Using average transcription level to understand the regulation of stochastic gene activation

Liang Chen¹²,†, Genghong Lin¹²,† and Feng Jiao¹²

¹Guangzhou Center for Applied Mathematics, and ²School of Mathematics and Information Sciences, Guangzhou University, Guangzhou, People’s Republic of China

†FJ, 0000-0002-5894-8094

Gene activation is a random process, modelled as a framework of multiple rate-limiting steps listed sequentially, in parallel or in combination. Together with suitably assumed processes of gene inactivation, transcript birth and death, the step numbers and parameters in activation frameworks can be estimated by fitting single-cell transcription data. However, current algorithms require computing master equations that are tightly correlated with prior hypothetical frameworks of gene activation. We found that prior estimation of the framework can be facilitated by the traditional dynamical data of mRNA average level \( M(t) \), presenting discriminated dynamical features. Rigorous theory regarding \( M(t) \) profiles allows to confidently rule out the frameworks that fail to capture \( M(t) \) features and to test potential competent frameworks by fitting \( M(t) \) data. We implemented this procedure for a large number of mouse fibroblast genes under tumour necrosis factor induction and determined exactly the ‘cross-talking n-state’ framework; the cross-talk between the signalling and basal pathways is crucial to trigger the first peak of \( M(t) \), while the following damped gentle \( M(t) \) oscillation is regulated by the multi-step basal pathway. This framework can be used to fit sophisticated single-cell data and may facilitate a more accurate understanding of stochastic activation of mouse fibroblast genes.

1. Introduction

Gene transcription is a random process in all genomic loci, wherein messenger RNA (mRNA) molecules for active genes are produced in a bursting fashion in which an episode of transcriptional activity is interrupted by irregular gene inactivation periods [1–3]. A central problem in studying stochastic gene transcription has been understanding the regulation scenarios that control random gene activation (on) and inactivation (off) in response to environmental conditions.
changes [4–7]. The exponentially distributed on period is one of the few universal features of transcription present in both prokaryotic and eukaryotic genes [6, 8, 9]. However, the duration of off state is highly gene-specific, with its transition to on state being modelled as a single rate-limiting step (e.g. figure 1(a)) [1, 3], a single pathway consisting of multiple sequential rate-limiting steps (e.g. figure 1(b)) [10], the parallel competitive rate-limiting pathways (e.g. figure 1(c)) [11] or a combination of both (e.g. figure 1(d)) [12]. These different gene activation frameworks, together with mRNA birth and death processes controlled by single rate-limiting steps, lead to different mathematical models (figure 1). These models have produced many important observations by fitting massive amounts of transcription data at the single-cell level [8, 13, 14].

Real-time imaging of transcriptional bursting makes it possible to count the durations of each gene on and off period along the entire timeline, which generates duration distributions for both gene on and off periods, respectively. For instances of the Escherichia coli P_{lac/ara} promoter [15] and yeast FLO11 genes [16], their on and off periods are all well fitted by single exponential distributions. These observations support the classical two-state model shown in figure 1(a), that the gene for turning genes on and off are all controlled by single rate-limiting biochemical steps [1, 3]. However, the duration of the off state may vary with different genes, manifested by the observed distribution with a unique peak for mouse fibroblast genes [8] and E. coli tetA promoters [17]. The unimodal distribution of gene off duration can be mathematically explained by assuming gene activation frameworks in the three-state model (figure 1(b)), cross-talking pathways model (figure 1(c)) and cross-talking three-state model (figure 1(d)). However, these theoretical approaches cannot determine which framework is the best to describe the observed gene off distribution data, although they have provided good approximations to the downstream distribution of gene transcription [18, 19].

The snapshot data for the distribution histogram of mRNA copy numbers in an isogenic cell population at different time points carry rich, dynamic information on the fluctuations in transcription [20]. When combined with mathematical models, the fit of mRNA (or other RNA types) distribution data has served as a powerful tool for revealing the multi-step regulation of the activation of different genes in bacteria, yeast and human cells [13, 20, 21]. However, the calculation of exact forms of dynamical mRNA distribution requires solving infinite arrays of chemical master equations under the whole parameter region of the models, which is beyond the scope of standard theoretical methods, even for the simplest two-state model (figure 1(a)) [22–24]. The fitting of mRNA distribution data must integrate several computational tools to determine the rate-limiting step numbers in the activation framework and search for suitable system parameters [13, 20]. However, current computation...
algorithms focus only on a class of prior hypothetical multi-step gene activation [13,21,25]; therefore, additional competent frameworks may be ignored. Moreover, some typical dynamical transition patterns among mRNA distribution profiles can be well exhibited by different models [24,26–28], preventing a direct way to rule out models that do not panoramically match the transition patterns of dynamical mRNA distribution.

The steady-state measurement of gene transcription under different cellular conditions has generated a large dataset of mRNA distribution and its mean level \( M \), the Fano factor \( \phi \) (the variance over \( M \)), and noise \( CV^2 \) (\( \phi \) over \( M \)) [1,14]. Under mathematical models, fitting steady-state data has revealed a large spectrum of regulation scenarios that cells use in response to environmental changes [3,14,29]. The steady-state mRNA distributions observed so far can be typically classified into three modes [1]. However, the models in figure 1 can only generate the three distribution modes shown at steady state [10,23,30,31], suggesting that the limited mRNA distribution modalities are insufficient to map reversely onto the diversified frameworks of gene activation. The steady-state data of noise \( CV^2 \), Fano factor \( \phi \) and mean level \( M \), when mapped as scattered points onto \( M-CV^2 \) and \( M-\phi \) planes, provide a diagram of trend lines of \( CV^2 \) and \( \phi \) against \( M \) under varying environments [2,14]. For a given gene of interest in \( E. \) coli, yeast or mammalian cells, the trend lines fitted by different models have revealed distinct regulation scenarios [32]. However, the scenario that plays a dominant role in gene regulation remains elusive.

In contrast to the time-consuming single-cell measurements that require RNA labelling and imaging with high sensitivity and resolution [15,21], the dynamical mRNA average level \( M(t) \) can be relatively easily captured by conventional methods at the cell population level [21,33]. Previous studies have revealed rich temporal profiles of \( M(t) \) for different genes and cellular conditions, such as monotonic increases in the \( E. \) coli promoter \( \text{Plac/ara} \) [15], up-and-down behaviour in the \( c\text{-Fos} \) gene in human osteosarcoma [21], multiple peaks in mouse fibroblast genes [33,34], and even oscillations in yeast stress-induced genes [35]. These observations give rise to the problem of whether such rich dynamical behaviours of \( M(t) \) can be mapped back to the diversified frameworks of gene activation. The key objective is to establish bijections between the dynamical features of \( M(t) \) and the parameter regions for certain mathematical models. This allows us to rule out models that do not capture the exhibited dynamical features of \( M(t) \) and to test the simplest of the remaining models based on their fit to \( M(t) \) data. In this study, we assumed that gene activation is regulated by a combination of sequential and parallel pathways, as shown in figure 1, and we illustrated the way that dynamic mRNA average level data could be used to help estimate the gene activation frameworks.

2. Results

2.1. Cross-talking three-state model

To make the paper easier to follow, we focused on analysing the mRNA average level \( M(t) \) data from a large group of mouse fibroblast genes under cytokine tumour necrosis factor (TNF) stimulation conditions [33,34]. Except for the simple monotonic growth of \( M(t) \) generated by late response genes, the rich non-monotonic behaviours of \( M(t) \) have also been determined, such as up-and-down for the \( Fos \) gene, up-down-up for the \( Cxcl1 \) gene, and damped oscillation with multiple peaks for the \( Nfkbia \) gene.

