which is true when there is only one active gene state

and multiple inactive gene states. Previous studies have primarily examined the steady-state or time-dependent behavior of mature mRNA or protein, and thus far, there has been still a lack of detailed studies about the steady-state behavior of nascent RNA with arbitrary promoter structure. Therefore, the aim of this study is twofold: one is to develop a general stochastic model of nascent RNA with complex promoter structure to include the studied models of nascent RNA as its special cases, and the other is to derive the nascent RNA distribution to show how molecular memory impacts stochastic kinetics of nascent RNA.

A general stochastic model of nascent RNA

Here we introduce a stochastic model of gene expression with complex promoter switching followed by a deterministic elongation process. Assume that the gene promoter has L distinct states, denoted by $S_1, S_2, ..., S_L$. Biophysically, these states correspond to different conformational states during chromatin remodeling or to different binding states for TFs. We use the matrix $\mathbf{W} = \left[\omega_{ij}\right]$ to describe the switching rates between promoter states with elements ω_{ij} representing the transition rate from state S_i to state S_j, satisfying the relationship $\omega_{ii} = -\sum_{j \neq i} \omega_{ij}$. We use the diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}\left(k_1, k_2, ..., k_L\right)$ to describe the rates of transcription initiation, where k_i is the transcription initiation rate when the promoter state is at S_i state. We further assume that the elongation of a single transcript is deterministic and the elongation speed is a constant k_{elo} kb/min. Consequently, the elongation time is $t_{\mathrm{elo}} = l_{\mathrm{gene}} / k_{\mathrm{elo}}$ where l_{gene} is the gene length. Although a transcription elongation process would be very complex, e.g., it would involve pausing and backtracking of polymerases along the gene body, our assumption of deterministic elongation is justified for a vast number of genes [22,39]. Then the survival function of the nascent RNA is given by

$$H(t) = \begin{cases} 1, & t \leq t_{\mathrm{elo}}, \\ 0, & \text{else.} \end{cases} \tag{1}$$

Let $P_i(n,t)$ denote the probability that the number of nascent RNAs is n at time t, given that promoter state is i at time 0 and the number of nascent RNA is 0 at time 0, i.e.,

$$P_i\left(n,t\right) = \Pr\left\{N(t) = n \middle| N(0) = 0, U(0) = i\right\}, \quad i = 1, 2, ..., L, \tag{2}$$

where $U(t)$ and $N(t)$ represent the promoter state and the nascent RNA number at time t, respectively. For convenience, we denote $\boldsymbol{P}\left(n,t\right) = \left(P_1\left(n,t\right), ..., P_L\left(n,t\right)\right)^{\mathrm{T}}$, where the symbol 'T' represents transpose. Using the method proposed in our recent studies [40], we can derive the following so-called generalized master equation for $\boldsymbol{P}(n,t)$

$$\frac{\partial}{\partial t} \boldsymbol{P}\left(n,t\right) = \left[\mathbf{W} - H(t)\mathbf{\Lambda}\right] \boldsymbol{P}\left(n,t\right) + H\left(t\right) \mathbf{\Lambda} \boldsymbol{P}\left(n-1,t\right). \tag{3}$$

Let $\boldsymbol{G}\left(z,t\right) = \left(G_1\left(z,t\right), G_2\left(z,t\right), ..., G_L\left(z,t\right)\right)^{\mathrm{T}}$ be the probability generating function corresponding to $\boldsymbol{P}\left(n,t\right)$, i.e., $\boldsymbol{G}\left(z,t\right) = \sum_{n=0}^{\infty} z^n \boldsymbol{P}\left(n,t\right)$. Then Eq. (3) can be transformed to

$$\frac{\partial}{\partial t}\boldsymbol{G}(z,t)=\left[\mathbf{W}+(z-1)H(t)\boldsymbol{\Lambda}\right]\boldsymbol{G}(z,t), \tag{4}$$

with the initial condition $\boldsymbol{G}(z,0)=\boldsymbol{e}$, where $\boldsymbol{e}=(1,1,...,1)^{\mathrm{T}}$.

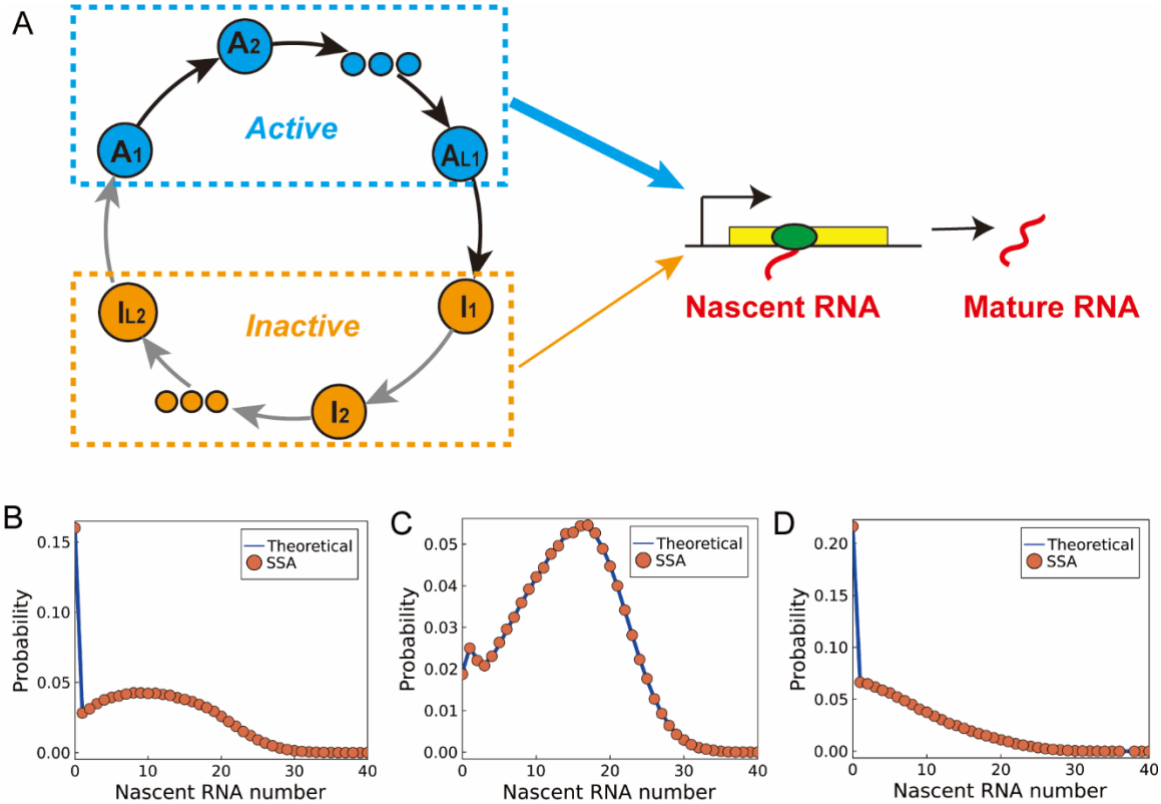


Fig. 1. Schematic diagram for a model of nascent RNA with the cyclic promoter structure, where the blue and red cycles represent transcription active (ON) and inactive (OFF) states, respectively. (B)(C)(D) The steady-state nascent RNA distributions obtained by the matrix-form solution and the stochastic simulation algorithm (SSA). The parameter values are set as: (A) $u=1, v=1.5, L_1=L_2=2, k_1=20, k_0=0$. (B) $u=0.7,\ v=1.5, L_1=L_2=2, k_1=20, k_0=1$. (C) $u=2, v=1.5, L_1=1, L_2=3, k_1=20, k_0=0$.