Theoretical bijections between the dynamical features of \( M(t) \) and three mathematical models were established (table 1) to determine the framework that can efficiently capture the rich transcription dynamics of mouse fibroblast genes. The two-state model shown in figure 1a can only generate monotonic increasing dynamics of \( M(t) \) [32] and thus is not suitable for the discussion of non-monotonic dynamical behaviours. The three-state model shown in figure 1b is proven to display damped oscillatory dynamics of \( M(t) \) under a certain parameter region [36,37]. However, such oscillation behaviour is almost invisible owing to its rapid exponential decay and only slightly decelerates the dynamic increase in \( M(t) \) [37]. The frameworks with two or more parallel pathways can capture the up-and-down dynamics of \( M(t) \) but fail to generate more complex transcription dynamics [32,38,39]. Moreover, the cross-talking pathway model (figure 1c) generates up-and-down \( M(t) \) only when the stronger pathway is frequently selected to activate the gene [32], which is incompatible with the robust up-and-down dynamics of \( M(t) \), even if the TNF induction level is extremely low [34]. Collectively, the activation frameworks of a single pathway or parallel pathways alone are insufficient to capture the rich dynamics of \( M(t) \) from mouse fibroblast genes (table 1).
Table 1. Bijection theory between $M(t)$ dynamical features and system parameter regions for the (a) two-state [32], (b) three-state [36,37], (c) cross-talking pathways [32,38] and (d) multiple pathways models [39]. $\alpha$, $\xi$, $\Lambda$, $\alpha_1$ and $x_1$ are the auxiliary numbers associated with the system parameters, and $\gamma$ and $\delta$ are the gene inactivation rate and mRNA degradation rate, respectively.

| mathematical models | $M(t)$ dynamical profiles | $\Longleftrightarrow$ parameter regions |
|---------------------|---------------------------|-----------------------------------------|
| (a) two-state model  | increase                  | all parameters                          |
| (b) three-state model| increase                  | $\alpha^2 \geq \xi$                    |
|                     | almost increase           | $\alpha^2 < \xi$                       |
| (c) cross-talking pathways model | increase | $\Lambda \geq \min\{\delta, \gamma\}$ |
|                     | up-and-down               | $\Lambda \leq \min\{\delta, \gamma\}$ |
| (d) multiple pathways model | increase  | $x_1 \geq \min\{\delta, \alpha_3\}$    |
|                     | up-and-down               | $x_1 < \min\{\delta, \alpha_3\}$      |

The three-state model generates a damped oscillatory $M(t)$ when $\alpha^2 < \xi$ [36,37]. However, such oscillation decays exponentially and displays visually increasing dynamics [37].

By combining the three-state model (figure 1b) and cross-talking pathways model (figure 1c), it is possible to generate new dynamic $M(t)$ features. The simplest combination is shown in figure 1d, which we call the cross-talking three-state model. This model can be viewed as adding a parallel pathway in the three-state model or decomposing a pathway of the cross-talking pathways model into two sequential steps. We assumed that two competitive pathways activate the gene. These are the weak basal pathway, which has a selection probability $q_1$ and consists of two sequential rate-limiting steps with strength rates $\kappa_1$ and $\kappa_2$, or the strong rate-limiting signalling pathway, which has a selection probability $q_2$ and strength rate $\lambda$, satisfying

$$0 < q_1, q_2 < 1, \quad q_1 + q_2 = 1 \quad \text{and} \quad 0 < \kappa_1, \kappa_2 < \lambda < \infty.$$  

The basal pathway is regulated independently by a spontaneous mechanism to maintain basal transcription levels under normal cellular growth conditions [40,41]. The assumption of two sequential steps and low strength rates for the basal pathway is in close agreement with the real-time imaging data of the off period of 16 mouse fibroblast genes [8]. Moreover, suppose $\kappa_1$ or $\kappa_2$ is relatively large. In that case, the basal pathway can be mathematically approximated by a single rate-limiting step [36], and the framework in figure 1d reduces to the cross-talking pathways model (figure 1c). A stronger signalling pathway is triggered when cells receive external cues, and downstream transcription factors (TFs) are activated by special signal transduction pathways to upregulate gene transcription [34,42]. For each target gene, its activation is ultimately mediated through the binding of downstream TFs in the basal or signalling pathways at the cognate DNA sites in the gene promoter or enhancer domains [5,42]. The selection probabilities $q_1$ and $q_2$ may then quantify the concentration and availability of activated TFs in each pathway to competitively form TF/DNA binding configurations while the inducible activation rate $\lambda$ of the signalling pathway quantifies the binding accessibility and strength between the corresponding TFs and DNA sites [5,12,39,42].

2.2. Dynamics of $M(t)$ and the bijection with parameter regions

To establish the bijection between the $M(t)$ profiles and the parameter regions of the model (figure 1d), we first need to calculate the exact forms of $M(t)$ in terms of system parameters. At time $t \geq 0$, let random variable $X(t) = X = o_1, o_1$, $o_2$, $o_2$, $e$, specify the states of gene off 1 for basal pathway, gene off 1 for signalling pathway, gene off 2, and gene on, respectively. Then define

$$P_{m,X}(t) = \text{Prob}\{\text{the system is residing at state } X \text{ with } m \text{ mRNA molecules at time } t\},$$

and the mass function

$$P_{m}(t) = P_{m,o_1}(t) + P_{m,o_2}(t) + P_{m,e}(t) + P_{m,e}(t), \quad m = 0, 1, \ldots$$

that quantifies the probability of $m$ mRNA transcripts at time $t$ in a single cell.
Following the standard procedure, we can obtain an infinite array of master equations with respect to $P_{m,X}(t)$ \[10,36,43\]

\[
P'_{m,011}(t) = - (\kappa_1 + m\delta) P_{m,011}(t) + (m + 1) \delta P_{m+1,011}(t) + q_1 \gamma P_{m,e}(t),
\]

\[
P'_{m,012}(t) = - (\lambda + m\delta) P_{m,012}(t) + (m + 1) \delta P_{m+1,012}(t) + q_2 \gamma P_{m,e}(t),
\]

\[
P'_{m,o2}(t) = - (\kappa_2 + m\delta) P_{m,o2}(t) + (m + 1) \delta P_{m+1,o2}(t) + \kappa_1 P_{m,011}(t),
\]

\[
P'_{m,e}(t) = - (\gamma + m\delta) P_{m,e}(t) + (m + 1) \delta P_{m+1,e}(t) + \kappa_2 P_{m,o2}(t) + \lambda P_{m,012}(t).
\]

Summing up (2.1)–(2.4) gives the master equation of $P_m(t)$

\[
P'_m(t) = (m + 1) \delta P_{m+1}(t) - \nu P_{m,e}(t) - m \delta P_m(t) + v P_{m-1,e}(t).
\]

We did not focus on solving (2.1)–(2.5), which are beyond the scope of current mathematical methods within all parameter regions \[22–24,26–28\]. However, the master equations set a basis for calculating analytical forms for the gene state probabilities $P_{011}, P_{012}, P_{o2}, P_e$ and mean transcript level $M(t)$, defined as

\[P_X(t) = \sum_{m=0}^{\infty} P_{m,X}(t), \quad X = 011, 012, o2, o_e \quad \text{and} \quad M(t) = \sum_{m=0}^{\infty} m P_m(t).\]

Using these definitions and (2.1)–(2.5), we derived the equations for the four state probabilities and the mean transcription level

\[
\begin{align*}
P'_{011}(t) &= - \kappa_1 P_{011}(t) + q_1 \gamma P_e(t), \\
P'_{012}(t) &= - \lambda P_{012}(t) + q_2 \gamma P_e(t), \\
P'_{o2}(t) &= \kappa_1 P_{011}(t) - \kappa_2 P_{o2}(t), \\
P'_e(t) &= \lambda P_{o2}(t) + \kappa_2 P_{o2}(t) - \gamma P_e(t)
\end{align*}
\]

and

\[M(t) = \nu P_e(t) - \delta M(t).\]

Because the system must reside on exactly one gene state at any time, we set an arbitrary initial condition for (2.6)

\[P_{011}(0) + P_{012}(0) + P_{o2}(0) + P_e(0) = 1 \quad \text{and} \quad M(0) \geq 0.\]

We utilized the Laplace transform method to solve the first-order differential system (2.6) and (2.7) (electronic supplementary material). We defined a polynomial function as

\[h(x) = \left[ (\kappa_1 \lambda + \kappa_2 \lambda + \kappa_1 \kappa_2) P_e(0) + \kappa_1 \kappa_2 P_{011}(0) + \lambda (\kappa_1 + \kappa_2) P_{012}(0) + \kappa_2 (\lambda + \kappa_1) P_{o2}(0) \right] x \]

\[+ P_e(0) x^3 + \left[ (\kappa_1 + \lambda + \kappa_2) P_e(0) + \lambda P_{o12}(0) + \kappa_2 P_{o2}(0) \right] x^2 + \kappa_1 \kappa_2 \lambda,\]

and two auxiliary numbers $c_1$ and $c_2$ in terms of the system parameters (electronic supplementary material, (7) and (8)). It can be verified that zero is a simple eigenvalue of the coefficient matrix for the system of the first four equations in (2.6). The other non-zero eigenvalues $\lambda_1$, $\lambda_2$ and $\lambda_3$ are calculated in terms of $c_1$, $c_2$ and system parameters (electronic supplementary material, (4)–(6)). Under the arbitrary initial condition (2.7), the average mRNA level $M(t)$ is found to be as follows:

1. If $c_1^2 < c_2, \lambda_1, \lambda_2$ and $\lambda_3$ are real numbers with $0 < \lambda_1 < \lambda_2 < \lambda_3$ (electronic supplementary material, (11)), and $M(t)$ takes the form of

\[
M(t) = \frac{\nu \kappa_1 \kappa_2 \lambda}{\delta \lambda_1 \lambda_2 \lambda_3} - \frac{\nu h(-\lambda_1) e^{-\lambda_1 t}}{\lambda_1 (\lambda_2 - \lambda_1) (\lambda_3 - \lambda_1) (\delta - \lambda_1)}
\]

\[+ \frac{\nu h(-\lambda_2) e^{-\lambda_2 t}}{\lambda_2 (\lambda_1 - \lambda_2) (\lambda_3 - \lambda_2) (\delta - \lambda_2)} - \frac{\nu h(-\lambda_3) e^{-\lambda_3 t}}{\lambda_3 (\lambda_1 - \lambda_3) (\lambda_2 - \lambda_3) (\delta - \lambda_3)}
\]

\[+ \frac{\nu h(-\delta) e^{-\delta t}}{\delta (\delta - \lambda_1) (\delta - \lambda_2) (\delta - \lambda_3)} + M(0) e^{-\delta t}.\]
We characterized the dynamic profiles of only the newly produced mRNA molecules. This gives the following initial values:

\[
M(t) = \frac{v}{\delta} \frac{k_1 k_2 \lambda}{a_1 a_2} (1 - e^{-\delta t}) - \frac{v h(-a_1)(e^{-a_1 t} - e^{-\delta t})}{a_1(\delta - a_1)(a_2 - a_1)^2} + \frac{v h(-a_2)e^{-a_2 t}}{a_2(a_2 - a_1)(\delta - a_2)}
\]

\[
+ \frac{v h(-\delta)(e^{-\delta t} - e^{-a_2 t})}{\delta(\delta - a_1)(\delta - a_2)^2} + M(0)e^{-\delta t}.
\]

(2.9)

If \( c_1^2 > c_2, a_1 > 0 \) is a real number, whereas \( a_2 \) and \( a_3 \) are conjugate complexes, then let \( a_r = \text{Re}(a_3) \) and \( a_i = \text{Im}(a_3) \) (electronic supplementary material, (15)). Then, \( M(t) \) takes the form of

\[
M(t) = \frac{v}{\delta a_1(a_2^2 + a_r^2)} - \frac{v h(-a_1)e^{-a_1 t}}{a_1(\delta - a_1)(a_1 - a_r)^2 + a_r^2} + \frac{v h(-\delta)e^{-\delta t}}{\delta(\delta - a_1)(\delta - a_1)^2 + a_r^2}
\]

\[
+ A \sqrt{1 + A^2 \cos(a_3 t + \theta)} e^{-a_3 t} + M(0)e^{-\delta t},
\]

where the constants \( \bar{A}, A \) and \( \theta \) correlate with the parameters and initial conditions (electronic supplementary material, (17)).

We started with an interesting case, \( c_1^2 > c_2 \). Then, \( M(t) \) was expressed in exact form (2.10), which contained a cosine function, and suggested possible oscillatory dynamics of \( M(t) \). However, the coefficient of the cosine function damped exponentially, which drastically weakened the oscillation visually, as manifested by our numerical examples and observations from the three-state model [37].

For the other case of \( c_1^2 \leq c_2 \), the eigenvalues \( a_1, a_2 \) and \( a_3 \) are real, and the exact forms (2.8) and (2.9) of \( M(t) \) do not contain oscillatory functions but contain multiple exponential functions. The case \( c_1^2 = c_2 \) rarely occurs in biology and can be viewed mathematically as a limiting case of \( c_1^2 < c_2 \). For \( c_1^2 < c_2 \), a rigorous statement of the \( M(t) \) profiles is inevitably technical. Let \( x_1 \) and \( x_2 \) denote the two roots of

\[
H(x) = vq_2 \lambda x^2 + v(q_1 k_1 k_2 + q_2 \lambda k_1 + q_3 \lambda k_2)x + v k_1 k_2 \lambda.
\]

When both \( x_1 \) and \( x_2 \) are complex numbers, there are a total of four parameter correlations, and we showed that \( M(t) \), expressed by (2.8), either increases monotonically for all \( t > 0 \), or develops an up-down-up profile. If \( x_1 \) and \( x_2 \) are real values, we can classify all 60 correlations among \( x_1, x_2, a_1, a_2, a_3 \) and \( \delta \) into three categories that correspond to three distinct dynamical behaviours: the increasing, up-and-down and up-down-up profiles of \( M(t) \). We illustrated detailed mathematical results and their proof in appendix theorem A.2. In summary, for the cross-talking three-state model (figure 1d), even if \( M(t) \) is expressed in different exact forms (2.8)–(2.10) and influenced by various parameter correlations, \( M(t) \) can exhibit, but exhibits only three features: increasing, up-and-down and up-down-up dynamics.

2.3. Fitting dynamical transcription data of mouse fibroblast genes

We demonstrated three dynamical profiles of \( M(t) \) generated by the cross-talking three-state model (figure 1d). These distinct behaviours correspond with the observed dynamic trends of average mRNA levels in mouse fibroblast genes in response to TNF [33,34]. For instance, Hao & Baltimore [33] divided 180 activated mouse fibroblast genes under TNF induction into three groups, separately characterized by the short, median and long half-lives of the transcripts. As shown in figure 2, they found that the group I genes responded quickly by forming a sharp dynamical peak at average transcription levels; group II genes did not respond quickly. However, most still formed a gentle transcription peak along the timeline, and group III mRNAs accumulated rather slowly and gradually increased in abundance during the observation window. The transcription data from the 12
representative mouse fibroblast genes shown in figure 2 contains four genes displaying the up-down-up trend of transcription dynamics (Edn1, Cxcl1, Ccl2, Icam1), as well as others displaying either dynamical up-and-down or monotonic increasing transcription.

The fit of $M(t)$ data in the case of measurement noise can be achieved by minimizing weighted least squares $J(\theta)$ with parameter values $\theta$, an objective function indicating the difference between experiment and simulation [44]. For the sake of simplicity, we assumed a standard normally distributed measurement noise and then $J(\theta)$ is given by

$$J(\theta) = \sum_{j=1}^{N} (M_j - M(t_j, \theta))^2.$$ 

Here $N$ is the number of measurement time points, $M_j$ is the experimental data at time $t_j$, and $M(t_j, \theta)$ is the simulated $M(t)$ at $t = t_j$ obtained by using exact forms (2.8)–(2.10), or by the efficient computational methods on master equations (2.1)–(2.4) [25] or ordinary differential equations (2.6) and (2.7) [44]. To reduce the dimension of parameter space, we assigned extremely small values to the activation rates $\kappa_1$ and $\kappa_2$ of the weak basal pathway. The initial values of the other parameters that initialize the minimization runs of $J(\theta)$ are randomly sampled as in [44]. For the sake of simplicity, we assumed a standard normally distributed measurement noise and then $J(\theta)$ is given by

$$J(\theta) = \sum_{j=1}^{N} (M_j - M(t_j, \theta))^2.$$ 

Here $N$ is the number of measurement time points, $M_j$ is the experimental data at time $t_j$, and $M(t_j, \theta)$ is the simulated $M(t)$ at $t = t_j$ obtained by using exact forms (2.8)–(2.10), or by the efficient computational methods on master equations (2.1)–(2.4) [25] or ordinary differential equations (2.6) and (2.7) [44]. To reduce the dimension of parameter space, we assigned extremely small values to the activation rates $\kappa_1$ and $\kappa_2$ of the weak basal pathway. The initial values of the other parameters that initialize the minimization runs of $J(\theta)$ are randomly sampled as in [44]. As shown in figure 2, the cross-talking three-state model provides good theoretical fits to all 12 datasets ($R^2 > 0.9$). The fitted system parameters are listed in electronic supplementary material, table S1.