As a special case, we consider the case of a cyclic promoter structure, i.e., the promoter proceeds sequentially through several irreversible active (ON) state and inactive (OFF) states, which altogether form a cycle (Fig. 1 (A)). Note that this cyclic model can be used to describe many processes involved in transcription such as recruitment of various polymerases and transcription factors, assembly of the pre-initiation complexes, chromatin remodeling, and histone modification [36]. For simplicity, we assume that the transcription initiation rate is k_1 at each active state and k_0 at each inactive state, where $k_1>k_0$. Specifically, we assume that the gene promoter has L states in total, which consist of L_1 active states and L_2 inactive states. Denoted by $A_1,...,A_{L_1}$ ON states and by $I_1,...,I_{L_2}$ OFF states. The corresponding biochemical reactions are listed in Table 1. Note that if $L_1=1$ and $L_2>1$, the cyclic model

becomes the multi-OFF model (i.e., the promoter has more than one OFF state and only one ON state). If $L_1 > 1$ and $L_1 = 1$, the cyclic model becomes the multi-ON model (i.e., the promoter has more than one ON state and only one OFF state). In particular, if $L_1 = 1$ and $L_1 = 1$, the cyclic model becomes the classical two-state model. Therefore, the introduced-above cyclic model includes the transcription model being previously studied.

Table 1 Biochemical reactions for a transcription model to be studied.

Reaction	**Description**
$\mathrm{A}_i \xrightarrow{\omega_{i,i+1}^{(\mathrm{on})}} \mathrm{A}_{i+1}$, $i = 1, 2, ..., L_1 - 1$	Transitions between active states
$\mathrm{I}_j \xrightarrow{\omega_{j,j+1}^{(\mathrm{off})}} \mathrm{I}_{j+1}$, $j = 1, 2, ..., L_2 - 1$	Transitions between inactive states
$\mathrm{A}_{L_1} \xrightarrow{\omega_{L_1,1}^{(\mathrm{on})}} \mathrm{I}_1$, $\mathrm{I}_{L_2} \xrightarrow{\omega_{L_2,1}^{(\mathrm{off})}} \mathrm{A}_1$	Transitions between active and inactive states
$\mathrm{A}_i \xrightarrow{k_1} \mathrm{A}_i + \mathrm{nascentRNA}$	Transcription initiation at the active state
$\mathrm{I}_j \xrightarrow{k_0} \mathrm{I}_j + \mathrm{nascentRNA}$	Transcription initiation at the inactive state
$\mathrm{nascent\ RNA} \underset{t_{\mathrm{elo}}}{\Rightarrow} \varnothing$	Nascent RNA elongation

To make an equitable comparison between models with different values of L, we assume that the transition rates at the ON (OFF) states are the same, i.e., $\omega_{1,2}^{(\mathrm{on})} = ... = \omega_{L_1-1,L_1}^{(\mathrm{on})} = \omega_{L_1,1}^{(\mathrm{on})} = L_1 u$, $\omega_{1,2}^{(\mathrm{off})} = ... = \omega_{L_2-1,L_2}^{(\mathrm{off})} = \omega_{L_2,1}^{(\mathrm{off})} = L_2 v$, where u and v are two constants. With these assumptions, the dwell time at the ON state, τ_{on}, obeys an Erlang distribution with shape parameter L_1 and mean $1/u$, and the dwell time at the OFF state, τ_{off}, follows an Erlang distribution with shape parameter L_2 and mean $1/v$. Note that if the noise of the dwell time is characterized by the coefficient of variation, it is $1/L_1$ for the dwell time τ_{on}, and $1/L_2$ for the dwell time τ_{off}. As $L_i \to \infty$, this noise tends to zero and the dwell time converges to a deterministic value and the molecular memory is the strongest in each case. In the following paper, L_1 and L_2 are called as memory index of promoter activation and deactivation, respectively.

Distribution of nascent RNA in steady state

In previous studies [40], we used the finite state projection (FSP) method to calculate a numerical nascent RNA distribution $\boldsymbol{P}(n,t)$ based on Eq. (3). Although the FSP method can ensure that the approximation error is within a range of some pre-specified tolerance level, it is inefficient to obtain an accurate nascent RNA distribution in the steady state since it often takes a long time when a system reaches the steady state. In this section, we give two forms of the exact solution for the stationary distribution of nascent RNA: Laplace-form solution and matrix-form solution.

Laplace-form solution

The closed-form solution of Eq. (4) is generally difficult to obtain. Under the assumption that the elongation time

is deterministic, the solution of Eq. (4) is given by

$$\boldsymbol{G}(z,t)=\begin{cases}\exp\left[\left(\mathbf{W}+(z-1)\boldsymbol{\Lambda}\right)t\right]\boldsymbol{e} & t\le t_{\text{elo}}\\ \exp\left[\mathbf{W}(t-t_{\text{elo}})\right]\exp\left[\left(\mathbf{W}+(z-1)\boldsymbol{\Lambda}\right)t_{\text{elo}}\right]\boldsymbol{e} & t>t_{\text{elo}}\end{cases}. \tag{5}$$

Note that matrix exponential function $\exp\left[\mathbf{W}(t-t_{\text{elo}})\right]$ tend to $\boldsymbol{e\pi}$ when $t\to\infty$, where the row vector $\boldsymbol{\pi}$ denotes the steady state of the promoter state satisfying $\boldsymbol{\pi}\mathbf{W}=\mathbf{0}$ and $\boldsymbol{\pi e}=1$. Since the steady-state distribution of nascent RNA is independent of the initial state: $P(n)\equiv P_1(n,\infty)=...=P_L(n,\infty)$, we have $G(z)\equiv G_1(z,\infty)=G_2(z,\infty)=...=G_L(z,\infty)$. Combining these results with Eq. (5), we can derive the analytical expression of the stationary nascent RNA distribution as follows

$$G(z)=\boldsymbol{\pi}\exp\left[\left(\mathbf{W}+(z-1)\boldsymbol{\Lambda}\right)t_{\text{elo}}\right]\boldsymbol{e}, \tag{6}$$

which should be a function of t_{elo}, re-denoted by $G(z,t_{\text{elo}})$. The Laplace transform of $G(z,t_{\text{elo}})$ with respect to t_{elo} is given by

$$G^*(z,s)=\boldsymbol{\pi}\left(s\mathbf{I}-\mathbf{U}-\boldsymbol{\Lambda}z\right)^{-1}\boldsymbol{e}, \tag{7}$$

where $\mathbf{U}=\mathbf{W}-\boldsymbol{\Lambda}$. Note that the function $G^*(z,s)$ has the following Taylor expansion

$$G^*(z,s)=\sum_{n=0}^{\infty}\boldsymbol{\pi}\left[\left(s\mathbf{I}-\mathbf{U}-\boldsymbol{\Lambda}z\right)^{-1}\boldsymbol{\Lambda}\right]^n\left(s\mathbf{I}-\mathbf{U}\right)^{-1}\boldsymbol{e}z^n. \tag{8}$$

On the other hand, according to the definition of probability generating function, we know that function $G^*(z,s)$ can also be written as $G^*(z,s)=\sum_{n=0}^{\infty}P^*(n,s)z^n$, where $P^*(n,s)$ is the Laplace transform of $P(n)$ (which should be also a function of t_{elo}) with respect to t_{elo}. Therefore, we obtain the Laplace-form solution of stationary nascent RNA distribution

$$P^*(n,s)=\boldsymbol{\pi}\left[\left(s\mathbf{I}-\mathbf{U}\right)^{-1}\boldsymbol{\Lambda}\right]^n\left(s\mathbf{I}-\mathbf{U}\right)^{-1}\boldsymbol{e}. \tag{9}$$

Then, the inverse Laplace transform gives the explicit expression of the probability distribution $P(n)$. For $n=0$, we have

$$P(n=0)=\boldsymbol{\pi}\exp\left(\mathbf{U}t_{\text{elo}}\right)\boldsymbol{e}. \tag{10}$$

For $n>0$, it is possible, albeit tedious, to derive the analytical expression of $P(n)$. Instead, we can numerically calculate the probability distribution.