While fitting the data shown in figure 2, we found that the estimated mRNA degradation rate $\delta$ of each gene falls within the $\delta$ region of the gene group to which they belong [33] (electronic supplementary material, table S1).
negatively correlated with gene groups I, II and III. The probability of the peak and suppressing the stationary values of $M(t)$ change from monotonic to up-down-up mode. An increase in the non-monotonic modes are displayed within most of the observation suggests that the simultaneous large in units of gene groups (electronic supplementary material, table S1). Different gene groups exhibit distinct temporal transcription $\delta$ supplementary material, table S1). Therefore, the fitted $\delta$ values exhibited a significant negative correlation with gene groups I, II and III because the gene groups themselves were classified by their transcript half-lives (figure 3a). Intriguingly, the freely fitted inactivation rate $\gamma$ and activation rate $\lambda$ of the signalling pathway also exhibited a negative correlation with gene groups (figure 3a). This observation suggests that the simultaneous large $\lambda$ and $\gamma$ may separately help in increasing the height of the peak and suppressing the stationary values of $M(t)$ in group I. By contrast, the small $\lambda$ and $\gamma$ implemented contrary functions that destroy the dynamical peak of $M(t)$ and lift the stationary mRNA numbers in group III. However, the probability $q_2$ of the signalling pathway did not correlate with the gene groups but varied for different genes (figure 3a). Note that the dynamics of $M(t)$ in the three gene groups are primarily discriminated by its first peak. Our observations suggest that, once cells receive external cues, the frequency of the signalling pathway directing gene activation may not play a crucial role in regulating the dynamical peak of transcription level.

To further understand the regulation of $M(t)$ profiles, we varied the parameters $\delta$, $\gamma$, $\lambda$ and $q_2$ under the fitted parameter sets of all 12 genes in figure 2. This procedure reveals a uniform regulation mode for each system parameter. As shown in figure 3b for Cxcl1 gene, the variation in $\delta$ behaves as a bilateral switch to regulate $M(t)$: there is a threshold value such that $M(t)$ increases monotonically. By contrast, $\delta$ stays below the threshold but switches to a non-monotonic profile once $\delta$ exceeds the threshold. Such bilateral regulation of $\delta$ has been observed to play an important role in controlling the temporal transcription mode in mouse fibroblasts [33,34]. In addition, both $\gamma$ and $\lambda$ play the same bilateral roles in regulating $M(t)$ dynamics (figure 3b), reinforcing previous observation of smaller $\delta$, $\gamma$, $\lambda$ for group

Figure 3. Regulation of transcription dynamics by system parameters. (a) Fitted parameters for 12 mouse fibroblast genes are listed in units of gene groups (electronic supplementary material, table S1). Different gene groups exhibit distinct temporal transcription modes [33]. The values of mRNA degradation rate $\delta$, inactivation rate $\gamma$, and activation rate $\lambda$ of the signalling pathway are all negatively correlated with gene groups I, II and III. The probability $q_2$ of the signalling pathway does not follow a clear correlation with the gene groups. (b) Based on the fitted parameter set for the Cxcl1 gene, increasing $\delta$, $\gamma$ and $\lambda$ transit transcription dynamics from monotonic to up-down-up mode. An increase in $q_2$ generates switches among multiple transcription dynamical modes, where the non-monotonic modes are displayed within most of the $q_2$ variation region (0, 1).
parameters in (b) the oscillatory trend of transcription data (black circles) for the mouse fibroblast enhances the following damped oscillatory behaviour. (limiting steps with probability $q$ in the basal pathway has almost no impact on the initial quick up-and-down dynamics of $M(t)$. However, exceptions need to be made for cases where transcription level increase in the step number prolongs the initial response lag. In addition, it enhances the damped oscillation of the dynamical peaks of $n$.

III genes that generate increasing $M(t)$ with larger $\delta$, $\gamma$, $\lambda$ for groups I and II genes that exhibit non-monotonic transcription dynamics (figure 3a). The regulation scenario of $q_2$ is different because it generates multiple switches among distinct $M(t)$ profiles, when $q_2$ increases from 0 to 1 (figure 3b). Exceptions need to be made for cases where $q_2$ approaches 0 or 1, which generates increasing transcription dynamics, because the dynamical peak of $M(t)$ seems to be robust in almost all the various regions of $q_2 \in (0, 1)$. Note that $q_2$ is closely related to signal strength. Our observations fit with the ubiquitous transcription dynamical peak of mouse fibroblast genes under TNF induction from the lowest to the highest levels [34].

2.4. Cross-talking $n$-state model for oscillatory transcription dynamics

Our bijection theory shows that the cross-talking three-state model (figure 1d) cannot generate multiple dynamical peaks of $M(t)$. Therefore, the model (figure 1d) can be ruled out when $M(t)$ exhibits oscillation [33,35]. We focused on transcription with constant kinetic rates under stable inductions to avoid complicated $M(t)$ dynamics generated by time-dependent rates under time-varying signals [45] or feedback regulations [46]. The bijection theory (table 1) shows that multiple parallel pathways induce at most one dynamical peak, and therefore introducing more parallel pathways in the model may not capture oscillatory $M(t)$. We then considered decomposing the basal pathway of the model (figure 1d) into multiple sequential steps, as shown in figure 4a for the cross-talking $n$-state model. The calculation of $M(t)$ can follow the same procedure for the model of figure 1d, which relies on solving a system of differential equations for which the coefficient matrix may have multiple pairs of eigenvalues expressed by conjugate complexes (electronic supplementary material). According to the classical theory of ordinary differential equations, the exact forms of $M(t)$ may contain multiple periodic cosine functions, and it is plausible to visualize oscillatory dynamics.
To test \( M(t) \) oscillation induced by the sequential multi-step gene activation, we compared the \( M(t) \) profiles for different step numbers. This procedure requires that all comparisons are restricted to a constant average duration \( T_{\text{off}} \) of the gene off state \([10,37]\). For the cross-talk \( n \)-state model (figure 4a), \( T_{\text{off}} \) is given by

\[
T_{\text{off}} = q_1 \sum_{i=1}^{n-1} \frac{1}{\kappa_i} + \frac{q_2}{\lambda}, \quad n \geq 2.
\]

To guarantee the unchanged \( T_{\text{off}} \), we set three parameter scaling conditions where activation rates \( \kappa_1, \ldots, \kappa_{n-1} \) are separately scaled identically \([17,47]\), differently \([8,17]\) and alternatively \([17]\)

- (C1) : \( \kappa_1 = \kappa_2 = \cdots = \kappa_{n-1} \), with \( \kappa_1 = \frac{q_1(n-1)}{T_{\text{off}} - q_2/\lambda} \);
- (C2) : \( \kappa_2 = 2\kappa_1, \kappa_3 = 3\kappa_1, \ldots, \kappa_{n-1} = (n-1)\kappa_1 \), with \( \kappa_1 = \frac{1 + (1/2) + \cdots + (1/(n-1))}{T_{\text{off}} - q_2/\lambda} \);
- (C3) : \( \kappa_1 = \kappa_3 = \cdots, \) and \( \kappa_2 = \kappa_4 = \cdots = 3\kappa_1 \), with \( \kappa_1 = \begin{cases} \frac{2q_1(n-1)}{3T_{\text{off}}-q_2/\lambda}, & n \text{ is odd,} \\ \frac{q_1(2n-1)}{3T_{\text{off}}-q_2/\lambda}, & n \text{ is even.} \end{cases} \)