We can also derive the analytical expression for the stationary moments of the nascent RNA with a general promoter switching. Let $u^{(k)}$ denotes the *k*-th factorial moments of nascent RNA number N in the steady state, i.e., $u^{(k)}=E\left[N(N-1)...(N-k+1)\right]$ Applying the formula $u^{(k)}=\partial_z^k G(z)|_{z=1}$ to Eq. (6), we obtain the first two factorial moments (see reference [41,42] for more detail)

$$u^{(1)} = \boldsymbol{\pi\Lambda e} t_{\text{elo}}, \tag{11}$$

and

$$u^{(2)} = t_{\text{elo}}^2 \left(\boldsymbol{\pi\Lambda e}\right)^2 + 2t_{\text{elo}} \left[\boldsymbol{\pi\Lambda}\left(\tau \boldsymbol{e\pi} - \mathbf{W}\right)^{-1} \boldsymbol{\Lambda e} - \tau^{-1}\left(\boldsymbol{\pi\Lambda e}\right)^2\right] + 2\boldsymbol{\pi\Lambda}\left(\exp(\mathbf{W}t_{\text{elo}}) - \mathbf{I}\right)\left(\tau \boldsymbol{e\pi} - \mathbf{W}\right)^{-2} \boldsymbol{\Lambda e}, \tag{12}$$

where $\tau \neq 0$. Therefore, the Fano factor, which is defined as the ratio of the variance over the mean, is given by

$$FF = 1 + 2\frac{\boldsymbol{\pi\Lambda}\left(\tau \boldsymbol{e\pi} - \mathbf{W}\right)^{-1} \boldsymbol{\Lambda e} - \tau^{-1}\left(\boldsymbol{\pi\Lambda e}\right)^2}{\boldsymbol{\pi\Lambda e}} + \frac{2}{t_{\text{elo}}}\frac{\boldsymbol{\pi\Lambda}\left(\exp(\mathbf{W}t_{\text{elo}}) - \mathbf{I}\right)\left(\tau \boldsymbol{e\pi} - \mathbf{W}\right)^{-2} \boldsymbol{\Lambda e}}{\boldsymbol{\pi\Lambda e}}. \tag{13}$$

Note that the value of FF does not depend on the choice of τ. Setting $\tau = -1$, we have

$$FF = 1 + 2\frac{\left(\boldsymbol{\pi\Lambda e}\right)^2 - \boldsymbol{\pi\Lambda}\left(\boldsymbol{e\pi} + \mathbf{W}\right)^{-1} \boldsymbol{\Lambda e}}{\boldsymbol{\pi\Lambda e}} + \frac{2}{t_{\text{elo}}}\frac{\boldsymbol{\pi\Lambda}\left(\exp(\mathbf{W}t_{\text{elo}}) - \mathbf{I}\right)\left(\boldsymbol{e\pi} + \mathbf{W}\right)^{-2} \boldsymbol{\Lambda e}}{\boldsymbol{\pi\Lambda e}}.$$

As $t_{\text{elo}} \to \infty$, we have

$$FF = 1 + 2\frac{\left(\boldsymbol{\pi\Lambda e}\right)^2 - \boldsymbol{\pi\Lambda}\left(\boldsymbol{e\pi} + \mathbf{W}\right)^{-1} \boldsymbol{\Lambda e}}{\boldsymbol{\pi\Lambda e}}. \tag{14}$$

Matrix-form solution

Although Eq. (9) allows us to calculate the form-closed solution of stationary nascent RNA distribution for some special cases, we find that it is inefficient to calculate the stationary distribution numerically in general cases. Therefore, we try to derive the matrix-form solution of the stationary nascent RNA distribution to facilitate numerical computation. Assuming that there exists a large N such that $P_i(n,t) = 0$ for $i = 1, 2, \ldots, K$ and $n \geq N+1$. We denote

$$\boldsymbol{Q}(t) = \left(P_1(0,t), \ldots, P_L(0,t), P_1(1,t), \ldots, P_L(1,t), \ldots, P_1(N,t), \ldots, P_L(N,t)\right)^{\mathrm{T}}. \tag{15}$$

For $t \leq t_{\text{elo}}$, Eq. (3) can be solved by the following ordinary differential equation (ODE) up to time t

$$\frac{\partial}{\partial t}\boldsymbol{Q}(t) = \mathbf{A}\boldsymbol{Q}(t), \tag{16}$$

where matrix $\mathbf{A}$ takes the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{W}-\boldsymbol{\Lambda} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Lambda} & \mathbf{W}-\boldsymbol{\Lambda} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}-\boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{W}-\boldsymbol{\Lambda} \end{bmatrix}_{L(N+1)\times L(N+1)}, \tag{17}$$

and the initial condition is

$$\boldsymbol{Q}(0) = \left(\underbrace{1, 1, \ldots, 1}_{L}, \underbrace{0, 0, \ldots, 0}_{L}, \ldots, \underbrace{0, 0, \ldots, 0}_{L}\right)^{\mathrm{T}}. \tag{18}$$

For $t \leq t_{\text{elo}}$, the solution of the above equation is

$$\boldsymbol{Q}(t)=\exp(\mathbf{A}t)\boldsymbol{Q}(0). \tag{19}$$

Therefore, we have

$$\boldsymbol{P}(n,t)=\mathbf{E}_n\exp(\mathbf{A}t)\boldsymbol{Q}(0), \tag{20}$$

where $\mathbf{E}_n=\boldsymbol{e}_n\otimes\boldsymbol{e}$, $\otimes$ represents the Kronecker production, $\boldsymbol{e}_n$ denotes the N-dimension vector where the n-th coordinate is 1 and the others are zero, i.e., $\boldsymbol{e}_n=(0,\ldots,1,\ldots,0)_{1\times N}$, and $\boldsymbol{e}=(1,1,\ldots,1)_{1\times N}$.

For $t>t_{\text{elo}}$, the solution of Eq. (3) is

$$\boldsymbol{P}(n,t)=\exp\left(\mathbf{W}(t-t_{\text{elo}})\right)\boldsymbol{P}(n,t_{\text{elo}}). \tag{21}$$

As $t\to\infty$, we obtain the following matrix-form solution of the stationary nascent RNA distribution

$$P(n)=\boldsymbol{\pi}\mathbf{E}_{\mathbf{n}}\exp(\mathbf{A}t_{\text{elo}})\boldsymbol{Q}(0). \tag{22}$$

To verify the validity of the theoretical solution given by Eq. (22), we consider a four-state cyclic model of gene expression with deterministic elongation time. Fig. 1 (B)(C)(D) exhibits numerical [43] and theoretical results for the steady-state distribution of nascent RNA for three different sets of parameter values. From these diagrams, we observe that both numerical and theoretical results are well consistent.