We initially examined the case of \( q_2 = 0 \), for which a single multi-step pathway directs gene activation \([10,17,27,28]\). Under parameter scaling condition (C1), we generated several \( M(t) \) curves under different activation step numbers, using numerical simulations from the corresponding system of differential equations (electronic supplementary material, (19)). As shown in figure 4b under the fitted parameters of the \( \text{Cxcl1} \) gene in figure 2, multi-step gene activation for large step numbers triggered significantly damped oscillations \( M(t) \), where the significance of the oscillation is positively correlated with the step number. However, the system displayed lag times of more than 8 h to reach the first peak of \( M(t) \). This slow transcription response contradicts the rapid peak of \( M(t) \) within 0.5–2 h for mouse fibroblast genes \([33,34]\). Moreover, the damped oscillation and response lag of \( M(t) \) were robust against the parameter scaling conditions (C2) and (C3) (electronic supplementary material, figure S2).

Collectively, the large number of sequential gene activation steps facilitates the oscillatory dynamics of \( M(t) \). At the same time, a single multi-step pathway could not induce quickly peaked transcription dynamics of mouse fibroblast genes.

We then quantified \( q_2 \neq 0 \) to introduce the cross-talking regulation of pathways on \( M(t) \) dynamics. We generated \( M(t) \) curves (electronic supplementary material, (18)) under the fitted parameters of the \( \text{Cxcl1} \) gene in figure 2 and conditions (C1)–(C3). As shown in figure 4c and electronic supplementary material, figure S3, the system generates oscillatory \( M(t) \) with two major features. Firstly, \( M(t) \) displays a quick and sharp first peak within the initial time region. In contrast, the height and sharpness of the first peak are nearly independent of the step number and parameter scaling conditions of the basal pathway. Secondly, \( M(t) \) displays a damped and gentle second and following peaks that are tightly correlated with the step number and parameter conditions, similarly to the oscillation of \( M(t) \) induced by a single multi-step pathway (figure 4b). These two features capture multiple dynamical peaks for the transcription of the mouse fibroblast \( \text{Nfkbia} \) gene \([33]\) (figure 4d). Collectively, the cross-talk between signalling and basal pathways plays a dominant role in generating the first quick up-and-down transcription dynamics. By contrast, the subsequent gentle and damped transcription oscillation is induced by the multi-step regulation in the weak basal pathway.

### 2.5. Cross-verification with single-cell measurements

Our method requires rich \( M(t) \) data dynamics to rule out frameworks that cannot capture the dynamic features sufficiently. When \( M(t) \) data displays simple monotonic dynamics, all the gene activation frameworks in figure 1 may facilitate good fit to \( M(t) \) data. Therefore, we cannot rule out any activation framework by capturing the monotonic \( M(t) \) dynamics. To solve this dilemma, we may introduce some single-cell transcription data which display smooth trend lines along the timeline, such as the noise CV\(^2\)(t), Fano factor \( \phi(t) \) and probability \( P_q(t) \) of the gene producing zero transcripts. These indexes may be used in conjunction with \( M(t) \) to cross-verify the gene activation frameworks \([25]\).
The classical two-state model used in single-cell studies posits that a gene will randomly transition between on (active) and off (inactive) states, with mRNA molecules being only produced when the gene is on (figure 1a). Compared to the universal feature of a single rate-limiting step turning gene off [6,8,9], the process of turning a gene on is typically influenced by multiple rate-limiting fluctuations [6,7,9]. The framework of gene activation has been modelled by listing rate-limiting steps sequentially [4,8,17,29] (figure 1b), in parallel [11,38,39] (figure 1c), or in the form of their combinations [12] (figure 1d). These activation frameworks, coupled with simple gene inactivation, mRNA birth and death processes, lead to different mathematical models of stochastic gene transcription (figure 1).

Recent studies have facilitated efficient computational method for fitting the transcription distribution data to search the optimal number of rate-limiting steps in gene activation and their kinetic parameters.
The workflow of these methodologies typically starts with generalized chemical master equations for which a single kinetic rate controls each reaction [13,25]. This assumption implicitly suggests a single pathway with sequential rate-limiting steps directing the gene activation (figure 1a,b), while ignores cross-talking parallel pathways (figure 1c,d) for which the kinetic rates in master equations may be multiplied by pathway selection probabilities $q_1$ and $q_2$ (equations (2.1) and (2.2)).

We noticed several controversial conclusions on the gene activation obtained by fitting transcription distribution data. For example, using the two-state model (figure 1a) led to an acceptable fit to dynamical mRNA distribution data for mammalian c-Fos gene [21]. However, the mRNA average level $M(t)$, extracted from the distribution data, exhibited a significant up-and-down dynamics [21] which cannot be captured by the two-state model as it only generates increasing $M(t)$ (table 1). By contrast, the cross-talking pathway model (figure 1c) can generate up-and-down $M(t)$ (table 1) and thus may serve as a better candidate model for the c-Fos gene. Another controversy arose from unimodal distribution data of off duration for 16 mouse fibroblast genes which have been fitted by the three-state model (figure 1b) [8]. However, the three-state model may not be appropriate as it fails to capture non-monotonic dynamical $M(t)$ data (table 1) for a large number of mouse fibroblast genes [33,34], whereas the cross-talking three-state model (figure 1d) provides good fit to those $M(t)$ data (figure 2d). Therefore, it is crucial to determine the confidential gene activation framework in parallel pathways and multiple steps before attempting to fit single-cell dynamical transcription data computationally.

This study demonstrated that the transcript average level $M(t)$ is competent for prior estimation of gene activation framework. Compared to the multiple uneven mRNA distribution profiles at discrete time points [13,20], a single smooth curve of $M(t)$ along the timeline presents easily discriminated dynamic features. Subsequently, frameworks that cannot capture the exhibited $M(t)$ dynamical features can be ruled out, while the other potential frameworks are further determined by fitting $M(t)$ data. The calculation of $M(t)$ and its fitting to the data under arbitrary initial conditions (equation (2.7)) are standard using ordinary differential equations [44] or master equations [25]. The challenge is determining whether the considered framework can or cannot display the exhibited $M(t)$ features. Solving this challenge requires rigorous but tedious mathematical analysis to establish bijection theory (table 1, appendix theorems A.1 and A.2) that shows one-to-one correspondence between parameter regions and $M(t)$ dynamical features for each activation framework of the model. We noticed that the initial condition of the system could significantly influence $M(t)$ dynamics [49]. Here we restricted our analysis under the condition of zero transcripts at time $t=0$ (equation (2.11)), consistent with the basal expression level of genes under normal cellular growth conditions before adding external inductions [15,21,33].

We illustrated our idea using $M(t)$ for many TNF-induced mouse fibroblast genes. These genes display rich transcription dynamics that can be categorized into three main features: increasing, up- and-down and up-down-up profiles of $M(t)$ [33,34]. Our bijection theories (table 1 and appendix theorems A.1 and A.2) show that these distinct dynamics cannot be achieved by the sequential or parallel rate-limiting steps alone but can be captured by the simplest form of combined sequential and parallel steps (figure 2). We call this the cross-talking three-state framework, as depicted in figure 1d, for which the gene is activated either by the weak basal pathway consisting of two sequential steps or by the strong signalling pathway. Furthermore, analysis of the freely fitted parameters of the cross-talking three-state model reveals regulation scenarios of combined gene activation, inactivation and mRNA degradation to navigate monotonicity of $M(t)$ dynamics (figure 3; electronic supplementary material, figure S1 and table S1). These observations may facilitate answering the proposed question in [33] of how the transcriptional control cooperates with transcript stability to determine the kinetics of TNF-induced gene expression.