The two-state gene expression model: A case study

As a case, we consider a general two-state gene expression model with a deterministic elongation time, which assumes that the promoter switches between two gene expression states, an inactive state A and an active state I. The transcription initial rate is k_1 at the inactive state and k_2 at the active state. The two-state gene expression model is a special case of the cyclic model with $L_1=1$ and $L_2=1$. The reaction scheme for this process is listed as following

$$\begin{aligned}
&\mathrm{I}\underset{\omega_{21}}{\overset{\omega_{12}}{\rightleftharpoons}}\mathrm{A},\\
&\mathrm{A}\xrightarrow{k_2}\text{nascent RNA},\\
&\mathrm{I}\xrightarrow{k_1}\text{nascent RNA},\\
&\text{nascent RNA}\underset{t_{\text{elo}}}{\Rightarrow}\varnothing
\end{aligned} \tag{23}$$

Previous studies have derived the analytical solution of nascent RNA distribution under the assumption that the transcription initiation only occurs at the active states, i.e., $k_2>0$ and $k_1=0$ In this section, we derive the steady state distribution of nascent RNA for a general two-state gene expression model with leaky transcription, i.e., $k_2>k_1\geq 0$.

Applying Eq. (10), and using Sylvster's matrix theorem to calculate $\left[(s\mathbf{I}-\mathbf{U})^{-1}\boldsymbol{\Lambda}\right]^n$, we finally obtain the Laplace-form solution of nascent RNA distribution for the general two-state model as following

$$P^*(0,s)=\frac{s+\lambda_1+\lambda_2-\left(k_2\omega_{12}+k_1\omega_{21}\right)/\left(\omega_{12}+\omega_{21}\right)}{\left(s+\lambda_1\right)\left(s+\lambda_2\right)}, \tag{24}$$

$$\begin{aligned}P^*(n,s)=&\frac{\gamma_1^n-\gamma_2^n}{\gamma_1-\gamma_2}\frac{k_1\omega_{21}\left(s+k_1+\omega_{12}+\omega_{21}\right)^2+k_2\omega_{12}\left(s+k_2+\omega_{12}+\omega_{21}\right)^2}{\left(\omega_{12}+\omega_{21}\right)\left[\left(s+\lambda_1\right)\left(s+\lambda_2\right)\right]^{n+1}}\\&-\frac{\gamma_1\gamma_2\left(\gamma_1^{n-1}-\gamma_2^{n-1}\right)}{\gamma_1-\gamma_2}\frac{s+\lambda_1+\lambda_2-\left(k_2\omega_{12}+k_1\omega_{21}\right)/\left(\omega_{12}+\omega_{21}\right)}{\left[\left(s+\lambda_1\right)\left(s+\lambda_2\right)\right]^{n}},\quad \text{for } n>1,\end{aligned} \tag{25}$$

where $-\lambda_1$ and $-\lambda_2$ are roots of the quadratic equation

$$\left(s+\lambda_1\right)\left(s+\lambda_2\right)=\left(s+k_1+\omega_{12}\right)\left(s+k_2+\omega_{21}\right)-\omega_{12}\omega_{21}, \tag{26}$$

and γ_1 and γ_2 are roots of the quadratic equation

$$\left(x-\gamma_1\right)\left(x-\gamma_2\right)=\left[x-k_1\left(s+k_2+\omega_{21}\right)\right]\left[x-k_2\left(s+k_1+\omega_{12}\right)\right]-k_1k_2\omega_{12}\omega_{21}. \tag{27}$$

When $k_1=0$, the roots of the above equation are $\gamma_1=0$, $\gamma_2=k_2\left(s+\omega_{12}\right)$. We have

$$P^*(0,s)=\frac{s+\lambda_1+\lambda_2-k_2\omega_{12}/\left(\omega_{12}+\omega_{21}\right)}{\left(s+\lambda_1\right)\left(s+\lambda_2\right)}, \tag{28}$$

$$P^*(n,s)=\frac{k_2\omega_{12}\left(s+\omega_{12}+\omega_{21}\right)^2\left[k_2\left(s+\omega_{12}\right)\right]^{n-1}}{\left(\omega_{12}+\omega_{21}\right)\left[\left(s+\lambda_1\right)\left(s+\lambda_2\right)\right]^{n+1}}. \tag{29}$$

We note that this result has been obtained in previous studies [18].

Using Eq.(31) and setting $\tau=\omega_{12}+\omega_{21}$, we obtain the Fanor factor of the nascent RNA number in the steady state

$$FF=1+\frac{2\omega_{12}\omega_{21}\left(k_1-k_2\right)^2}{\left(\omega_{12}+\omega_{21}\right)^2\left(\omega_{12}k_1+\omega_{21}k_2\right)}+\frac{2\omega_{12}\omega_{21}\left(k_1-k_2\right)^2}{\left(\omega_{12}+\omega_{21}\right)^2\left(\omega_{12}k_1+\omega_{21}k_2\right)}\frac{e^{-\left(k_{12}+k_{21}\right)t_{\text{elo}}}-1}{t_{\text{elo}}}. \tag{30}$$

When $k_2>0$ and $k_1=0$, we have

$$FF=1+\frac{2\omega_{12}k_2}{\left(\omega_{12}+\omega_{21}\right)^2}+\frac{2\omega_{12}k_2}{\left(\omega_{12}+\omega_{21}\right)^3}\frac{e^{-\left(\omega_{12}+\omega_{21}\right)t_{\text{elo}}}-1}{t_{\text{elo}}}. \tag{31}$$

We note that these results have been obtained in previous studies [23].

Effect of molecular memory on nascent RNA distribution

Previous studies have demonstrated that molecular memory can modulate the distribution of mature RNA in both steady state and transient state [31-34]. However, the effect of molecular memory on the stationary distribution of nascent RNA remain incompletely understood. In this section, we investigate the influence of molecular memory on the stationary nascent RNA distribution based on the cyclic model. Within this model, two types of molecular memory are considered: active memory during the active period (characterized by index L_1) and inactive memory

during the inactive period (characterized by index L_2).

To understand how active and inactive memory affect the stationary nascent RNA distribution, we first focus on the cyclic model with fast or slow promoter switching. Under the slow-switching condition, i.e., $u \to 0, v \to 0$, Eq. (6) can be written as $G(z) = \pi_0 \exp(k_0 t_{\text{elo}}(z-1)) + \pi_1 \exp(k_1 t_{\text{elo}}(z-1))$, where $\pi_0 = v/(u+v)$ and $\pi_1 = u/(u+v)$, indicating that the stationary distribution is a mixture Poisson distribution. In contrast, under the fast-switching condition, i.e., $u \to \infty, v \to \infty$, Eq. (6) can be written as $G(z) = \exp((\pi_0 k_0 + \pi_1 k_1) t_{\text{elo}}(z-1))$, suggesting that the stationary nascent RNA distribution is a Poisson distribution. Consequently, both active and inactive memory do not alter the stationary distribution of nascent RNA in the limiting cases characterized by fast and switching regimes.