We note that a small number of mouse fibroblast genes display damped transcription oscillation, with the first peak forming rapidly within the initial period [33,34]. We first ruled out the widely used framework of a single multi-step pathway as it triggers a long lag reaching the first peak of $M(t)$ (figure 4b; electronic supplementary material, figure S2), which contradicts the observed rapid peak of $M(t)$. Also, the cross-talking three-state framework cannot generate oscillatory dynamics of $M(t)$ (appendix theorems A.1 and A.2). However, when we developed a cross-talking $n$-state framework by decomposing the basal pathway of a cross-talking three-state framework into multiple steps (figure 4a), the oscillation appeared, and the first peak of $M(t)$ formed quickly (figure 4c; electronic supplementary material, figure S3). Together with the good fit to transcription data (figure 4d), we confirmed that the cross-talk between pathways is crucial to trigger the first rapid, sharp peak of $M(t)$, while the multi-step regulation facilitates the following damped and gentle oscillatory dynamics, for mouse fibroblast genes.
Our conclusions suggested that all the representative datasets on $M(t)$ dynamics of mouse fibroblast genes can be universally fitted by the cross-talking n-state ($n \geq 3$) activation framework. By contrast, the other simpler frameworks fail to capture several $M(t)$ dynamical features. The cross-talking n-state framework may also describe similar non-monotonic transcription dynamics of $c$-Fos genes after serum induction [21] or the genes in the innate immune system of insects when fighting pathogen invasions [50]. Therefore, before the fit of sophisticated distribution data for those genes on off duration or mRNA copy numbers, the cross-talking n-state framework should be introduced into the master equations as in equations (2.1) and (2.2). Otherwise, the computational methods may only estimate optimal parameters and step numbers using an inappropriate gene activation framework.

Our procedure relies on the rich dynamics of $M(t)$. When $M(t)$ behaves monotonically, the other transcription indexes at single-cell level are required to cross-verify the activation frameworks. For instance, the dynamical Fano factor or the probability of no transcript being produced may help discriminate frameworks regulated by a single pathway or cross-talking pathways (figure 5). Future work may use additional data for different genes and transcription indexes to test and develop our procedure. In addition, we noted the masking effect of small mRNA degradation rate on estimates of optimal parameters and step numbers using an inappropriate gene activation framework.

Appendix A. Mathematical analysis of $M(t)$ dynamics for cross-talking three-state model

In this section, we use exact forms (2.8)–(2.10) of $M(t)$ to understand its dynamical features. We consider the initial condition (2.11) for which the gene is silent, and there is almost no transcript at the initial time [15,21,33]. The following two theorems illustrate the cases of the initial condition (2.11) for which the gene is silent, and there is almost no transcript at the initial time.

We divided the discussion into two cases of $r$ and $\alpha$ for $M(t)$ displays only one of increasing, up-and-down and up-down-up dynamical profiles.

Proof. The derivative of $M(t)$ gives

$$M'(t) = -a r e^{-at} - \beta c e^{-\beta t} - r c e^{-rt} \implies M'(t)e^{rt} = -[ae^{(r-\alpha)t} + \beta c e^{(r-\beta)t}].$$

We divided the discussion into two cases of $a > 0$ and $a < 0$.

(i) If $a > 0$, then $M'(t) < 0$ by noting that the sign of $M'(t)$ is dominated by the term $-ae^{-at}$ when $t \to \infty$. Taking the derivative of $M'(t)e^{rt}$ in (A 1) gives

$$M'(t)e^{rt} = -a(r-\alpha)e^{(r-\alpha)t}M_1(t), \quad \text{with } M_1(t) = 1 + \frac{\beta c e^{(r-\beta)t}}{ae^{(r-\alpha)t}}.$$ (A 2)

If $b \geq -a(r-\alpha)/\beta(r-\beta)$, (A 2) indicates

$$M_1(t) > 0 \implies [M'(t)e^{rt}]' < 0 \implies M'(t)e^{rt} \text{ decreases}, \quad \text{for } t > 0.$$

There are two probabilities of $M'(0) > 0$ and $M'(0) = 0$. If $M'(0) > 0$, together with $\lim_{t \to \infty} M'(t) < 0$, there exists $t_1 > 0$ such that

$$M'(t)e^{rt} > 0, \quad \text{for } t \in (0, t_1), \quad M'(t_1)e^{rt_1} = 0 \quad \text{and } \quad M'(t)e^{rt} < 0, \quad \text{for } t > t_1.$$
Therefore, \( M(t) \) increases for \( t \in (0, t_1) \) and decreases for \( t > t_1 \) which presents up-and-down dynamics. If \( M'(0) = 0, M'(t)e^{rt} < 0 \) for \( t > 0 \), and thus \( M(t) \) decreases with \( M(t) < 0 \) all the time. This contradicts to the assumption of \( M(t) \geq 0 \).

If \( b < -a(a(r-a)/\beta(r-\beta), (A2) \) suggests that there is \( t_2 > 0 \) such that
\[
M_1(t) < 0, \quad t \in (0, t_2), \quad M_1(t_2) = 0 \quad \text{and} \quad M_1(t) > 0, \quad t > t_2,
\]
where
\[
t_2 = \frac{1}{\alpha - \beta} \ln \left( \frac{-a(a(r-a)}{b\beta(r-\beta)} \right).
\]

Then, \( A(2) \) suggests that \( [M'(t)e^{rt}] < 0 \) for \( t \in (0, t_2) \) and \( [M'(t)e^{rt}] < 0 \) for \( t > t_2 \). Therefore, \( M'(t)e^{rt} \) increases in \((0, t_2)\) while decreases in \((t_2, \infty)\). As \( M'(0) \geq 0 \) and \( \lim_{t\to\infty} M'(t) < 0 \), there exists \( t_3 \) such that \( M'(t_3)e^{rt_3} = 0 \). Thus \( M'(t)e^{rt} > 0 \) for \( t \in (0, t_3) \) and \( M'(t)e^{rt} < 0 \) for \( t > t_3 \). This indicates that \( M(t) \) increases for \( t \in (0, t_3) \) and decreases for \( t > t_3 \), which is the up-and-down dynamics.

(ii) If \( a < 0 \), the \( A(1) \) indicates that \( \lim_{t\to\infty} M'(t) > 0 \). If \( b < -a(a(r-a)/\beta(r-\beta), (A2) \) gives
\[
M_1(t) > 0 \implies [M'(t)e^{rt}] > 0 \implies M'(t)e^{rt} \quad \text{for} \quad t > 0.
\]

As \( M'(0) \geq 0 \), we have \( M'(t)e^{rt} \geq 0, t > 0 \) which suggests that \( M(t) \) increases for all the time.

Contrastingly, if \( b > -a(a(r-a)/\beta(r-\beta), \) from \( A(2) \) we find that there exists \( t_4 > 0 \) such that
\[
M_1(t) < 0, \quad t \in (0, t_4), \quad \quad M_1(t_4) = 0 \quad \text{and} \quad M_1(t) > 0, \quad t > t_4,
\]
where
\[
t_4 = \frac{1}{\alpha - \beta} \ln \left( \frac{-a(a(r-a)}{b\beta(r-\beta)} \right).
\]

Together with \( A(2) \), this further leads to \([M'(t)e^{rt}] < 0 \) for \( t \in (0, t_4) \) and \([M'(t)e^{rt}] > 0 \) for \( t > t_4 \). Therefore, \( M'(t)e^{rt} \) decreases in \((0, t_4)\) while increases in \((t_4, \infty)\). As \( M'(0) \geq 0 \), there are two possibilities of \( M'(t_4)e^{rt_4} \geq 0 \) and \( M'(t_4)e^{rt_4} < 0 \). For \( M'(t_4)e^{rt_4} \geq 0 \), the fact of \( \lim_{t\to\infty} M'(t) > 0 \) indicates \( M'(t)e^{rt} \geq 0 \) for all \( t \geq 0 \) and thereby \( M(t) \) increases all the time. If it is \( M'(t_4)e^{rt_4} < 0 \), there exists \( 0 < t_5 < t_4 < t_6 \) such that \( M'(t_5)e^{rt_5} = M'(t_6)e^{rt_6} = 0 \) and
\[
M'(t)e^{rt} > 0, \quad t \in (0, t_5), \quad M'(t)e^{rt} < 0, \quad t \in (t_5, t_6), \quad \text{and} \quad M'(t)e^{rt} > 0, \quad t > t_6.
\]

This indicates that \( M(t) \) displays an up-down-up dynamics that \( M(t) \) increases for \( t \in (0, t_5) \cup (t_6, \infty) \) and decreases within \( t \in (t_5, t_6) \). The proof is completed.