We further investigate the impact of active and inactive memory on the stationary nascent RNA distribution in cases where promoter switching is neither too fast or too slow. To illustrate this, we present the $u-v$ phase diagram for the two-state model ($L_1 = 1$, $L_2 = 1$) (Fig. 2(A)), multi-ON model ($L_1 = 5$, $L_2 = 1$) (Fig. 2(B)) and multi-OFF model ($L_1 = 1$, $L_2 = 5$) (Fig. 2(C)). In all models, the stationary nascent RNA distribution exhibit three main types: unimodal distribution with zero peak (Region I or Fig. 2(D)), unimodal distribution with non-zero peak (Region II or Fig. 2(E)), and bimodal distribution with a zero peak and a non-zero peak (Region III or Fig. 2(F)). Comparing these phase diagrams across the different models, we observe molecular memory can influence the size of each region: (1) increasing the active memory L_1 leads to a transformation from unimodal distribution of type *I* to bimodal distribution, resulting in a significant reduction in region *I* as active memory L_1 increases; (2) increasing inactive memory L_2 cause a transformation from bimodal distribution to the unimodal distribution of type *II*, leading to an enlargement of Region *II* as inactive memory L_2 increases. In addition, we discover that increasing the inactive memory L_2 in the mult-OFF model induce a trimodal distribution with a zero peak and two non-zero peaks ((type IV or Fig. 2(G)), which has not been previous reported.

Further more, we demonstrate the stationary nascent RNA distribution under different choices of L_1 and L_2 (Fig. 2(H)). From this figure, it is evident that increasing inactive memory L_2 tends to eliminating the zero peak and transform the bimodal distribution into an unimodal distribution. When $L_1 = 1$, increasing L_2 can induce a new non-zero peak (the upper row in Fig. 2(H)), leading to the trimodal distribution in Region *IV*. When L_2 is sufficient large, i.e., $L_2 = 15$, the peak at zero almost disappears and the trimodal distribution becomes a bimodal distribution with two non-zero peaks. When both active and inactive memory are larger than 1, i.e., $L_1 > 1$ and $L_2 > 1$, two non-zero peaks of the bimodal distribution tend to merge together, leading to an unimodal distribution.

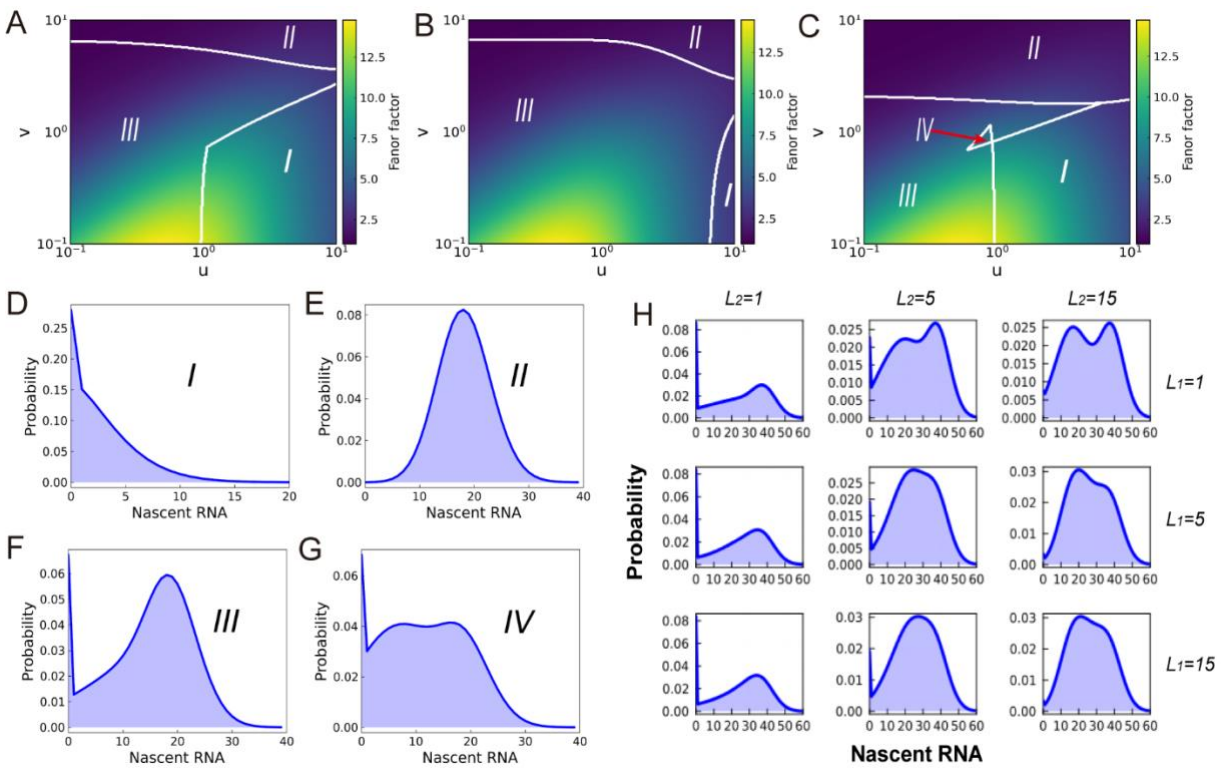


Fig. 2. The effect of molecular memory in the promoter switching on the nascent RNA distribution in the steady state. (A) The *u*-*v* phase diagram for the two-state model $(L_1 = 1, L_2 = 1)$; (B) The *u*-*v* phase diagram for the multi-ON model $(L_1 = 5, L_2 = 1)$; (C) The *u*-*v* phase diagram for the multi-OFF model $(L_1 = 1, L_2 = 5)$; other parameter values are set as $k_1 = 20, k_0 = 0$; (D)(E)(F)(G) The distribution of different type. The parameter values of each type are set as Type *I* $u = 12, v = 2$; Type *II* $u = 1, v = 10$; Type *III* $u = 0.5, v = 1.5$; Type *IV* $u = 0.8, v = 1.2, L_2 = 8$. Other parameters are the same as those in (A); (D) the steady state of nascent RNA distribution under different choices of L_1 and L_2. Other parameters values are set as: $u = 0.8, v = 1.5, k_1 = 40, k_0 = 0$.

To understand why increasing the active memory L_1 can transform the unimodal distribution of type *I* to the bimodal distribution, we investigate the stationary nascent RNA distribution in the burst regime ($u \ll v$). From Fig. 3(C)(D), it is evident that increasing active memory L_1 diminishes the probability $P(n=0)$ and $P(n=1)$ simultaneously, and concurrently gives rise to a valley between two peaks. Consequently, active memory L_1 can induce bimodal distribution in the burst regime. However, within the burst regime, increasing L_2 significantly reduce the probability $P(n=0)$ while increase the probability $P(n=1)$ (Fig. 3(A)(B)), and the unimodal distribution remains unchanged. Hence, active and inactive memory exhibit distinct modulation effects on nascent RNA distribution. It should be noted that this characteristic may serve as an indicator for distinguishing active memory form station nascent RNA distribution; i.e., when the stationary nascent RNA displays bimodal distribution in the bursting regime, it suggests that the active memory may be larger than 1.