**Theorem A.2.** Let condition \( c_1^2 < c_2 \) hold. Let
\[
H(x) = vh(x) = vq_2\lambda x^2 + v[q_1\kappa_1\kappa_2 + q_2\lambda(\kappa_1 + \kappa_2)]x + v\kappa_1\kappa_2\lambda, \quad (A3)
\]
and \( x_1 \) and \( x_2 \) be two roots of \( H(x) \). Without loss of generality, assume \( \delta \neq a_1, a_2, a_3 \) given by electronic supplementary material, (11). If \( x_1 \) and \( x_2 \) are complex valued, either \( M(t) \) increases monotonically for all \( t > 0 \), or \( M(t) \) develops up-down-up dynamics. If \( x_1 \) and \( x_2 \) are real valued and \( x_1 < x_2 \), we have:

(1) If one of the following occurs: (i) \( -x_2 < [a_1, a_2, a_3, \delta] < -x_1, \) (ii) \( -x_2 < [a_1, a_2, \delta] < -x_1 < a_3, \) (iii) \( -x_2 < [a_1, \delta] < -x_1 < a_3, \) (iv) \( -x_2 < a_1 < a_3 < a_2 < -x_1 < \min(a_3, \delta), \) (v) \( -x_2 < a_1 < a_2 < -x_1 < \min(a_3, \delta), \) (vi) \( -x_2 < a_1 < -x_1 < \min(a_3, \delta), \) (vii) \( -x_2 < \delta < -x_1 < a_1, \) (viii) \( \delta < -x_2 < a_1 < a_2 < a_3 < -x_1, \) \( M(t) \) increases monotonically for all \( t > 0 \).

(2) If one of the following occurs: (i) \( a_1 < -x_2 < a_2 < a_3 < -x_1 < \delta, \) (ii) \( a_1 < -x_2 < a_3 < a_2 < -x_1 < \delta, \) (iii) \( a_1 < -x_2 < a_2 < -x_1 < \min(a_3, \delta), \) (iv) \( a_1 < -x_2 < a_3 < -x_1 < a_2 < -x_1 < \min(a_3, \delta), \) (v) \( -x_2 < a_1 < a_2 < -x_1 < a_3, \) (vi) \( -x_2 < a_3 < -x_1 < a_1, \) (vii) \( -x_2 < \delta < -x_1 < a_1, \) (viii) \( \delta < -x_2 < a_1 < a_2 < a_3 < -x_1, \) \( M(t) \) increases monotonically for all \( t > 0 \), or \( M(t) \) develops up-down-up dynamics.

(3) If one of the following occurs: (i) \( -x_2 < \min(a_1, \delta), \) (ii) \( a_1 < -x_2 < -x_1 < \min(a_3, \delta), \) (iii) \( a_2 < -x_2 < -x_1 < \min(a_3, \delta), \) (iv) \( a_3 < -x_2 < -x_1 < \delta, \) (v) \( \max(a_1, \delta) < -x_2 < a_1 < a_2, \) (vi) \( \max(a_1, \delta) < -x_2 < a_2 < a_1 < a_3, \) (vii) \( \max(a_1, \delta) < -x_2 < a_3 < a_1 < a_2, \) (viii) \( \delta < -x_2 < a_1 < a_2 < a_3 < -x_1, \) (ix) \( a_3 < \delta < -x_1 < a_1, \) (x) \( a_2 < -x_2 < a_3 < -x_1 < \delta, \) (xi) \( a_2 < -x_2 < [a_3, \delta] < -x_1, \) (xii) \( a_2 < [a_3, \delta] < -x_2 < a_3 < -x_1, \) (xiii) \( a_2 < -x_2 < [a_3, \delta] < -x_1, \) (xiv) \( a_3 < [a_3, \delta] < -x_2 < a_3 < -x_1, \) \( M(t) \) increases monotonically for all \( t > 0 \), or \( M(t) \) develops up-down-up dynamical profile.

**Proof.** The substitution of the initial condition (2.11) into the differential system (electronic supplementary material, (11)) gives \( M'(0) = 0, P'_c(0) > 0 \), and in turn, \( M'(0) = v P'_c(0) > 0 \). Consequently,
$M(t)$, $M'(t)$ and $M''(t)$ are all positive for $t > 0$ sufficiently small. The initial values (2.11) simplify the exact form of $M(t)$ (equation (2.9) of main text) in the form of

$$M(t) = \frac{v}{\delta} \frac{\kappa_1 \kappa_2 \lambda}{a_1 a_2 a_3} \frac{\beta_1}{a_1} e^{-a_1 t} - \frac{\beta_2}{a_2} e^{-a_2 t} - \frac{\beta_3}{a_3} e^{-a_3 t} - \frac{\beta_4}{\delta} e^{-\delta t},$$

where

$$\beta_1 = \frac{H(-a_1)}{(\delta - a_1)(a_2 - a_1)(a_3 - a_1)}, \quad \beta_2 = \frac{H(-a_2)}{(\delta - a_2)(a_1 - a_2)(a_3 - a_2)},$$

and

$$\beta_3 = \frac{H(-a_3)}{(\delta - a_3)(a_1 - a_3)(a_2 - a_3)}, \quad \beta_4 = \frac{H(-\delta)}{(a_1 - \delta)(a_2 - \delta)(a_3 - \delta)}.$$

Thus

$$M'(t) = \beta_1 e^{-a_1 t} + \beta_2 e^{-a_2 t} + \beta_3 e^{-a_3 t} + \beta_4 e^{-\delta t}. \tag{A 4}$$

We first consider the situation of complex valued $x_1$ and $x_2$. There are total four parameter cases and we only present the proof for the case $\delta < a_1 < a_2 < a_3$, as the other cases $a_1 < \delta < a_2 < a_3$, $a_1 < a_2 < \delta < a_3$ and $a_1 < a_2 < a_3 < \delta$ can be treated similarly. If $x_1$ and $x_2$ are complex valued, $H(x) > 0$ for all $x$. In the view of $\delta < a_1 < a_2 < a_3$, we obtain

$$\beta_1 < 0, \quad \beta_2 > 0, \quad \beta_3 < 0 \quad \text{and} \quad \beta_4 > 0. \tag{A 5}$$

For $t > 0$ sufficiently large, $M'(t)$ is dominated by $\beta_4 \exp (-\delta t)$ in (A 4) and is therefore positive. It follows from (A 4) that

$$[e^{\delta t} M'(t)]' = e^{(\delta - \delta_1) t} g(t),$$

with $g(t) = (a_3 - \delta) \beta_4 + (a_3 - a_1) \beta_1 e^{(\delta - a_1) t} + (a_3 - a_2) \beta_2 e^{(\delta - a_2) t}.$

It is seen that

$$g(0) = M''(0) > 0 \quad \text{and} \quad \lim_{t \to \infty} g(t) = (a_3 - \delta) \beta_4 > 0. \tag{A 7}$$

We can see from (A 6) that

$$M'(t) = e^{-a_1 t} G(t), \quad \text{where} \ G(t) = \int_0^t e^{(a_3 - \delta) s} g(s) ds, \quad t > 0. \tag{A 8}$$

It follows from (A 4) and (A 8) that

$$G(0) = 0, \quad \lim_{t \to \infty} G(t) = +\infty \quad \text{and} \quad G'(t) = e^{(a_3 - \delta)t} g(t). \tag{A 9}$$

Differentiating $g(t)$ gives

$$g'(t) = \frac{H(-a_1)}{a_2 - a_1} e^{(a_2 - a_1) t} \left[ e^{(a_2 - a_1) t} \frac{H(-a_2)}{H(-a_1)} \right].$$