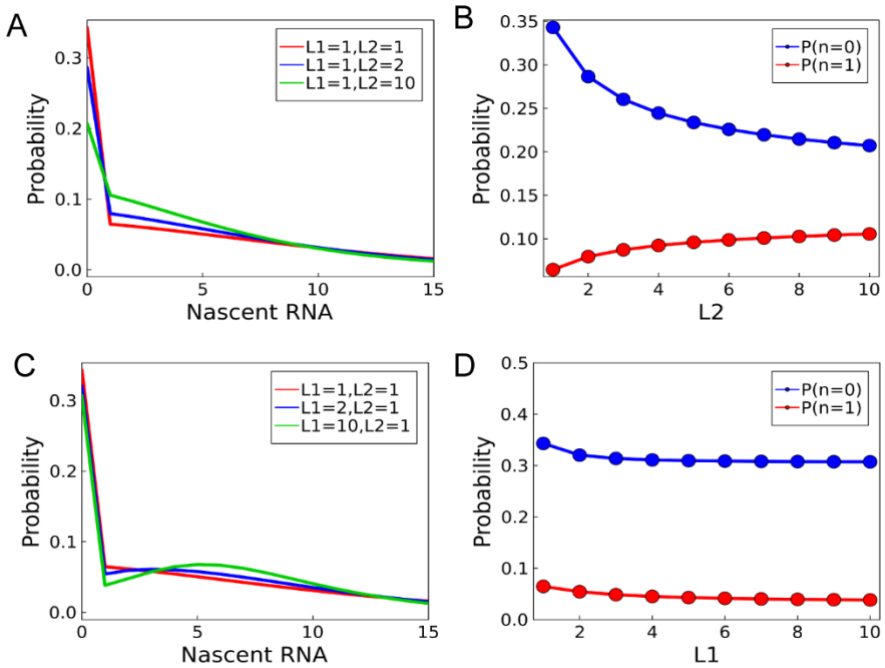


Fig. 3. The effect of memory on the stationary nascent RNA distribution in the burst regime $(u \gg v)$. (A) The steady-state distribution of nascent RNA for different values of L_2; (B) The probability of $P(n=0)$ and $P(n=1)$ as a function of L_2; (C) The steady-state distribution of nascent RNA for different value of L_1; (D) The probability of $P(n=0)$ and $P(n=1)$ as a function of L_1. Other parameter value are set as: $u=3, v=1, k_1=20, k_0=0$.

In summary, we observe that the memory in both active and inactive period can modulate the steady-state distribution of nascent RNA cooperatively. Increasing the inactive memory L_1 tends to disrupt the zero peak in the stationary nascent RNA distribution, while enhancing the active memory L_2 can induce a bimodal distribution within the bursting regime. Depending on the level of the active and inactive memory, distinct shapes of stationary nascent RNA distribution may emerge, including bimodal distribution with a zero peak and non-zero peak, bimodal distribution with two non-zero peaks, and trimodal distribution, unimodal distribution with zero peak, and unimodal distribution with non-zero peak.

Inferring the memory of promoter switching from a nascent RNA distribution

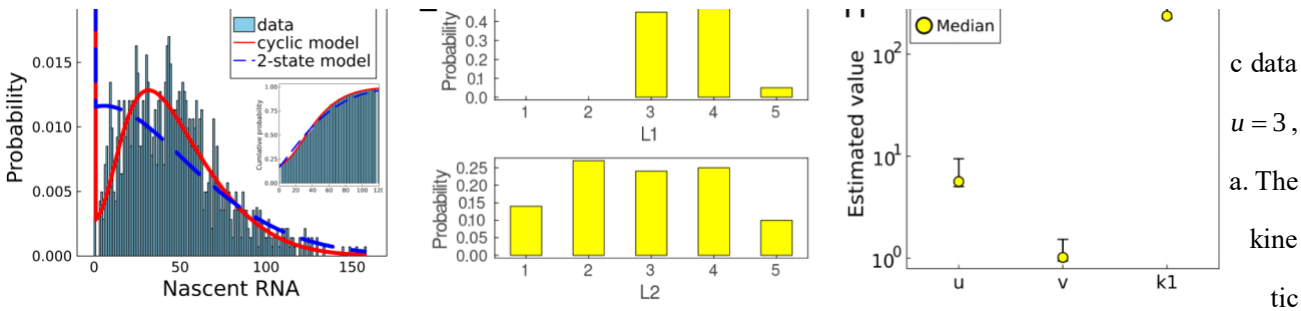
The nascent RNA distribution can be obtained through experimental techniques such as mRNA-FISH labeling or single-cell RNA sequencing (scRNA-seq). An intriguing question arises as to whether the promoter memory can be inferred from the stationary nascent RNA distribution. The question of how model parameters are inferred from a given distribution has been explored in previous studies [44-46]. The two-state non-Markov model has commonly

been employed to fit the nascent RNA distribution and elucidate the the underlying mechanism of gene expression [14,15]. In this section, we proposed a method based on the cyclic model for inferring the promoter memory from the stationary nascent RNA distribution.

We assume that M cells are measured simultaneously at a one-time experiment and cell i has n_i copies of nascent RNA in the steady state, and the stationary distribution of nascent RNA is $P(n,\theta)$. We also assume that the promoter consists of L_1 active states and L_2 inactive states. The transition rates at the ON (OFF) states are the same, i.e., $\omega_{1,2}^{(\mathrm{on})} = \ldots = \omega_{L_1-1,L_1}^{(\mathrm{on})} = \omega_{L_1,1}^{(\mathrm{on})} = L_1 u$, $\omega_{1,2}^{(\mathrm{off})} = \ldots = \omega_{L_2-1,L_2}^{(\mathrm{off})} = \omega_{L_2,1}^{(\mathrm{off})} = L_2 v$. The initiation of nascent RNA only occurs in the active state with rate k_1. Since it is impossible to infer the elongation time and other parameters simultaneously from a stationary distribution, we fix the elongation time $t_{\mathrm{elo}} = 1$. Let $\theta = (L_1, L_2, u, v, k_1)$ be the parameter vector of the cyclic model to be inferred. We define the total likelihood function as

$$\mathrm{L}(\theta) = \prod_{i=1}^{M} P(n_i, \theta),$$

where $P(n_i;\theta)$ is calculated according to Eq. (22). We use an efficient gradient-free optimization algorithm [47] (i.e., adaptive differential optimizer; referring to BlackBoxOptim.jl [ref]) to find the optimal parameters by maximizing the likelihood function $\mathrm{L}(\theta)$.



c data $u = 3$, a. The kine tic parameters estimated by bootstrap resampling for the synthetic data. The red circle represents the true value. (D) The fitted distribution for experimental data of *HIV-1* gene using the cyclic model and the two-state model. The estimated parameters of the cyclic model are: $L_1 = 3$, $L_2 = 4$, $k_1 = 242.66$, $u = 5.12$, $v = 0.95$; the estimated parameters of the two-state model are: $k_1 = 122.21$, $u = 3.41$, $v = 1.50$. The inset figure in (D) is the cumulative probability function that corresponds to distribution estimated by the cyclic and two-state model. (E) the distribution of promoter memory estimated by bootstrap resampling for the experimental data. (H) the 95% confidence interval of kinetic parameters estimated by bootstrap resampling for the experimental data.

To evaluate the accuracy of the method, we generated synthetic data comprising 2000 cells using the stochastic simulation method [ref]. It is worth noting that the parameter set for this synthetic data here was sampled from the

burst regime, satisfying the condition $u \gg v$. By fitting both the cyclic model and classical two-state model to this synthetic data, we observe that the cyclic model outperformed the two-state model in terms of goodness of fit (Fig. 4(A)). In order to assess estimation accuracy, bootstrapping resampling (random sampling samples with replacement) was performed to obtain a distribution of estimated parameters, which can be utilized to estimate the standard errors and confidence interval for these estimated parameters [48]. As depicted in Fig. 4(B), we find that the active memory L_1 can be accurately estimated. Furthermore, Fig.4(C) demonstrate that all true parameters fall within the 95% confidence interval estimated by the bootstrapping, indicating precise inference of true parameters from the stationary nascent RNA distribution in the burst regime. However, extensive numerical simulations revealed limitations in accurately inferring promoter memory solely based on stationary nascent RNA distribution under certain conditions. For instance, when comparing promoter switching rates with elongation process rates—either being sufficiently fast or slow—the nascent RNA distribution becomes independent of promoter memory.

Finally, we apply the aforementioned inference method to a realistic example. Specifically, we aim to infer the promoter memory from the stationary nascent RNA distribution for the *HIV-1* gene in cells expressing a high level of Tat which is measured by Katjana et.al. [20]. It has been reported that the expression of the *HIV-1* gene exhibit bursting dynamic, which are believed to play a crucial role in the control of latency. We fit the experimental data using the two-state model and cyclic model, and present the distribution of estimated parameters derived by bootstrapping resampling in Fig. 4(E)(H). Notably, this analysis reveals that the active memory L_1 is larger than 1. By comparing with Fig. 4(D), it is evident that our proposed cyclic model outperforms its two-state counterpart in terms of fitting accuracy. Furthermore, although inferring an exact distribution for dwell time at the active state remains challenging, our estimations strongly suggest non-exponential behavior for such dwell time.

Conclusions and discussion

In this study, we have derived the stationary nascent RNA distribution for the stochastic gene expression with complex promoter switching. As a special case, we consider a cyclic model where the promoter proceeds sequentially through several irreversible active (ON) states and inactive (OFF) states. The cyclic model can be seen as a extension of the classical two-state model, allowing for non-exponential distribution in dwell times during both active and inactive period. Additionally, our model allows two distinct transcriptional rates, one corresponding to the active periods of the promoter and another corresponding to the inactive periods. Following our previous work, we describe the system using the so-called generalized master equation framework and obtain exact and computationally efficient

solution in Laplace-form and matrix-form for determining stationary nascent RNA distribution.

Using the analytical solution of the cyclic model, we investigated how active and inactive memory affect the shape of the stationary nascent RNA distribution. Previous study[14] have shown that the two-state model with deterministic elongation time can display bimodal distribution with a zero peak and non-zero peak. Our study revealed that increasing inactive memory can destroy the zero peak, resulting in a bimodal distribution with two non-zero peaks. Additionally, we found that increasing the initiation rate $k_0$at the inactive state also eliminates the zero peak. This suggests that the zero peak in stationary nascent RNA distribution is unstable, which may explain why the bimodal distribution is rarely observed in experiments. Furthermore, our study shows that by increasing active memory, delay-induced bimodality or delay-induced zero-inflation phenomenon for nascent RNA can be further enlarged within burst switching regime as previous reported by Qingchao et.al. [17] However, our study has limitations since we only consider a cyclic promoter proceeding sequentially through several irreversible active and inactive state while real-life promoters may have more complex structure such as competitive cross-talking pathways [49] or reversible processes [31]affecting stationary nascent RNA distributions.

Based on the analytical solution of the stationary nascent RNA, we have also developed an effective method to infer the promoter memory and other transcriptional kinetics from a known nascent distribution. Our results indicate that in the burst regime, where the promoter spends most of its time in the inactive period, it is possible to infer promoter memory, particularity active memory, from the stationary nascent RNA distribution. We applied this method to the *HIV-1* gene [20], which is reported to exhibit bursting dynamic. By fitting the experimental data to the cyclic model, we observed that the active memory is larger than 1, suggesting a non-exponential dwell time distribution during activation. However, it should be note that accurately inferring promoter memory from stationary nascent RNA distribution may not always be feasible as different promoters with distinct active and inactive memories can generate identical stationary distribution [31,50]. Further investigation is required to address potential bias in inferring promoter memory from nascent RNA distribution. Additionally, intermediate processes of gene expression such as the partitioning of mRNA/protein during cell division [5,51,52], would contribute noise of gene expression and potentially skew the inferred results; however, these processes are not considered in our study. How details of intermediate processes affect the inferred results is worth further study.

Author contributions

C.S., J.Z., and T.Z. designed the research. C.S. performed the research and analyzed the data. X.Y. analyzed the data. C.S., J.Z., and T.Z. wrote the article.

Acknowledgments

This work was supported by grants 12001129, 12171494, 11931019, 12371483 and 11775314 from the Natural Science Foundation of P. R. China, and by grants 2019B110233002, 202007030004 from Key-Area Research and Development Program of Guangzhou, P.R. China, by grants 2022ZDZX2045 from the Special Projects in Key Fields for Colleges and Universities in Guangdong Province.

Data Availability

The nascent data of *HIV-1* gene for cells expression high level of Tat, were downloaded from [20].

Declaration of interests

The authors declare no competing interests.

References

1. Raj, A., C.S. Peskin, D. Tranchina, D.Y. Vargas, and S. Tyagi. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4:e309.
2. Leloup, J.C., and A. Goldbeter. 2003. Toward a detailed computational model for the mammalian circadian clock. *Proc. Natl. Acad. Sci. U. S. A.* 100:7051–7056.
3. Sepúlveda, L.A., H. Xu, J. Zhang, M. Wang, and I. Golding. 2016. Measurement of gene regulation in individual cells reveals rapid switching between promoter states. *Science*. *351:1218–1222.*
4. Jones, D.L., R.C. Brewster, and R. Phillips. 2014. Promoter architecture dictates cell-to-cell variability in gene expression. Science. 346:1533–1536.
5. Cao, Z., and R. Grima. 2020. Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proc. Natl. Acad. Sci.* 117:4682–4692.
6. Shahrezaei, V., and P.S. Swain. 2008. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci.* 105:17256–17261.
7. Zoller, B., S.C. Little, and T. Gregor. 2017. Diverse spatial expression patterns emerge from common transcription bursting kinetics.
8. Wan, Y., D.G. Anastasakis, J. Rodriguez, M. Palangat, P. Gudla, G. Zaki, M. Tandon, G. Pegoraro, C.C. Chow, M. Hafner, and D.R. Larson. 2021. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell*. 184:2878-2895.e20.
9. Filipowicz, W., S.N. Bhattacharyya, and N. Sonenberg. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* 9:102–114.
10. Halpern, K.B., I. Caspi, E. Elinav, I. Ulitsky, K.B. Halpern, I. Caspi, D. Lemze, M. Levy, S. Landen, E. Elinav, and I. Ulitsky. 2015. Nuclear Retention of mRNA in Mammalian Tissues Report Nuclear Retention of mRNA in Mammalian Tissues. *CellReports*. 13:2653–2662.
11. Zenklusen, D., D.R. Larson, and R.H. Singer. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15:1263–1271.