If $H(-a_2) \leq H(-a_1)$, $g'(t) \geq 0$, and so $g(t) > 0$ for all $t > 0$. Hence (A 8) gives $M'(t) > 0$ for all $t > 0$. If $H(-a_2) > H(-a_1)$, $g'(t) < 0$ in $(0, t_0)$ for

$$t_0 = \frac{1}{a_2 - a_1} \left[ \ln(H(-a_2)) - \ln(H(-a_1)) \right],$$

and becomes positive for $t > t_0$. If $g(t_0) \geq 0$, $g(t) > 0$ for all $t > 0$ and $t \neq t_0$. Thus (A 8) gives again $M'(t) > 0$ for all $t > 0$. If $g(t_0) < 0$, (A 7) implies that $g(t)$ has two zeros $t_1 > 0$ and $t_2 > 0$ with $g(t) > 0$ in $(0, t_1) \cup (t_2, +\infty)$, and $g(t) < 0$ in $(t_1, t_2)$. It follows from (A 9) that

$$G'(t) > 0 \quad \text{for} \ t \in (0, t_1) \cup (t_2, +\infty) \quad \text{and} \quad G'(t) < 0 \quad \text{for} \ t \in (t_1, t_2). \tag{A 10}$$

If $G(t_2) > 0$ for all $t > 0$ and $t \neq t_2$, and so $M'(t) > 0$ for all $t > 0$ and $t \neq t_2$. We recall that $G(t) > 0$ for $t > 0$ both sufficiently small and large. Thus, if $G(t_2) > 0$, (A 10) implies that $G(t)$ has two zeros $t_3 > 0$ and $t_4 > 0$ with $G(t) > 0$ in $(0, t_3) \cup (t_4, +\infty)$, and $G(t) < 0$ in $(t_3, t_4)$. It follows from (A 8) that $M'(t) > 0$ in
We present the proof for the case \( x_2 < a_1 < a_2 < \delta < -x_3 < a_3 \), as the other cases can be dealt with by the same idea. As
\[
-x_2 < a_1 < a_2 < \delta < -x_3 < a_3 \Rightarrow \beta_1 < 0, \beta_2 > 0, \beta_3 < 0, \text{ and } \beta_5 < 0,
\]
using (A 4) again, we obtain
\[
[e^{a_3}M'(t)]' = e^{(a_3 - a_1)} g_1(t),
\]
\[
g_1(t) = (a_2 - a_1) \beta_3 + (a_2 - \delta) \beta_4 \delta^{(a_2 - \delta)} + (a_2 - a_3) \beta_5 \delta^{(a_2 - a_3)}, \tag{A 12}
\]
with \( f(0) = M''(0) > 0 \) and \( \lim_{t \to \infty} g_1(t) = (a_2 - a_1) \beta_1 < 0 \). In terms of (A 11) and (A 12), we have
\[
(a_2 - \delta)(a_1 - \beta_3) < 0 \text{ and } (a_2 - a_3)(a_1 - \beta_3) > 0, \text{ and so}
\]
\[
f'(t) = (\delta - a_2)(a_1 - a_2) \beta_2 \delta^{(\delta - a_2)} + (\delta - a_3)(a_1 - a_3) \beta_3 \delta^{(\delta - a_3)} < 0.
\]
Hence, there exists a \( t_5 > 0 \) such that \( g_1(t) > 0 \) in \((0, t_5)\), \( g_1(t_5) = 0 \), and \( g_1(t) < 0 \) in \((t_5, \infty)\). For \( t > 0 \) sufficiently large, \( M'(t) \), dominated by the term \( \exp (a_2 - a_1) \), is negative. We also recall that \( M'(t) > 0 \) for \( t > 0 \) sufficiently small. Thus, there exists a finite \( t_6 > 0 \) such that \( M'(t) > 0 \) in \((0, t_6)\), and \( M'(t_6) = 0 \). Then, \( M'(t_6) \leq 0 \), and (A 12) gives \( f(t_6) = \exp (a_1 t_6) M''(t_6) \leq 0 \). It follows that \( t_6 \geq t_5 \) and \( f(t) < 0 \) for all \( t > t_6 \). Hence, by using (A 12) again, for each \( t > t_6 \), we have
\[
M'(t) = e^{-a_1} \int_{t_6}^{t} e^{(a_2 - a_3) s} f(s) ds < 0.
\]
(2) We present the proof for the case \( a_1 < a_2 < x_2 < \delta < -x_1 < a_3 \), as the other cases can be handled similarly. With this specification, we have
\[
\beta_1 > 0, \quad \beta_2 < 0, \quad \beta_3 < 0 \quad \text{and} \quad \beta_5 < 0. \tag{A 13}
\]
By (A 4) and (A 13), we obtain for all \( t > 0 \)
\[
[e^{a_1}M'(t)]' = (a_1 - \delta) \beta_2 \delta^{(a_1 - \delta)} + (a_1 - a_2) \beta_3 \delta^{(a_1 - a_2)} + (a_1 - a_3) \beta_5 \delta^{(a_1 - a_3)} > 0.
\]
As \( M'(0) = 0 \), it shows that \( e^{a_1}M'(t) > 0 \), and therefore, \( M'(t) > 0 \) for all \( t > 0 \).

(3) We give a short description for the proof with the additional condition that \( x_1 < x_2 < \delta < x_1 < a_2 < a_3 \), as the other cases can be proceeded in an analogous manner. Under this extra condition, (A 5) holds again. Thus, the remaining discussion is the same as the proof given above for two complexes valued \( x_1 \) and \( x_2 \), and we omit it here.

References

1. Munkiv B, Neuret G, van Oudenaarden A. 2012 Using gene expression noise to understand gene regulation. Science 336, 183–187. (doi:10.1126/science.1216379)
2. Dar RO. 2012 Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc. Natl Acad. Sci. USA 109, 17454–17459. (doi:10.1073/pnas.1215330109)
3. Larsson AIM. 2019 Genomic encoding of transcriptional burst kinetics. Nature 565, 251–254. (doi:10.1038/s41586-018-0836-1)
4. Choudry S, Kondrov J, Sanchez A. 2015 Deciphering transcriptional dynamics in vivo by counting nascent RNA molecules. PloS Comput. Biol. 11, e1004345. (doi:10.1016/j.pcbi.2013.11.2123)
5. Sepúlveda LA, Xu H, Zhang J, Wang M, Golding I. 2016 Measurement of gene regulation in individual cells reveals rapid switching between promoter states. Science 351, 1218–1222. (doi:10.1126/science.aad0635)
6. Lloyd-Price J. 2016 Dissecting the stochastic transcription initiation process in live Escherichia coli. DNA Res. 23, 1–12. (doi:10.1093/dnares/dsw009)
7. Bhogale PM, Sorg RA, Veening JW, Berg J. 2013 What makes the lac-pathway switch: identifying the fluctuations that trigger phenotype switching in gene regulatory systems. Nucleic Acids Res. 42, 11321–11328. (doi:10.1093/nar/gkt139)
8. Suter DM, Molina N, Gatliff D, Schneider K, Schieber U, Naef F. 2011 Mammalian genes are transcribed with widely different bursting kinetics. Science 332, 472–474. (doi:10.1126/science.1198817)
9. Larson DR. 2011 What do expression dynamics tell us about the mechanism of transcription.
10. Zhou T, Zhang J. 2012 Analytical results for a multisstate gene model. SIAM J. Appl. Math. 72, 789–818. (doi:10.1137/110852887)
11. Sun Q, Tang M, Yu J. 2012 Temporal profile of gene transcription noise modulated by crosstalkin signal transduction pathways. Bull. Math. Biol. 74, 375–398. (doi:10.1007/s11538-011-9683-2)
12. Sun Q, Tang M, Yu J. 2012 Modulation of gene transcription noise by competing transcription factors. J. Math. Biol. 64, 469–491. (doi:10.1007/s00285-011-0420-x)
13. Munsky B, For Z, Neuret G. 2015 Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics.