12. Larson, D.R., D. Zenklusen, B. Wu, J.A. Chao, and R.H. Singer. 2011. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science (80-.).* 332:475–478.
13. Halpern, K.B., S. Tanami, S. Landen, M. Chapal, L. Szlak, A. Hutzler, A. Nizhberg, and S. Itzkovitz. 2015. Bursty gene expression in the intact mammalian liver. *Mol. Cell*. 58:147–156.
14. Xu, H., S.O. Skinner, A.M. Sokac, and I. Golding. 2016. Stochastic kinetics of nascent RNA. *Phys. Rev. Lett*. 117:128101.
15. Fu, X., H.P. Patel, S. Coppola, L. Xu, Z. Cao, T.L. Lenstra, and R. Grima. 2022. Quantifying how post-transcriptional noise and gene copy number variation bias transcriptional parameter inference from mRNA distributions. *Elife*. 11:e82493
16. Szavits-Nossan, J., and R. Grima. 2022. Mean-field theory accurately captures the variation of copy number distributions across the mRNA life cycle. *Phys. Rev. E*. 105:1–15.
17. Jiang, Q., X. Fu, S. Yan, R. Li, W. Du, Z. Cao, F. Qian, and R. Grima. 2021. Neural network aided approximation and parameter inference of non-Markovian models of gene expression. *Nat. Commun.* 12:1–12.
18. Szavits-Nossan, J., and R. Grima. 2023. Steady-state distributions of nascent RNA for general initiation mechanisms. *Phys. Rev. Res*. 5.
19. Dobrzyński, M., and F.J. Bruggeman. 2009. Elongation dynamics shape bursty transcription and translation. *Proc. Natl. Acad. Sci. U. S. A.* 106:2583–2589.
20. Tantale, K., E. Garcia-Oliver, M.C. Robert, A. L'Hostis, Y. Yang, N. Tsanov, R. Topno, T. Gostan, A. Kozulic-Pirher, M. Basu-Shrivastava, K. Mukherjee, V. Slaninova, J.C. Andrau, F. Mueller, E. Basyuk, O. Radulescu, and E. Bertrand. 2021. Stochastic pausing at latent HIV-1 promoters generates transcriptional bursting. *Nat. Commun*. 12:1–20.
21. Filatova, T., N. Popovic, and R. Grima. 2021. Statistics of nascent and mature RNA fluctuations in a stochastic model of transcriptional initiation, elongation, pausing, and termination. *Bull. Math. Biol.* 83:1–62.
22. Choubey, S., J. Kondev, and A. Sanchez. 2015. Deciphering Transcriptional Dynamics In Vivo by Counting Nascent RNA Molecules. *PLoS Comput. Biol*. 11:1–21.
23. Choubey, S. 2018. Nascent RNA kinetics: Transient and steady state behavior of models of transcription. *Phys. Rev. E*. 97:1–11.
24. Chong, S., C. Chen, H. Ge, and X.S. Xie. 2014. Mechanism of transcriptional bursting in bacteria. *Cell*. 158:314–326.
25. Suter, D.M., N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 332:472–474.
26. J. M. G. Vilar and L. Saiz. CplexA: a mathematica package to study macromolecular-assembly control of gene expression. *Bioinformatics* 26:2060-2061 (2010).
27. G. Hornung, R. Bar-Ziv, …, N. Barkai. Noise-mean relationship in mutated promoters. Genome Res. 22:2409-2417 (2012).
28. A. Halme, S. Bumgarner, …, G. R. Fink. Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. *Cell* 116:405-415 (2004);
29. L. M. Octavio, K. Gedeon, and N. Maheshri. Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression. *PLoS Genet*. 5:e1000673 (2009);
30. L. Weinberger, Y. Voichek, …, N. Barkai. Expression noise and acetylation profiles distinguish HDAC functions. *Mol. Cell* 47:193-202 (2012).
31. Zhang, J., L. Chen, and T. Zhou. 2012. Analytical distribution and tunability of noise in a model of promoter progress. *Biophys. J.* 102:1247–1257.

32. Zhou, T., and J. Zhang. 2012. Analytical results for a multistate gene model. *SIAM J. Appl. Math.* 72:789–818.
33. Shi, C., Y. Jiang, and T. Zhou. 2020. Queuing Models of Gene Expression: Analytical Distributions and Beyond. *Biophys. J.* 119:1606–1616.
34. Cao, Z., T. Filatova, D.A. Oyarzún, and R. Grima. 2020. A stochastic model of gene expression with polymerase recruitment and pause release. *Biophys. J.* 119:1002–1014.
35. Pedraza, J.M., and J. Paulsson. 2008. Effects of molecular memory and bursting on fluctuations in gene expression. *Science*. 319:339–343.
36. Yang, X., Z. Wang, Y. Wu, T. Zhou, and J. Zhang. 2021. Kinetic characteristics of transcriptional bursting in a complex gene model with cyclic promoter structure. *bioRxiv*. 2012–2021.
37. Yang, X., S. Luo, Z. Zhang, Z. Wang, T. Zhou, and J. Zhang. 2022. Silent transcription intervals and translational bursting lead to diverse phenotypic switching. *Phys. Chem. Chem. Phys.* 24:26600–26608.
38. Jia, C. 2022. Analytical time-dependent distributions for gene expression models with complex promoter switching mechanisms. 1–28.
39. Adelman, K., A. La Porta, T.J. Santangelo, J.T. Lis, J.W. Roberts, and M.D. Wang. 2002. Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci.* 99:13538–13543.
40. Shi, C., X. Yang, J. Zhang, and T. Zhou. 2023. Stochastic modeling of the mRNA life process: A generalized master equation. *Biophys. J*. 122:4023–4041.
41. Fischer, W., S. Ag, K. Meier-hellstern, and T.B. Laboratories. 1992. The Markov-modulated Poisson process (MMPP) cookbook. 18:149–171.
42. Neuts, M.F. 1979. Versatile Markovian Point Process. *J. Appl. Probab.* 16:764–779.
43. Fu, X., X. Zhou, D. Gu, Z. Cao, and R. Grima. 2022. DelaySSAToolkit. jl: stochastic simulation of reaction systems with time delays in Julia. *bioRxiv*.
44. Luo, S., Z. Wang, Z. Zhang, T. Zhou, and J. Zhang. 2023. Genome-wide inference reveals that feedback regulations constrain promoter-dependent transcriptional burst kinetics. *Nucleic Acids Res.* 51:68–83.
45. Cao, Z., and R. Grima. 2019. Accuracy of parameter estimation for auto-regulatory transcriptional feedback loops from noisy data. *J. R. Soc. Interface*. 16:20180967.
46. Gómez-Schiavon, M., L.F. Chen, A.E. West, and N.E. Buchler. 2017. BayFish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells. *Genome Biol.* 18:1–12.
47. Feldt, R., and A. Stukalov. 2018. Blackboxoptim. jl. *See https//github. com/robertfeldt/BlackBoxOptim. jl*.
48. Endo T, Watanabe T, Yamamoto A. Confidence interval estimation by bootstrap method for uncertainty quantification using random sampling method. *Journal of Nuclear Science and Technology*, 2015, 52(7-8): 993-999.
49. Jiao F, Zhu C. Regulation of gene activation by competitive cross talking pathways. *Biophys. J*, 2020, 119(6): 1204-1214.
50. Kepler, T. B., and T. C. Elston. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81:3116–3136.
51. Beentjes, C. H. L., R. Perez-Carrasco, and R. Grima. 2020. Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics. *Phys. Rev. E.* 101:032403.
52. Skinner, S.O., H. Xu, S. Nagarkar-Jaiswal, P.R. Freire, T.P. Zwaka, and I. Golding. 2016. Single-cell analysis of transcription kinetics across the cell cycle. *Elife*. 5:1–24.